

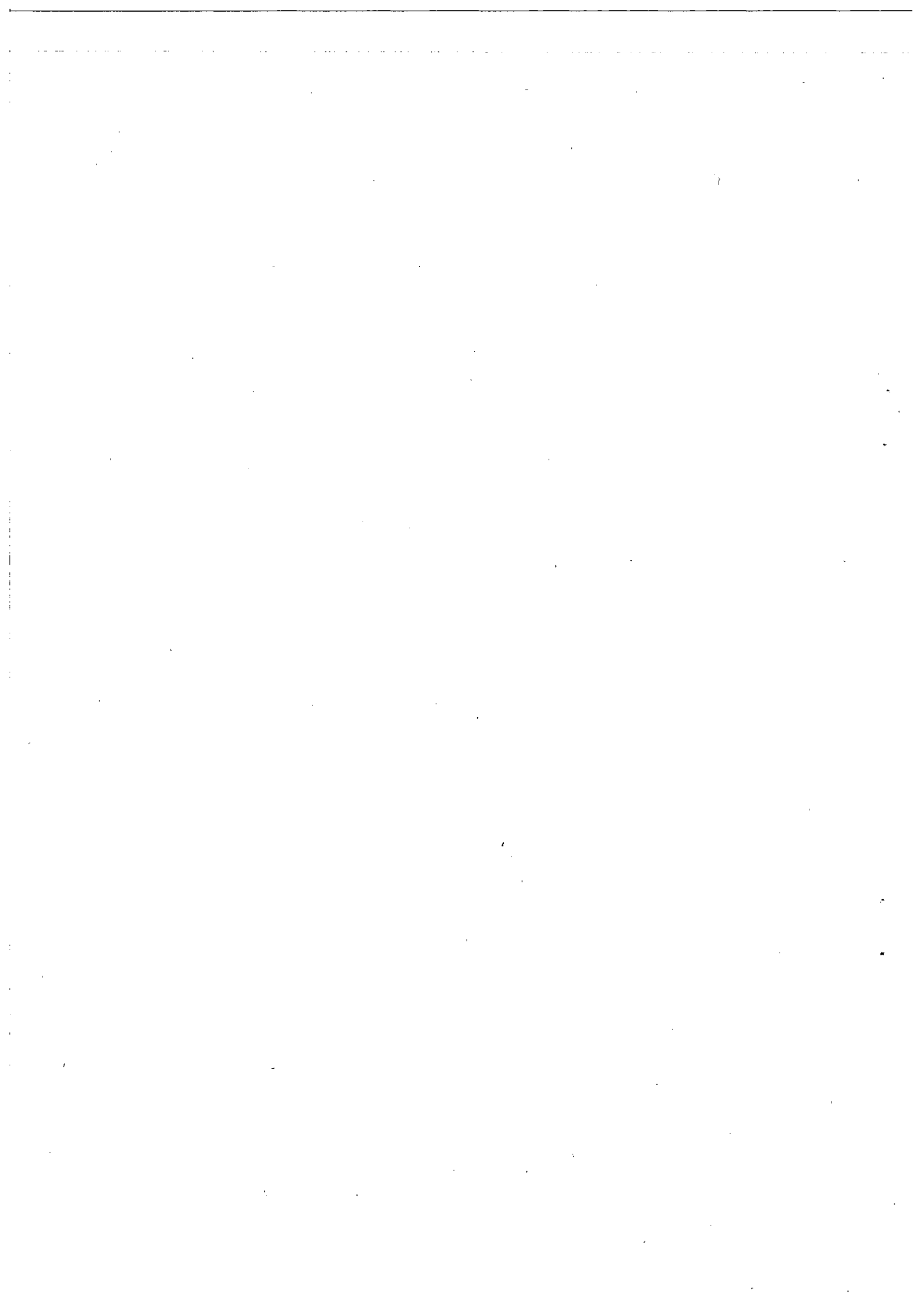
No.970

**False Beliefs and Game Theory: Implications from
the Japanese Comic Story Konnyaku Mondô**

by

Mamoru Kaneko and J. Jude Kline

January 2002



False Beliefs and Game Theory: Implications from the Japanese Comic Story Konnyaku Mondô*

Mamoru Kaneko[†] and J. Jude Kline[‡]

January 30, 2002 (diamj)

Abstract

This is a dialogue between two professional people on the new field called “epistemic logics and game theory”. One speaker is a specialist who has been working in this field for a long time, and the other is a game theorist who is quite new to the field and younger than the specialist. They start discussing the Konnyaku Mondô and find that it has many implications for game theory in terms of its foundations and its scope.

[*Setting:* Jan Hummer, a lecturer from a foreign land, is visiting a well known institution in a remote research village north-east of Tokyo. Kurai Shinzuki, a prominent professor at the institute, has agreed to engage in discussions with the visitor. Acts 1 and 2 take place in a laboratory with dim lighting and inadequate heating]

1. Act: *Konnyaku Mondô*

Jan: Hey, Shinzuki, I heard from your graduate students that you had discussed a Japanese comic story, called something like the Cognac Melon. They claimed it is quite interesting and important for game theory. Can you tell me that story?

Shinzuki: Ha, ha. I think it must be the Konnyaku Mondô, instead of the Cognac Melon. I can, of course, explain the story to you. Didn't my graduate students give an adequate explanation?

Jan: No, they didn't. They seemed unable to explain it since they did not read the story, but heard it only from you. One of them claimed it is written in

*The authors are partially supported by Grant-in-Aids for Scientific Research No.106330003, Ministry of Education, Science and Culture.

[†]Institute of Policy and Planning Sciences, University of Tsukuba, Ibaraki 305-8573, Japan (kaneko@shako.sk.tsukuba.ac.jp)

[‡]Department of Economics, School of Business, Bond University, Gold Coast, QLD 4229, Australia, (Jeffrey_Kline@bond.edu.au)

classical Japanese which is only understandable by some old professors. Was it written so long ago?

Shinzuki: No, no, it was written from an oral presentation only about 100 years ago.¹ It is difficult for young people, in fact, I confess I feel it is slightly difficult for me too.

Jan: Huhm..., Japanese has changed a lot in only 100 years, hasn't it? Anyway, let's have the Cognac Melon..., sorry, Konnyaku Mondô.

[Shinzuki assumes a more professorial mode of speaking]

Shinzuki: The literal translation of the Konnyaku Mondô is "Devil's Tongue Jelly Dialogue".

Jan: What? What is "devil's tongue jelly"?

Shinzuki: It is a food product, like a jelly, made from roots of the Konnyaku plant. You had it in the sake restaurant we went to last week. It is usually brown and looks like a devil's tongue.

