Department of Computer Science,
University of Otago

Technical Report OUCS-2001-10

Reflections on 'The Importance of Being Formal'

Author:

Hans van Ditmarsch

Status: submitted to *Mathematical Intelligencer*

Department of Computer Science,
University of Otago, PO Box 56, Dunedin, Otago, New Zealand

http://www.cs.otago.ac.nz/trseries/

Reflections on "The Importance of Being Formal"

Hans van Ditmarsch*[†]

In (Makarychev and Makarychev, 2001) the following problem was discussed:

> *From a pack of seven known (and all different) cards, two players each draw three cards and the third player gets the remaining card. How can the players with three cards openly inform each other about all their cards, without the third player learning from any of their cards who holds it?*

The authors solve the problem, analyze what they call a 'bad solution', and give a procedural requirement for ('good') solutions. From a background including formal specification of information states involving card deals (van Ditmarsch, 2000) I had encountered the same cards problem (van Ditmarsch, 2001). I propose a perspective that is more in accordance with epistemic logic (Fagin et al., 1995). Central to this are the notions of 'update' and 'common knowledge'. These notions have precise interpretations in relational structures representing information states. I also give other solutions to the problem, and some generalizations.

Requirements for a solution of the cards problem

The players are called A, B and C (referred to as 'she', 'he', and 'it', respectively), the cards are named $0, 1, 2, 3, 4, 5, 6$. A postcondition for a solution of the cards problem is that A knows B's cards (a), B knows A's cards (b), and C doesn't know any of A's or B's cards (c). Assume w.l.o.g. that A holds $\{0, 1, 2\}$, B holds $\{3, 4, 5\}$ and C holds 6. For that card deal, the 'bad solution' in (Makarychev and Makarychev, 2001) is:

> *A says "If you don't have 0, then I have $\{0, 1, 2\}$" and B says "If you don't have 3, then I have $\{3, 4, 5\}$".* (i)

Now imagine that there is a fourth person D present who can look into every player's cards, an 'insider' so to speak, and that A's and B's announcements had actually both been made by that insider:

> *D says "If B doesn't have 0, then A has $\{0, 1, 2\}$" and D says "If A doesn't have 3, then B has $\{3, 4, 5\}$".* (ii)

*Computer Science, University of Otago, New Zealand, hans@cs.otago.ac.nz

[†] I thank Alexander Shen, and Konstantin and Yuri Makarychev, for helpful and interesting discussions, and for their translations.

We then reach an information state where a, b, and c all hold. This suggests a solution of the problem. However, not this insider but the players themselves made these announcements, and this is informative. As the Makarychevs correctly state, player A can only truthfully make her announcement if she actually holds cards $\{0, 1, 2\}$. And C knows that! In other words, we may assume i to have been:

> *A says "I know that if you don't have* 0*, then I have* $\{0, 1, 2\}$*" and B says "I know that if you don't have* 3*, then I have* $\{3, 4, 5\}$*".* (iii)

By 'an agent knows something' we mean 'an agent knows something to be true'. In fact, after iii the players have common knowledge of the deal of cards! Now iii is a more likely interpretation of i than ii is, but to make their (valid) point the Makarychevs need ii. That confuses issues somewhat. The following example doesn't have this confusion, and from here on we diverge from their approach:

> *A says "I have* $\{0, 1, 2\}$ *or I don't have any of these cards." and B says "I have* $\{3, 4, 5\}$ *or I don't have any of these cards.".* (iv)

Even when we assume that players only say what they know (to be true), we reach an information state where c holds. But although c holds, *A doesn't know that c holds* (A doesn't know that C doesn't know any of A's and B's cards). Also, even though a holds, *C doesn't know that a holds* (C doesn't know that A knows B's cards): if in the resulting state A says "I know B's cards" this will be *that* informative, that once more the card deal becomes common knowledge.

