

# Imperfect Information in Multi-agent Epistemic Logic

Extended abstract

March 5, 1998

## 1 Introduction

Although first-order multi-agent epistemic logic is an expressive and powerful formalism for representing many types of knowledge, there are simple and natural situations that cannot be captured in it. A case in point is mutually imperfect *de dicto/de re* knowledge between two agents. If one agent knows *de dicto* something that the other agent knows *de re* and conversely, the situation cannot be expressed in first-order epistemic logic.

This kind of knowledge can be, however, represented in first-order epistemic logic where imperfect information is allowed. As a basic framework for representing imperfect information we shall, therefore, employ a first-order logic of imperfect information similar to the one proposed by Hintikka [3], Hintikka & Sandu [4] and Sandu [5]. In such a logic, it is possible to control which semantic information passes down to subformulas and which does not. Essentially, a mechanism is exploited which relieves some subformulas of the syntactic scope load of the quantifiers further out in the formula.

On several occasions ([2],[3, p. 83–87],[4, p. 376–378],[5, 47–49]) Hintikka and Sandu suggest that the same mechanism is also useful in the case of first-order epistemic logic. According to Hintikka [2, p. 261], it in fact yields to the logic where “all types of knowledge expressible in natural languages can be expressed by knowledge statements, and their logical relationships to other sentences are the correct ones”.

Our task is to extend the idea behind knowledge statements to epistemic logic with many

agents. Consequently, a mutually imperfect knowledge between agents is expressible in such a logic. However, the original semantics for epistemic logic of imperfect information, which was given by means of semantic games poses additional difficulties. We shall put forward two further questions. First, what semantic game can be played on a formula where quantifiers and epistemic operators enjoy cyclic dependencies? Second, if the game-theoretical apparatus for epistemic languages of imperfect information is employed, can the existence of winning strategies really be taken to mean that the formula in question is true or false?

## 2 Epistemic logic of imperfect information

In some cases it is instructive to think of epistemic operators as universal quantifiers ranging over the ‘multiverse’ of possible worlds. Hence an analogous treatment with that of Henkin’s [1] partially ordered quantification or Hintikka’s “independence-friendliness” [2],[3],[4],[5] results in interesting extension of traditional first-order epistemic logic. Hintikka’s suggestion is to use an outscoping device “/” to exempt quantifiers from the scope of the epistemic operators. For example, the intended meaning of the formula

$$K_1(\exists x/K_1)S(x) \quad (1)$$

is that the evaluation for the variable “ $x$ ” is to be made independently of the evaluation for the epis-

temic operator  $K_1$ . As with the case of independence of existential quantifier from the universal quantifier, (1) is seen to reduce to the expression

$$(\exists x)K_1 S(x) \quad (2)$$

The formula (2) represents knowledge *de re*, often taken to mean ‘knowing wh-’ (who, what, which, where etc.). Hence (2) can be read as “an agent 1 knows who (say  $x$ ) is such that  $S(x)$ ”. In contrast, the formula

$$K_1(\exists x) S(x) \quad (3)$$

is knowledge *de dicto*, and it can be read as “an agent 1 knows that there exists  $x$  such that  $S(x)$ ”.

However, in epistemic logic of imperfect information there are expressions that do not reduce to any first-order form without the “/”. The simplest example is

$$K_1(\forall x)(\exists y/K_1) S(x, y) \quad (4)$$

Here  $\exists y$  needs to be in the scope of  $\forall x$  but not in the scope of  $K_1$ . Furthermore,  $\forall x$  may not quantify into the scope of  $K_1$ . Hence no linear ordering is possible. Hintikka intends the formula (4) to mean statements like “an agent 1 knows whom each person admires most” [2, p. 263],[4, p. 377].

A well-known device for giving semantics to the logics of imperfect information is Hintikka’s game-theoretical semantics (GTS). In GTS there are two players, call them  $\forall$ belard (falsifier) and  $\exists$ loise (verifier), who make moves according to the logical ingredients encountered in the formula.  $\forall$ belard is trying to falsify the sentence and  $\exists$ loise is trying to verify it. The logical ingredients  $\forall$  and  $\wedge$  prompt a move by  $\forall$ belard, and  $\exists$  and  $\vee$  prompt a move by  $\exists$ loise. When the players come across negation, they change rôles. Each move reduces the complexity of a formula and hence, when the atomic formula is finally reached, its truth-value (established by a given interpretation) determines which player wins the game. If the atomic formula is true,  $\exists$ loise wins, if it is false,  $\forall$ belard wins. The original formula is true iff there exists a winning strategy for the initial verifier, and false iff there exists a winning strategy for the initial falsifier.