Jan: Yes, I remember it. But, why is that funny jelly food important to game theory?

Shinzuki: The jelly food itself is not relevant to game theory, of course. The story of the Konnyaku Mondô is related to game theory in many respects. It is about false beliefs and the subjective nature of peoples' thoughts. It is an example of how inconsistent subjective thoughts may evolve, even with common knowledge and communication.

Jan: Is that true? It sounds incorrect, or if it is true, it has serious implications for game theory. Please explain the story to me?

Shinzuki: I am not sure I will be able to reconstruct the entire story in a manner meaningful to you. Nonetheless, I shall make an attempt.

[Shinzuki now assumes an even more professorial, but somewhat dramatic, tone of a storyteller]

"There was a temple without a monk for some period. A Konnyaku maker, named Rokubei, who had lived next to the temple, moved into the temple and started pretending to be a monk. One day, a Zen Buddhist monk visited the temple and challenged the new master of the temple to a dialogue on Buddhist thoughts. Since Rokubei had no idea about Buddhist dialogue, he refused at first but eventually agreed."

Jan: OK, we have the food producer, Robeki, who has no knowledge on Buddhist thoughts about to engage in a dialogue on Buddhist thoughts with a Buddhist monk. It looks like a sure victory. How possibly could there be any implications for game theory? By the way, are these two the only characters in the story? No beautiful princes?

¹pp.61-70 in [14].

Shinzuki: First of all, the jelly maker's name is Rokubei, not Robeki. Secondly, you should be patient. A lot of implications will be found after our story concludes. Finally, another character is involved. His name is Hachigoro and his role is to be a witness to the entire story. I am sorry you are disappointed by finding no love stories or battles of the sexes included here.

Let me continue the story.

"Since Rokubei did not know how to communicate with the monk on Buddhism, he did not answer the monk's questions. The monk thought that being silent might be a style of dialogue, and tried asking questions in several different fashions. After some time, Rokubei began answering the monk's questions with hand gestures. Taking this as a style of dialogue, the Monk began responding with hand gestures as well. Eventually, both Rokubei and the monk agreed that Rokubei had defeated the monk. Then the monk left as a loser."

Jan: Why? I do not understand how Rokubei defeated the monk. What happened?

Shinzuki: I am glad to hear you have found some strange nature in the story. Perhaps here you will find the implications to game theory you are searching for. As you shall see, Hachigoro both witnesses the event and plays an important role by questioning the outcome. The concluding part of the story is as follows:

"After the dialogue, Hachigoro followed the monk and asked him about the dialogue. The monk answered that the master had expressed great Buddhism thoughts through his gestures and he should be respected. Hachigoro wondered about when, if ever, Rokubei had learned Buddhism thoughts, and he returned to the temple to ask him about it. Rokubei said he had never learned Buddhism thoughts. Rather, the monk started talking badly about his jelly products with his gestures. This angered Rokubei, and thus he beat the monk."

Jan: Wait a second. I understand the monk talked to Rokubei with gestures about Buddhism thoughts. Did the gestures by Rokubei have different meanings? What are they?

Shinzuki: In fact, in the full form of the story, Hachigoro heard the meanings of the gestures intended by the monk, as well as by Rokubei. For example, a gesture of creating a circle with one's arms means the universe for the monk, but it means a round Konnyaku product for Rokubei. All the gestures are meaningful to each person.

Jan: I understand the story itself, but where are the implications for game theory?

Shinzuki: All right, now, perhaps you are ready to understand the game theoretical implications. Each of Rokubei and the monk believed that they had a perfectly meaningful dialogue, and that the victory of Rokubei was common knowledge. However, the monk believed that they had a Buddhism dialogue, while Rokubei believed that they had discussed his jelly products.

Jan: Great. I understand your explanations about the story, but I am still not able to see any implications for game theory. May I ask some questions?

[Without waiting for a response, Jan charges on]

First of all, did the gestures they exchanged become common knowledge? Secondly, is the victory of Rokubei common knowledge between them? I believe that your answer to these questions will be yes.

However, these people attached completely different meanings to their gestures. They do not share a basic understanding of what the gestures mean. In such a situation, can we say that they share the common knowledge of the victory of Rokubei as well as the gestures exchanged?

[Shinzuki becomes slightly overwhelmed]

Shinzuki: How can you ask such sharp questions to me? Let me think about them, uh...

I would like to answer "Yes" to your first questions. The gestures exchanged and Rokubei's victory became common knowledge. Here, it is at least necessary to assume that their visions are normal. The speed of the light is almost infinity, which is actually about 300,000km/s. Through their visions, they are able to verify in an instant, almost infinite observations of a gesture and the other's observation of that gesture. Hence, the common knowledge of the gestures and of Rokubei's victory, expressed and understood with gestures, is an adequate assumption. Incidentally, I found this argument while reading the "analogy of the sun" in Chapter 6 of Plato's *Republic*.

Now more to the point, the symbolic forms of the gestures can be assumed to be common knowledge. But these symbolic forms have completely different meanings to Rokubei and the monk. Therefore, the gestures and victory are common knowledge at a superficial level. In fact, they lead those people to profound misunderstandings of each other. Nevertheless, they still have the common knowledge of the gestures and victory.

Jan: Yes, you understand the point of my questions. When we talk about common knowledge or even about knowledge, the truthfulness is somehow assumed. In our story, the common knowledge of the gestures cannot be true since the two people attached different meanings to them.

Shinzuki: Isn't the common knowledge of the gestures true? Huhm..., I think your reasoning may be correct, but your conclusion is not true. The objects of common knowledge in our case are the symbolic forms of the gestures exchanged. The common knowledge here does not include the interpretations of the gestures. May I use another analogy?

Jan: Of course, but do not digress too much.

Shinzuki: A Japanese apple farmer would like to sell his apples, which are dear to him and should be eaten fresh. An American visiting Japan wants apples to make apple pies. Without communicating their intentions, the exchange of apples and money occurs between these two people. With communication between

the two, the exchange might not have occurred. But my point here is that apples are apples independent of the intentions attached to them by the farmer and the visitor. I would like to call this the “analogy of apples” in analogy with Plato’s “analogy of the sun”.

Jan: Once again, you have told a nice story, but where is the relevance to our discussion?

Shinzuki: I am sorry, perhaps I should return to the truthfulness of the common knowledge of gestures. The truthfulness of the knowledge of a gesture is the accurate observation and recollection of the gesture. The intended interpretations of gestures are irrelevant as are the intentions of the buyers and sellers of apples.

Jan: Wow. You have brought out many interesting and puzzling aspects of the story. I will need some time to digest them. I should go home now and it is getting cold. Let’s continue our discussion tomorrow morning, and could we start talking about game theory?

Shinzuki: Yes, but tomorrow I will have a class until 10:30. Let’s meet around 10:45. I will bring another heater.

2. Act: False Beliefs and Decision Making

[In the lab, Jan is waiting for Shinzuki, and then Shinzuki appears in a good mood carrying a heater]

Jan: How was your class?

Shinzuki: Good, of course.

Jan: Last night I thought about the story more. Now I feel it may be related to game theory. I am anxious about hearing the implications of the story for game theory. However, since you have a tendency to digress, I would like to impose a constraint on our discussions. We should directly discuss the implications to game theory and its relationship to the story. The use of analogies should be avoided.

Shinzuki: Hhum..., You are harsh, but I will try my best to play within your rules.

Let us consider a very simple situation where players 1 and 2 play the prisoner’s dilemma game. Could you please draw its payoff matrix on the blackboard and name it Figure 0?

Jan: Sure, it is an easy task. I name the prisoner’s dilemma game $g^0 = (g_1^0, g_2^0)$, since I expect you want to have other games.

	s_{21}	s_{22}
s_{11}	(5, 5)	(1, 6)
s_{12}	(6, 1)	(3, 3)*

Figure 0: $g^0 = (g_1^0, g_2^0)$

Shinzuki: Yes, you are right. I want to give two more games now, which I myself write on the blackboard.

	s ₂₁	s ₂₂
s ₁₁	(5, 0)	(1, 0)
s ₁₂	(6, 0)	(3, 3)*

Figure 1: $g^1 = (g_1^1, g_2^1)$

	s ₂₁	s ₂₂
s ₁₁	(0, 5)	(0, 6)
s ₁₂	(0, 1)	(3, 3)*

Figure 2: $g^2 = (g_1^2, g_2^2)$

Now, suppose that player 1 thinks that the game is given as $g^1 = (g_1^1, g_2^1)$, while player 2 thinks that the game is $g^2 = (g_1^2, g_2^2)$. Each of them thinks about the common knowledge of a different game. Although the standard game theory literature assumes the common knowledge of the game structure, it would be apparently impossible to maintain this assumption in the current context. We modify this assumption to require that each player $i = 1, 2$ believes the common knowledge of the game $g^i = (g_1^i, g_2^i)$.

Jan: What? What do you mean by “Each player i believes the common knowledge of $g^i = (g_1^i, g_2^i)$ ”? Perhaps, I should verify if I understand you correctly. Huhm..., it is difficult to understand your statement. I should paraphrase it into a longer form: “Player i personally believes that it is common knowledge that $g^i = (g_1^i, g_2^i)$ is the game to be played.” Is this paraphrasing accurate?

Shinzuki: Yes, certainly.

Jan: All right. Now, it is getting clearer. You seem to allow the falsity of a belief, since otherwise the personal belief part could be redundant and the assumption set as a whole would fall apart. The truthfulness of the common knowledge of $g^i = (g_1^i, g_2^i)$ is assumed only in the belief of player i . Is my understanding OK?

Shinzuki: You correctly understand my assumptions.

Jan: Great.. Now, I should ask what kind of mathematical language you will use to represent your game theoretical argument. I would be very surprised if you talked about such a delicate problem without having any mathematical formulation.

Shinzuki: I will talk about that problem in the mathematical language called *epistemic logic*. I borrow the logic language from [5] and [7].² Using this language, the game theoretical assumptions I uttered and one logical conclusion derived from them are described as follows on the blackboard:

$$(1) \vdash g^0, B_1C(g^1), B_2C(g^2) \rightarrow C(\text{Nash}(s_{12}, s_{22})).^3$$

²Epistemic logic is a branch of modal logic. For a general introduction to modal logic, see Chellas [1] and Hughes-Cresswell [3]. Fagin et al [2] and Meyer-van der Hoek [10] are introductory books treating epistemic logics.

³The provability \vdash depends upon a logical system. In this paper, various logical systems are involved without specifying any. For details, see Kaneko [5], Kaneko-Suzuki [7] and Kaneko et al [8].

Some explanations about symbolic expressions should be given now. The first symbol, \vdash , means the entire statement following it is provable. The left-hand side following the first symbol \vdash and continuing before the arrow \rightarrow consists of assumptions, or axioms. To the right of the arrow \rightarrow we find another statement which is logically concluded from the left-hand side.

Jan: Let me interpret what you wrote on the blackboard. The first symbol in the left-hand side after \vdash seems to mean that the game to be played is the prisoner's dilemma game g^0 . The next two statements following commas seem to be that: player 1 believes⁴ the common knowledge of g^1 , and player 2 believes the common knowledge of g^2 . The right-hand side seems to mean that it is common knowledge that (s_{12}, s_{22}) is a Nash equilibrium. You said that the arrow means that the right-hand side is logically concluded from the left-hand side, didn't you? You also said that the symbol \vdash means that the entire statement following it is provable. Isn't \vdash redundant?

Shinzuki: Without the symbol \vdash , we have only the statement "if the left-hand side is assumed, then the right-hand side holds". The additional \vdash means that the "if-then" clause is provable in an epistemic logic.⁵

Jan: OK, OK. (1) states that it is provable that if g^0 , $B_1C(g^1)$ and $B_2C(g^2)$, then the strategy pair (s_{12}, s_{22}) as a Nash equilibrium is common knowledge. It sounds alright, but the beliefs and knowledge are mixed. In the left-hand side, the beliefs of common knowledge are assumed, while in the right-hand side, the common knowledge of a Nash equilibrium is obtained. The beliefs seem to have somehow vanished and been replaced entirely by knowledge as if by magic. How did it happen?

Shinzuki: In fact, (1) is true. This is parallel to the Konnyaku Mondô in that both statements include false beliefs, but the conclusions are common knowledge. However, I admit that some unparalellism between beliefs and knowledge is involved in (1). For a better understanding, I retreat here from (1) to the following slightly weaker assertion:

$$(2) \vdash g^0, B_1C(g^1), B_2C(g^2) \rightarrow B_1C(\text{Nash}(s_{12}, s_{22})) \wedge B_2C(\text{Nash}(s_{12}, s_{22})).$$

We shall discuss the relationship of these to the Konnyaku Mondô later.

Jan: Now, (2) is nicer than (1), since each "believes" the common knowledge of something in both the left-hand and right-hand sides. Huhm..., the belief of

⁴In the game theory literature, belief is often understood as subjective probability in the sense of Savage [12]. The concept of belief here differs from subjective probability in many respects. The main difference is that we discuss beliefs in association to the logical abilities of players, while beliefs are treated as a black box given together with tastes in the Savage approach.