Hintikka has proposed a straightforward extension of GTS to give an account of what epistemic formulas like (1)–(4) mean. In addition to the basic rules of GTS for first-order logic, if an epistemic operator  $K_i$  is encountered,  $\forall$ belard chooses a possible world, and the game continues with respect to this chosen world. If a dual of knowledge,  $M_i$  (regarded as possibility) is encountered,  $\exists$ loise makes a choice of a possible world.

In contrast to the games for traditional first-order logic or first-order epistemic logic where games are of perfect information, games associated with the formulas (1) and (4) are of imperfect information: players may occasionally fail to know which possible world was chosen by the opponent. However, the plausibility of a semantics like this should be investigated more carefully, for players may fail to know in which world they are to operate and thus lose track of their position in the game. In order to follow the game, they need information about what course of events was chosen previously; consequently, hiding the course of events is a rather radical thing to do. We shall return to this semantic problem after giving an argument for what is achieved by incorporating imperfect information in epistemic logic.

### 3 Multi-agent case

The formula (4) already provides an example of an expression that does not reduce to any first-order expression. In order to write such irreducible formulas, either universal quantifier and dependency between existential and universal quantifier is needed, or there must be more than one agent.

Let there, thus, be at least two agents, named  $1, 2, \dots, m$ . Now if imperfect information is allowed, something peculiar happens, and the logic becomes very expressive indeed. Let us consider a statement for two agents:

Agents 1 and 2 know that there exist  $x$  and  $y$  respectively, and 1 knows who (the individual denoted by)  $x$  is whereas 2 knows who (the individual denoted by)  $y$  is.

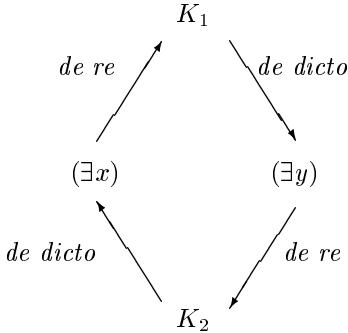
This statement is not expressible in first-order epistemic logic! Since in ‘knowing that’ an existential quantifier should come after a knowledge operator, but in ‘knowing who’ it should precede it, no first-order linear ordering is possible. The statement is, however, expressible with the aid of outscoping:

$$K_1(\exists x)K_2(\exists y/K_1)[\heartsuit_1(x) \wedge \heartsuit_2(y)]. \quad (5)$$

Since  $\exists y$  should now precede  $K_1$ , we also have *de re* knowledge for the agent 1. In colloquial language, (5) could express the mental state of both agents in a way not unlike the following ( $\heartsuit_i(x) = “x$  is a love of  $i$ ’s life”):

*I know that there is a love of my life  
and I know who (s)he is for you.*

The formula (5) is not unproblematic, however. It is as if  $\exists y$  should exist in two places simultaneously. It is due to the subtle bidirectional dependence between  $\exists y$  and  $K_i$ . Rewriting the prefix of (5) in a graph-like dependency diagram reflects the situation even better:



Note the graph prefix is intransitive, and hence  $\exists y$  is not allowed to quantify into the scope of  $K_2$ , nor is  $\exists x$  allowed to quantify into the scope of  $K_1$ . This means that the outscoping device in (5) is not really an adequate notation. We could remedy the situation by introducing also an in-scoping device “//” (not to be confused with the double-slash notation discussed in [3, p. 53–54] as an alternative syntax for the first-order logic of imperfect information) and using *both* devices in the same sentences. Here the intended meaning of

an expression  $A//B$  is that  $A$  quantifies only into the scope of  $B$ . If  $B = \emptyset$ ,  $A$  functions as a normal quantifier. Hence, the following formula gives a correct one-liner substitute for the graph:

$$K_1(\exists x)K_2(\exists y/K_1//K_1)[\heartsuit_1(x) \wedge \heartsuit_2(y)]. \quad (6)$$

It is unclear, however, what game in the sense of GTS could be played on these kind of cyclic quantifier prefixes, since there is no natural starting point where the game could begin. This situation seems to be very common when the methods of information hiding are added into multi-agent logic.

A general observation here is that whenever the dependency graph contains cycles the formula is not any more expressible in first-order epistemic logic. But then it is not, without resorting to the additional notation of in-scoping, expressible in “independence-friendly” epistemic logic of Hintikka either, nor it is playable by normal rules of GTS for epistemic logic.