⁵Note that it may be the case that neither the "if-then" clause nor the "if-then not" are provable, i.e., they are undecidable. This may mean that a player can find neither some conclusion nor its negation from his basic beliefs (the left-hand side). The logic approach often evaluates such provability or unprovability.

the common knowledge of a Nash equilibrium is derived from the individual belief of the common knowledge of the game g^i . This logical calculation takes place in the mind of each player. Then what is the role of the objective statement g^0 ?

Shinzuki: Oh my dear, Jan, I am pleased to find you understand the problem perfectly. In fact, (2) is equivalent to the following two separated statements, neither of which involves g^0 :

$$(2a) \vdash B_1C(g^1) \rightarrow B_1C(\text{Nash}(s_{12}, s_{22})).$$

$$(2b) \vdash B_2C(g^2) \rightarrow B_2C(\text{Nash}(s_{12}, s_{22})).$$

The derivation of (2a) and (2b) from (2) needs an advanced technique of logic, while the converse is just a calculation.

Jan: It sounds wonderful, though I don't appreciate the difficulty in the derivation from (2) to (2a) and (2b). Now on the issue of game theoretical implications, consider the decision making of player 1. He believes that his payoff function is g_1^1 , which is the same as his payoff function g_1^0 of the objective game g^0 . One might argue that he chooses his strategy s_{12} , because it is a dominant strategy rather than because it is part of a Nash equilibrium. Following this line of reasoning, we seem to be moving even further from common knowledge.

Shinzuki: All right, I can change (2a) into

$$(3a) \vdash B_1(g^1) \rightarrow B_1(\text{Dom}_1(s_{12})).$$

Jan: I am surprised you changed your statement so easily. However, (3a) is in some sense nicer than (2a), since it includes no common knowledge. Huhm..., (3a) still includes player 2's payoff function in $B_1(g^1)$. Can you change (3a) into the following?

$$(3a') \vdash B_1(g_1^1) \rightarrow B_1(\text{Dom}_1(s_{12})).$$

Shinzuki: Yes, of course. Are you satisfied by having a purely personalized version?

Jan: Yes, I am very satisfied. Uhum..., it raises two opposite thoughts in my mind. On the one hand, (3a') makes sense perfectly for me since player 1 simply derived s_{12} as a dominant strategy from the belief of his payoff function g_1^1 . On the other hand, (3a') no longer involves false beliefs since $g_1^1 = g_1^0$.

Shinzuki: Sorry about it. I change (3a'), again into

$$(3a'') \vdash B_1(g_1^2) \rightarrow B_1(\text{Dom}_1(s_{12})).$$

Please notice I changed the true payoff function $g_1^1 = g_1^0$ into the false payoff function g_1^2 of Figure 2.

Jan: Once again you have flippantly adjusted the assumptions. I would like to understand the whole problem, instead of each small piece and it is difficult to concentrate when you keep changing assumptions. We seem to be moving further and further away from our initial objective. While I applaud you for refraining from the use of analogies, I must admit you are weaving an impossible web in my head with your incessant manipulation of assumptions. Please, stop and return to something concrete. I almost forget the point.

Shinzuki: Aha, I return to a full statement now:

$$(3) \vdash g^0, B_1(g_1^2), B_1C(g_1^2, g_2^1) \rightarrow B_1(\text{Dom}_1(s_{12})) \wedge B_2C(\text{Nash}(s_{12}, s_{22})).$$

Surely, we can talk about the falsity of beliefs of players in (3).

Jan: Although this looks slightly different from our starting point, it does seem to be moving back in the right direction. Let me try to make sense of (3). I start with the left-hand side of the arrow \rightarrow . Uhum..., this involves a lot of falsities. Probably, for my own sanity, I should translate (3) into English. First, it says the objective situation is described by the prisoner's dilemma game g^0 . Next, it says that player 1 believes his payoff function is given by g_1^2 , which we know to be false relative to g^0 . Next, it says that player 2 believes the game (g_1^2, g_2^1) is common knowledge between the players, which we know to be false both relative to the objective game g^0 and the content of player 1's belief.

In this formulation, player 1 ignores, or does not discuss, player 2's payoff in the formation of his beliefs, while player 2 believes they have common knowledge of (g_1^2, g_2^1) which we know to be false relative to g^0 . Is my understanding of the left-hand side of (3) accurate?

Shinzuki: Yes, it is.

Jan: Good. Now let's attack the right-hand side. The first piece is the derivation by player 1 that his strategy s_{12} is a dominant strategy. This derived statement happens to be objectively true for the game g^0 . We also, find the derivation by player 2 that the pair (s_{12}, s_{22}) as a Nash equilibrium is common knowledge. The pair (s_{12}, s_{22}) to be a Nash equilibrium is also true relative to g^0 . However, player 2 believes that this fact is common knowledge between 1 and 2, while player 1 does not think about player 2 at all. This is another kind of falsity, which seems different from falsity relative to the objective game.

Shinzuki: Indeed, we meet a lot of falsities in beliefs, in particular, both sides of (3) involve falsities. Actually, I deliberately wrote (3) in a puzzling form.

[Jan answers in a tone of disagreement]

Jan: Yes, (3) has a lot of falsities. Hhum..., however, the falsities on both sides of the arrow \rightarrow are parallel. The previous asymmetry of falsities in (1) is more interesting, since the true common knowledge is derived from the false beliefs of common knowledge. As a game theorist, I would typically assume that the objective truths on the right-hand side were obtained from objective truths

in the minds of the players. You have shown that this need not be the case. Is this distinction important for game theory? Can you connect this finding in (1) to the Konnyaku Mondô discussed yesterday?

But now my head is reeling..., and noises are rising within me.

Shinzuki: I think you are saying you are hungry. Why don't we go to the Mexican Restaurant? The lunches there are cheap and large. Although I would like to continue our discussions right after lunch, I will have another stupid meeting until 3:00PM. Is it convenient for you to have more discussions after 3 o'clock?

Jan: Yes, of course. Anyway, it is time to eat. Let's go.

3. Act: More on Decision Making

[After lunch, Shinzuki went to a meeting, and Jan has been taking a nap on a couch in the lab.]

Shinzuki: Hi Jan, I am back from my hell meeting. Ah, I am sorry to wake you up.

Jan: No, no, thank you for waking me up. I wanted to wake up by myself. I was very sleepy because of the intense discussions in the morning and good food in the restaurant. How was your meeting?

Shinzuki: Nothing important as usual. Actually, I didn't sleep well in the meeting because of the loud voices of the chairperson.

Jan: Ha ha, you are in almost the same state as me. By the way, I thought about our discussions during my nap. Your explanations diverged quite a lot.

Shinzuki: I have a more general statement including the previous ones, which is written on the blackboard as:

$$(4) \vdash \Gamma^0, B_1(\Gamma_1), B_1(\Gamma_2) \rightarrow B_1(D_1(s_1)) \wedge B_2(D_2(s_2)).$$

Here, the first assumption set Γ^0 is the objective description of the situation, and $B_i(\Gamma_i)$ is the beliefs owned by player $i = 1, 2$. The right-hand side means that each player i derives his decision s_i . Again, the symbol \vdash means that the entire sentence is provable.

Jan: Shinzuki, please stop for a moment and let me digest your current statement. First, I like the fact that it does not seem to involve common knowledge at all. Our previous discussions led me to find that common knowledge is only in the mind of a player. This suggests that we can do away with common knowledge operators and focus only on beliefs. This appears to be what you have done. However, you also replaced g^0 by Γ^0 , and that worries me. If you included some "common knowledge" there, then I would not be so happy.

Shinzuki: We can suppose that Γ^0 contains purely objective statements and no instances of "common knowledge" or beliefs.

Jan: Good. Then I can go further and try to understand (4) now. Let me put (1), (2), (3), and (4) next to each other on the blackboard.

$$(1) \vdash g^0, B_1C(g^1), B_2C(g^2) \rightarrow C(\text{Nash}(s_{12}, s_{22})).$$

$$(2) \vdash g^0, B_1C(g^1), B_2C(g^2) \rightarrow B_1C(\text{Nash}(s_{12}, s_{22})) \wedge B_2C(\text{Nash}(s_{12}, s_{22})).$$

$$(3) \vdash g^0, B_1(g_1^2), B_1C(g_1^2, g_2^1) \rightarrow B_1(\text{Dom}_1(s_{12})) \wedge B_2C(\text{Nash}(s_{12}, s_{22})).$$

$$(4) \vdash \Gamma^0, B_1(\Gamma_1), B_1(\Gamma_2) \rightarrow B_1(D_1(s_1)) \wedge B_2(D_2(s_2)).$$

Mathematically speaking, both (2) and (3) appear to be special cases of (4), but (1) is not.

Shinzuki: Uhum.., (1) is not a special case. However, we can include (1) by changing (4) slightly.

Jan: Ha, ha, ha. I do not care about generality. Rather, I am curious about the treatment of strategies here. I find that in (4) each player i has a decision s_i , while in (1) and (2) each player thinks about the Nash strategy pair (s_{12}, s_{22}) , and in (3), player 1 thinks about s_{12} , but 2 thinks about (s_{12}, s_{22}) . In this sense, they are asymmetric. More specifically, in (3) s_{12} is 1's decision, but what is (s_{12}, s_{22}) for player 2? I ask this because player 2 cannot choose s_{12} .

Shinzuki: That is a good question, since I know the answer. The strategy s_{12} appears as a prediction of 1's decision making in 2's mind. Player 2 doesn't choose s_{12} , but he thinks about 1's decision making since he requires it to predict 1's decision making. To a large extent, this argument is rather standard from the game theoretical analysis of Nash [11].

Jan: But if (4) is a generalization of (2) and (3), then why doesn't it include predictions?

Shinzuki: Aha, (4) may include predictions in $D_i(s_i)$, though they are not explicitly written. In the case of (3a), since a dominant strategy is the decision criterion, the prediction part is not included at all.

Jan: OK, so again we move further from common knowledge. Do we really need common knowledge?

[Shinzuki speaks now in an authoritative tone]

Shinzuki: Your question is naive, and the issue has many aspects. First, we need to think about the treatment of common knowledge and common beliefs, and more fundamentally, the difference between knowledge and belief. Second, we should think about the relationship between common knowledge and decision making. We found by our analysis of (3) where player 1 has a dominant strategy, that common knowledge is "not necessarily" needed for decision making. Are you interested to find when common knowledge is necessary for decision making?

Jan: I would like to hear about the distinction between knowledge and belief now, but OK since you seem to want to, why don't you tell me about when common knowledge is needed?

Shinzuki: Yah, yah..., I promise to discuss the distinction between knowledge and belief later. But for now, let's discuss when common knowledge is needed. You will find that the argument is amazingly standard in the game theory literature. Consider the following form of prediction-decision making:

- (a) 1 maximizes his payoff predicting that 2's decision is governed by (b);
- (b) 2 maximizes his payoff predicting that 1's decision is governed by (a).

Jan: Hold on..., (a) appears in (b), and (b) appears in (a). This seems to involve circular reasoning. The sentence (a) can be plugged into (b), and (b) into (a), and so on, forever. This process yields an infinite regress, and I have a great head ache with such circularities and infinite regress. Is there any meaningful solution?

Shinzuki: Your concern is valid. To treat that infinite regress, we need a common knowledge extension of an epistemic logic. In such an extension, we find a complete solution which turns out to be the common knowledge of Nash equilibrium. If (a) and (b) are restricted to occur in the mind of a single player, then the solution is the personal belief of the common knowledge of Nash equilibrium, i.e., $B_i C(\text{Nash}(s_1, s_2))$.⁶

Exactly speaking, (a) and (b) are properties required for a type of decision criterion. It is my claim that if such a decision criterion is adopted, then common knowledge is necessarily involved.

[Shinzuki start speaking in a low voice with a critical tone]

Some people argued that common knowledge is unnecessary for Nash equilibrium. But that is not the issue! The issue is the necessity of common knowledge for decision making, not Nash equilibrium. Nash equilibrium is Nash equilibrium, as an apple is an apple in the analogy of apples, and Nash equilibrium should be distinguished from the description of decision making.

Jan: You seem to mean that the necessity of common knowledge may depend upon the decision criterion used. Hummm,...the dominant strategy criterion does not involve common knowledge, but it might not give a decision to a player in some games. Is there any decision criterion that always gives a decision and does not involve common knowledge?

Shinzuki: Actually, there is the simplest example called the *default decision criterion*. It recommends a prespecified strategy, say, the first strategy, without requiring the decision maker to think about anything. If you think it is simple, I can give more examples based on this.

⁶See Kaneko [5].

Jan: I am getting tired...., but the last point sounds very nice. Maybe we should have a coffee break. But before that, I have one thought I would like to bring out. Common knowledge involves an infinite regress, or infinite depths of nested beliefs structures. On the other hand, if the dominant strategy decision criterion or the default criterion is adopted, then no common knowledge is required, since there is no nesting of beliefs, or depth of beliefs is only 1.

Shinzuki: Yes, you are correct, as already I said there are many based on the default criterion. Would you like to see an example involving a finite depth of beliefs greater than 1 without using the default criterion?

Jan: I am tempted by coffee, but please go on.