## 4 Winning strategies

We might, therefore, need a new definition of game for the multi-agent epistemic logic of imperfect information, but we might also need a new account of what it means for there to exist a winning strategy for one of the players. In game-theoretical semantics, the existence of a winning strategy for Eloise means that the sentence is true. However, consider the formula (5) and imagine that the game could somehow start with the outermost ingredient  $K_1$ .  $\forall$ belard chooses a possible world, say  $w_i$ , where the variable  $x$  is to be evaluated. We can label the world as a superscript to the variable and hence the game continues with the formula  $(\exists x)K_2(\exists y/K_1)[\heartsuit_1(x^{w_i}) \wedge \heartsuit_2(y)]$ . The next choice is for Eloise, and so she chooses an individual in the world  $w_i$ , call it “ $a$ ”. The game continues for  $\forall$ belard, and he chooses now some possible world where  $y$  is to be evaluated, say  $w_j$ . The game has reached the position  $(\exists y/K_1)[\heartsuit_1(a^{w_i}) \wedge \heartsuit_2(y^{w_j})]$ , but this time Eloise has to make a selection *without knowing* the world  $\forall$ belard chose for  $x$ . An additional re-

quirement is also that there is no signalling: when, for example making a move for  $\exists y$ , Eloise may not use the information she has obtained during her previous move prompted by  $\exists x$ .<sup>1</sup> Now the game has reached the formula  $\heartsuit_1(a^{w_i}) \wedge \heartsuit_2(b^{w_j})$ , an agent 1 knowing that somebody exists in the world  $w_i$  and an agent 2 knowing that somebody exists in the world  $w_j$ . Furthermore, 1 knows the identity of an individual in  $w_i$  and 2 knows the identity of an individual in  $w_j$ . Certainly it might also happen that the two individuals are one and the same (which is unfortunate, if the reading is like above). The last selection is for  $\forall$ belard to choose one of the conjuncts, say  $\heartsuit_1(a^{w_i})$ . Game semantics tells us that if  $\heartsuit_1(a^{w_i})$  is true then Eloise wins. But when is the whole sentence true?

In the language of possible worlds semantics, knowledge is truth in all possible worlds. This intuition, however, obscures the rôle of winning strategies. Imperfect information clearly restricts the allowable set of winning strategies for Eloise. Similarly with the previous example, taking a tiny two-element domain  $\{a, b\}$  and only two possible worlds  $\{w_i, w_j\}$  there are now 16 winning strategies for Eloise, instead of 64 in the perfect information case. In the last choice for Eloise, she should have a strategy which can produce an individual (“ $a$ ” or “ $b$ ”) no matter what world  $\forall$ belard produced earlier (the last choice of conjunct does not matter for winning strategies for Eloise). But what does the existence of a winning strategies show? Does it show that the sentence in question is true? Or does it merely indicate that Eloise knows the formula? If the alleged truth-making choice for Eloise is operational without knowledge of the worlds, how can we establish that the sentence is true?

---

<sup>1</sup>In game theory this kind of forgetting is known as *imperfect recall*, and a related class of games is known as games of imperfect recall. Games of imperfect recall seem to be, in fact, a more accurate class of games for the languages of information hiding described here than the class of imperfect information.

## 5 Conclusion

We have provided a case study of epistemic logic of imperfect information in the case of two agents. It seems to result in an intriguing language where mutually imperfect *de dicto/de re* knowledge can be captured. However, a semantic treatment by means of games poses two problems. First, how can we play a semantic game on a formula where there exists cyclic dependencies? Second, what does the existence of winning strategies really mean?

## References

- [1] HENKIN, L. (1961) *Some remarks on infinitely long formulas*. **Infinistic Methods**. Pergamon Press, NY, pp. 167–183.
- [2] HINTIKKA, J. (1993) *Independence-friendly Logic as a Medium of Knowledge Representation and Reasoning about Knowledge*. In OHSUGA, S. et al. (Eds.) **Information Modelling and Knowledge Bases III: Foundations, Theory and Applications**. IOS Press, Amsterdam, pp. 258–265.
- [3] HINTIKKA, J. (1996) **Principles of Mathematics Revisited**. Cambridge University Press, NY.
- [4] HINTIKKA, J. and SANDU, G. (1997) *Game-theoretical Semantics*. In VAN BENTHEM, J. and TER MEULEN, A. (Eds.) **Handbook of Logic and Language**. Elsevier, Amsterdam, chapter 6, pp. 361–410.
- [5] SANDU, G. (1993) *On the logic of informational independence and its applications*. **Journal of Philosophical Logic**, Vol. 22, pp. 29–60.