Shinzuki: The example is very simple and obtained from changing only one payoff of player 2 in the prisoner's dilemma:

	s ₂₁	s ₂₂
s ₁₁	(5, 5)	(1, <u>2</u>)
s ₁₂	(6, 1)	(3, 3)*

Figure 0': $g^{0'} = (g_1^{0'}, g_2^{0'})$

In this game, player 2 has no dominant strategy, but if player 2 believes that player 1 follows the dominant strategy criterion, then 2 predicts s₁₂ as 1's decision, and based on this prediction, he chooses s₂₂. This is written as

$$(5b) \vdash B_2(g_2^{0'} \cup B_1(g_1^{0'})) \rightarrow B_2(B_1(\text{Dom}_1(s_{12})) \wedge \text{Best}_2(s_{22} | s_{12})),$$

where I did not write 1's statement since it becomes too long.

Jan: OK, $B_1(\text{Dom}_1(s_{12}))$ in the scope of B_2 is the prediction of player 2 about 1's decision making. The depth of nested beliefs is 2 here, isn't it? Then, $\text{Best}_2(s_{22} | s_{12})$ means s₂₂ is the best strategy to the prediction s₁₂. It all makes sense. I think, I can now find more examples which require finite depths of nested belief structures. This appears to capture some aspect of bounded rationality, doesn't it?

Shinzuki: Yes, it is one aspect of bounded rationality. You will find more examples in [7].

Jan: Now I recall that I am already very tired. I want to ask one more small question before coffee, since I feel we have reached some deep implications for game theory. My last question is about the multiplicity of... uhum..., what?

Shinzuki: You are really tired. I think you want to say the multiplicity of candidates for decisions.

Jan: That's right, in fact your wording is better than what I was thinking. Suppose we have a game like the battle of the sexes with two Nash equilibria:

	s_{21}	s_{22}
s_{11}	$(2, 1)^*$	$(0, 0)$
s_{12}	$(0, 0)$	$(1, 2)^*$

Figure 4: $g^4 = (g_1^4, g_2^4)$

Even if each has a belief of common knowledge of a Nash equilibrium, each may have a different Nash equilibrium in mind. Then it might seem that each can make a decision to use his equilibrium strategy. But the combination is not a Nash equilibrium. How do you treat such a situation?

Shinzuki: You need an additional assumption such as the common knowledge of one Nash equilibrium, e.g., (s_{11}, s_{21}) . This may be included in Γ_i of (4). If both players have the same common knowledge assumption, then the resulting outcome is (s_{11}, s_{21}) .⁷ However, if only player 1 has this, the outcome (s_{11}, s_{21}) may not be reached, yet player 1 chooses s_{11} and predicts s_{12} to be chosen by player 2. To obtain such common beliefs of one strategy combination needs communication, and such communication is external to epistemic logic.

Jan: I should say, "Let's go for coffee now" to avoid more divergence. Do you know any good places?.

Shinzuki: That is a very difficult question. Anyway, we can find coffee in the village.

4. Act: General Principles

[Shinzuki and Jan are coming back from a coffee shop]

Jan: It was good coffee, but the coffee shop was too crowded. I have a good idea about how to control the number of customers. Can you imagine it?

Shinzuki: Yes, I can, of course. It is that the price should be increased according to the economic principle that when price rises, demand will fall.

Jan: That is one way. Another way is to decrease the quality of coffee. Then some customers would leave.

Shinzuki: Ha ha, who takes such a policy?

Jan: Yes, yes, a coffee shop in my town took this really. It succeeded in decreasing the number of customers. I myself went there only once after the reduction in quality. Can you believe this control?

Shinzuki: It is difficult to believe. Now I should impose a constraint to you not to digress from our main problem.

Jan: Sorry about it. In fact, this is my revenge to you.

Shinzuki: Indeed, it is a nice help to understand the transgression of digression. I will think about a revenge to you, too.

⁷See Kaneko [4] for the treatment of games with uninterchangeable Nash equilibria.

Jan: Now, let me try to recall what should be discussed. Uhum..., you promised to discuss the distinction between knowledge and belief.

Shinzuki: Good, I remember our problem, too. In the European tradition of philosophy it is standard to define

(6): knowledge is justifiable true belief.

In this definition, belief is more basic than knowledge. For the moment please take the “belief” as something held in the mind of the believer. I want to talk first about the word “true”. This notion of truth refers to an objective observer, or at least some thinker other than the one holding the belief.

Jan: OK, you are saying that the truth in (6) is external to the believer. There might also be a truth included within the belief of a person. But this truth is beyond dispute for him and we cannot discuss the truth or falsity of such beliefs within his perspective. We can only objectively distinguish between false and true beliefs if we take the word “true” written in (6) to be an external one. A true belief in this sense is a candidate for knowledge, isn’t it? I think this is compatible with what we discussed before.

Shinzuki: Very good. In my previous statements, such as (2) and (3), there are at least three references to truth. One is from the perspective of the outside objective observer, and then there is one from the perspective of each player. In (2), g^0 is the situation true for the outside observer. From this perspective, the content, g^i , of $B_i C(g^i)$ is false, but, from the perspective of each player i his belief $B_i C(g^i)$ is true and beyond dispute. We discussed the falsities involved in (3) already.

Jan: It might help my understanding to consider the case where all beliefs are true even from the perspective of the outside observer. Consider the following:

(7): $\vdash g^0, B_1 C(g^0), B_2 C(g^0) \rightarrow ??$

where I put ?? in the right-hand side since the conclusion part is irrelevant to the point I want to make now. According to your explanation, the assumption part of (7) is indistinguishable from the common knowledge of g^0 . Is there any meaningful difference between this and simply writing $C(g^0)$?

Shinzuki: Actually, it can be proved that the set $g^0, B_1 C(g^0), B_2 C(g^0)$ is equivalent to $C(g^0)$, where the notation $g^0, B_1 C(g^0), B_2 C(g^0)$ is an abbreviation of $g^0 \cup B_1 C(g^0) \cup B_2 C(g^0)$. Moreover, this is equivalent to $CC(g^0)$, and to $CC\dots C(g^0)$ with any finite number of applications of C .⁸

Jan: Really? It is interesting. In this case, everything is OK with me and it also seems compatible with the standard “implicit” game theoretical arguments.

⁸There are two approaches to extending epistemic logic with common knowledge: fixed-point approach and infinitary approach. For the former, see Fagin et al [2], Meyer-van der Hoek [10] and [9]. For the latter as well as the relationship to the former, see Kaneko et al [8].

With this new understanding, I would like to consider more cases of false beliefs. Let's return to the assumption part, $g^0, B_1C(g^1), B_2C(g^2)$, of (2). Because there are false beliefs involved, it seems possible that one of the players might notice this inconsistency and the assumption set would fall apart. How do you avoid it, or guarantee that the assumption part is all compatible and might persist?

Shinzuki: I am glad you are now getting through the gate of epistemic logic.

Jan: You are flattering me, but I am still pleased at my finding, and perhaps at your noticing my finding. Anyway, please continue.

Shinzuki: OK. The assumption set g^0, g^1, g^2 is inconsistent, and so is the set $g^0, C(g^1), C(g^2)$.⁹ If we start with either of these assumption sets, the theory is nonsense since anything, including absurdities, can be derived from a set of inconsistent assumptions. To prevent such absurdities, and allow the persistence of a set of assumptions, it suffices to prove that the set is consistent. If consistency is shown, then no player will ever notice the inconsistency and all will be fine. This is really the starting point of the research in the fields of "epistemic logic and game theory". We use some developments from proof theory and model theory to guarantee the consistency of $g^0, B_1C(g^1), B_2C(g^2)$ which can be found in [5].

One notable finding is that to obtain the consistency of this set, we need to drop one famous axiom from our epistemic logic:

$$(8): B_i(A) \rightarrow A.$$

This is often called the *truth axiom* or *veridicality axiom*. If (8) is assumed, every belief is assumed to be true in the eyes of the outside observer, and false beliefs cannot be discussed in the epistemic logic.

Jan: Another point which I want to mention now, but perhaps leave for a later discussion, is about the meaning or interpretation of the consistency of an assumption set, and the viewpoint from which it is derived.

I can now think of four meaningful different perspectives and assumption sets. The assumption set for player 1 is clearly $B_1C(g^1)$, and for player 2 it is $B_2C(g^2)$. The last two perspectives involve the outside observer who may have either g^0 alone or the entire set $g^0, B_1C(g^1), B_2C(g^2)$. While we can imagine an outside observer who knows only the true game g^0 , we can also imagine an objective observer who knows all of $g^0, B_1C(g^1), B_2C(g^2)$. This second type of objective observer is rather close to the position I am taking in my analysis, and the one you seem to be taking in your analysis.

Shinzuki: Your last distinction is very nice. The consistency of $g^0, B_1C(g^1), B_2C(g^2)$ is proved from the viewpoint of the second objective observer you mentioned, even

⁹Throughout this paper, we assume that $C(A)$ means the common knowledge of A rather than the common belief of A . In our context, the common belief of A is definable as $B_1C(A) \wedge B_1C(A)$. Conversely, if we start with the common belief, $C_B(A)$, of A as a primitive, then the common knowledge of A is definable as $C_B(A) \wedge A$. See Kaneko et al [8] as well as Section 6 of Kaneko [5].

though I may have seemed to take the perspective of the first observer you mentioned since I stressed the comparisons of g^0 with g^1 and g^2 .

[Jan is pleased]

Jan: I am slightly curious about how to prove the consistency of $g^0, B_1C(g^1), B_2C(g^2)$, and perhaps this will help me understand what I just tried to describe. Can you give a brief explanation of the method of proof?

Shinzuki: Yes, I can. A proof needs a lot of details. However, believing its consistency is much like believing the logical possibility of the Konnyaku Mondô. The Konnyaku Mondô involves a set of false beliefs that is just like the assumption set $g^0, B_1C(g^1), B_2C(g^2)$. By admitting the logical possibility of the Konnyaku Mondô, you also admit a belief in the possibility of its consistency. When you heard the story, you most likely searched for some way to make sense of it all. This is what a proof of consistency does. The situation arrived at for making sense of it all is called a *model*. If we can construct a model of $g^0, B_1C(g^1), B_2C(g^2)$, then we can use the *soundness theorem* to obtain the consistency of $g^0, B_1C(g^1), B_2C(g^2)$.

Jan: Stop, stop, I understand at least a lot are needed for that step. Please postpone your explanation of the construction of a model and the soundness theorem to sometime in the future. I am learning a lot now from your description and by thinking about the story in this new light. Can we change slightly our direction and discuss the word “justifiable” in (6).

[Shinzuki shows his disappointment by raising his hands to his head]

Shinzuki: Arghhh..., I am slightly disappointed, since that is my favorite part. Perhaps later you will appreciate it more. “Justifiable” is not very exciting, technically speaking. Usually, “justifiable” is quite ambiguous, but it is clear-cut in the epistemic logic approach since it unearths everything. By a justifiable belief, say A , for player i , we mean that player i has some argument, or justification for A from his basic beliefs. The basic beliefs are written in Γ_i of (4). For example, if player i has a mathematical proof for A from Γ_i , then this A is a justifiable belief.

Jan: It is clear-cut. A belief is justified when, for example, a player has a proof of it, uhum..., this proof is taken from his basic beliefs, right? So mightn't we ask for a justification for a basic belief?

Shinzuki: No, basic beliefs are not justified in this sense.

Jan: But you claimed that the epistemic logic unearths everything.

Shinzuki: Sorry. I should say it unearths everything that can be unearthed.

[A bit surprised by Shinzuki's statement, Jan retorts]

Jan: Ha ha, that sounds like a tautology. Is it everything or nothing? It seems to me, it refuses to dig.

Shinzuki: You are right. I retreat by saying epistemic logic unearths everything before the consideration of basic beliefs Γ_i . We need to look for some other sources for basic beliefs.

Jan: What other sources for basic beliefs do you have in mind?

Shinzuki: This part is not well developed, though it has been discussed since the age of Plato. One source is one's experiences. This process is often called "induction", i.e., to obtain a general law from finite experiences. Please note that this "induction" differs from the mathematical induction principle. While the latter principle has the same name, it is essentially a principle of deduction. In fact, the latter may be included in Γ_i ,¹⁰ but the former is external to the logical framework and is the process of constructing and revising the basic beliefs in Γ_i .¹¹

Jan: I am surprised to hear that the mathematical induction principle is one for deduction. The other "induction" you speak of is not deductive at all. When I learned the distinction between induction and deduction at university, both were described as processes of scientific reasoning.

Shinzuki: One difference here from what you learned is that our target is a human inductive process rather than a scientist's inductive one. The latter requires careful statistical treatments, but the former is rather a bold process since a great generalization is often made from a few experiences. We all have this tendency to generalize. For example, you went to the coffee shop only once after it decreased the coffee quality, but the shop might have increased the quality later. Are you sure that the coffee quality there is still bad?

Jan: OK, I understand both your distinction between human and scientific inductive processes. Also, the last example seems to be your revenge to me.

Shinzuki: I admit that both your statements are correct. I would like to give one more example in this line of reasoning to deepen your understanding. Suppose that you meet two people of some ethnic group new to you. You observe differences in their appearances as well as their behavior relative to the standard in your community. You might conclude that all people of that ethnicity have such appearances and behavior.

Jan: I understand that these generalizations are rampant in our lives, but on the positive side, they allow us to expand and revise our basic belief set Γ_i .

Shinzuki: Your point is taken, but I would rather like to point out that once again we have arrived at some false beliefs. The generalities involve false beliefs in the sense that they are not valid from an objective viewpoint. These false beliefs are related to sociological problems of discrimination and prejudice. The inductive method I have been speaking of is formulated in [6], which the authors call the *inductive game theory*. In the paper, the emergence of racial prejudices is discussed, but connections to the epistemic logic are not yet fully reached.

Jan: It sounds interesting. I will look at the paper. However, I doubt I can attribute all my beliefs to my past experiences. My experiences are very limited,

¹⁰To include the mathematical induction principle in Γ_i , we need predicate epistemic logics. For such predicate epistemic logics including common knowledge, see Kaneko et al [8].

¹¹The revision of beliefs has been discussed in the literature on belief revision. See Schulte [13] for a good survey.

yet I seem to have a limitless fountain of beliefs, even some on things I never experienced, like what it would be like to fly.

Shinzuki: Yes, experiences are limited for an individual being, but we also have communication and education. The experiences of former generations have been communicated to later generations through discussions, books, and other sources. A lot of "trial and error" has been taken across the many generations of man and many of these experiences have been passed on. For example, our distinction between poisonous and edible mushrooms is possible only after many trials and errors by previous generations.

Jan: The accumulation of such experiences are taught to us. Great! We have been making progress, and will continue to progress.

Shinzuki: You are right to the extent that what we have discussed is a quite standard view of progression. However, I like to point out that communication and education are powerful, but have their own limitations. For communication and learning as well as for one's own memory of experiences, language is a useful and necessary tool. Language consists of primitive symbols and grammars, just as a logical system consists of symbolic expressions and inference rules. In this sense a logical system can be regarded as the idealization of a language.

[Having tuned out slightly, Jan notices Shinzuki is staring at him waiting for a reply]

Jan: What is your point?

Shinzuki: My point is as follows. Suppose somebody has experiences. However, he translates his experiences into language in order to communicate those experiences to others or even just to memorize them for himself. In this translation, something is lost and something is gained. Something is lost by the fact that the communicator must choose some part of his experience in order to be understood by others. Something is gained by adding the structure required to express the experience in a language understandable by others. Raw experiences can never be transferred to other people.

Jan: It would be nice if you could relate it directly to what we have been discussing.

Shinzuki: OK, I continue. People having experiences construct a simple story explaining experiences, and other people are taught this story through communication or education. This is a source of the basic beliefs Γ_i of (4).

[Shinzuki's voice becomes louder]

Now, I stress the story is coherent, but must also be simple so it can be memorized or communicated. This implies that Γ_i is innately subject to some falsities. If the story is extremely accurate and has every element of experiences, it is useless since it is not understandable. Also, everybody has a conscious or unconscious tendency to adjust the story so it is more comfortable for him or her. Such adjustments were described in a funny, but sad, manner in Akira Kurosawa's

movie "Rashoumon".¹²

[Jan feels slightly taken aback, and Shinzuki is roaring with enthusiasm]

Jan: I understand what you want to say. My view of progress may be too naive. Does your argument have any relationship to the Konnyaku Mondô?

Shinzuki: Yes, it does! You are finally becoming very sensitive to the contradictory elements in our discussions. In the Konnyaku Mondô, the exchange of gestures is a communication. It represents the symbolic nature of language. Each person constructs a story from the gestures, which includes falsities. This example is an extreme one, but it shows that communication is never entirely free from fallacious elements.

Jan: Yes, indeed, but still a lot can be communicated by discussions like the ones we have been having.

Shinzuki: I am afraid we might be having nothing more than a Konnyaku Mondô. Just like you, I believe we have had meaningful discussions, but in actuality, the whole time I have been talking about Konnyaku products in epistemic logic terms..., and you, my friend, have been discussing the battle of the sexes in game theoretical terms.

Jan: You are too cynical! I think we should stop for today. Why don't we go to the public bath tonight? The public bath increases the quality of our life here very much. I hope the town never decides to decrease the quality of the water to control crowding problems.

Shinzuki: Ha, ha, it is a great idea to meet at the public bath. It helps you understand the Japanese culture, doesn't it? Shall we meet there around 8pm?

References

- [1] Chellas, B., (1980), *Modal Logic*, Cambridge University Press, Cambridge.
- [2] Fagin, R., J.Y. Halpern, Y. Moses and M. Y. Verdi, (1995), *Reasoning about Knowledge*, The MIT Press, Cambridge.
- [3] Hughes, G. E. and Cresswell, M.J., *A Companion to modal logic*. London: Methuen & Co. (1984).
- [4] Kaneko, M., (1999), Epistemic considerations of decision making in games. *Mathematical Social Sciences* 38 (1999), 105-137.
- [5] Kaneko, M., (2002), Epistemic Logics and Game Theory: Introduction, *Economic Theory* 9, 7-62.
- [6] Kaneko, M., and A. Matsui, (1999), Inductive Game Theory: Discrimination and Prejudices, *Journal of Public Economic Theory* 1, 101-137.

¹²The original short story for the movie "Rashoumon" is called "Yabunonaka" by Ryunosuke Akutagawa. He wrote a different story with the title "Rashoumon".

- [7] Kaneko, M., and N-Y. Suzuki, (2002), Bounded Interpersonal Inferences and Decision Making, *Economic Theory* 9, 63-104.
- [8] Kaneko, M., T. Nagashima, N.-Y. Suzuki and Y. Tanaka, (2001), A Map of Common Knowledge Logics. To appear in *Sudia Logica*.
- [9] Lismont, L. and Mongin, P.,(1994), On the logic of common belief and common Knowledge. *Theory and Decision* 37, 75-106.
- [10] Meyer, J.-J. Ch. and van der Hoek, W., *Epistemic logic for AI and computer science*. Cambridge: Cambridge University Press (1995).
- [11] Nash, J. F., (1951), Noncooperative Games, *Annals of Mathematics* 54, 286-295.
- [12] Savage, L., (1954), *Foundations for Statistics*, Dover Publication Inc. New York.
- [13] Schulte, O., (2002), Minimal Theory Change and the Pareto Principle, *Economic Theory* 9, 105-144.
- [14] Terutoshi, Y. Okitsu K. and Enomoto, S., (eds.), *Collections of meiji-taisho comic stories* (in Japanese). Vol. 3. Tokyo: Kodansha (1980)