Why is iv also not a solution? Once more, the message is more informative than it appears to be: not only do players only say what they know, but players A and B will also avoid saying something that may result in C learning some of their cards. In other words: we may use that the announcements are part of the execution of a protocol to solve the cards problem. A truly interesting scenario unfolds:

> *After iv player C thinks: "Suppose that A does not have card 0. Then A doesn't know whether I have 0. If I had 0, I would have learnt from A's announcement that A doesn't have the cards 1 and 2. Therefore A will not make the announcement. But A just made the announcement! Therefore A has card 0. And, incidentally, therefore also cards 1 and 2."*

In other words: C learns A's cards from the assumption that A only makes announcements from which C does not learn A's cards. Worse than that: C can *only* learn A's cards from that assumption.

There is no real contradiction involved. A similar scenario occurs in the puzzle known as the 'wisemen', 'hats' or 'muddy children' problem: from the announcement that no child knows whether it is muddy, some children may learn that they are muddy. See (Fagin et al., 1995). As in i, we can be more explicit than iv:

> *A says "I know that I have $\{0,1,2\}$ or I don't have any of these cards, and that after saying this C does not know any of my cards." and B says "I know that I have $\{3,4,5\}$ or I don't have any of these cards, and that after saying this C does not know any of my cards."* (v)

The first 'this' in v only refers to "I know that I have $\{0,1,2\}$ or I don't have any of these cards." and not to "I know that I have $\{0,1,2\}$ or I don't have any of these cards, and that after saying this C does not know any of my cards." Similarly for the second 'this', in B's part of v.

Where does this end? We only have to go a little bit further along this road. If we execute a protocol to solve the cards problem and a player, say player A, says "φ", we may assume that she actually means "I know that φ and that after saying φ it is common knowledge that C doesn't know any of my of B's cards." It turns out to be the case, and we have just seen that this is *not* trivial, that after that entire statement it *remains* common knowledge that c holds ('C doesn't know any of my of B's cards'). A sequence of such announcements is a solution of the cards problem, if a and b are common knowledge after that. Something is commonly known, if and only if it is true and everybody knows that it is commonly known (it is the least fixed point of this operation). So common knowledge of some proposition ψ implies that everybody knows that ψ, i.e., in the cards problem, A knows that ψ and B knows that ψ and C knows that ψ, but it also implies that everybody knows that everybody knows that ψ, etc.

One more thing: in information states for card games something is common knowledge if it holds in 'all deals that are still consistent with the information given so far'. Checking that, is the general method to validate concrete solutions.

Epistemic logic for multiagent systems

The notions of state, knowledge, common knowledge, and announcement, can be represented by and interpreted on relational structures. We illustrate that by means of a simple example.

There are three cards, red (r), white (w), and blue (b), and three players, 1, 2 and 3. Each player gets one card. There are six such deals. Write rwb for the deal where player 1 holds red, 2 holds white, and 3 holds blue, etc., and suppose rwb is the actual deal of cards. Two deals are the same for a player if he holds the same card in both deals, e.g. $rwb \equiv_1 rbw$. This equivalence relation induces a partition on the set of six deals. E.g., the partition for player 1 is $\{\{rwb, rbw\}, \{wbr, wrb\}, \{bwr, brw\}\}$.

Facts about card ownership and knowledge of players about the cards and about each other, are represented by a relational structure that one may call a *modal state* (a.k.a. a pointed Kripke model or a pointed possible worlds model). Given a set N of agents and a set P of atoms, a modal state $(\langle S, R, V\rangle, s)$ – among logicians, this structured notation is preferred over a quadruple – consists of a domain S of abstract objects called *factual states* (or 'states of the world', or ‘worlds’, or simply ‘states’), a function $R : N \to \mathcal{P}(S \times S)$ that assigns a binary relation $R(n)$ to each agent n, a function $V : P \to \mathcal{P}(S)$ that assigns a unary relation (or valuation) $V(p)$ to every fact p, and one special object $s \in S$ called the *point*. If all the relations are equivalences $\equiv_n$, and from now on they always are, we call the structure an *information state*.

In the example, the domain is the set $\{rwb, rbw, wbr, wrb, bwr, brw\}$ of six deals, the point is the actual deal rwb, the binary relations are the three equivalence relations as distinguished above, and the (nine) unary relations correspond to subsets of the domain where facts hold, e.g. the fact ‘1 holds red’ is interpreted as the subset $\{rwb, rbw\}$. This information state $(Hexa, rwb)$ is visualized on the left in Figure 1. The point is underlined. Deals connected by a labelled link are the same for that player.

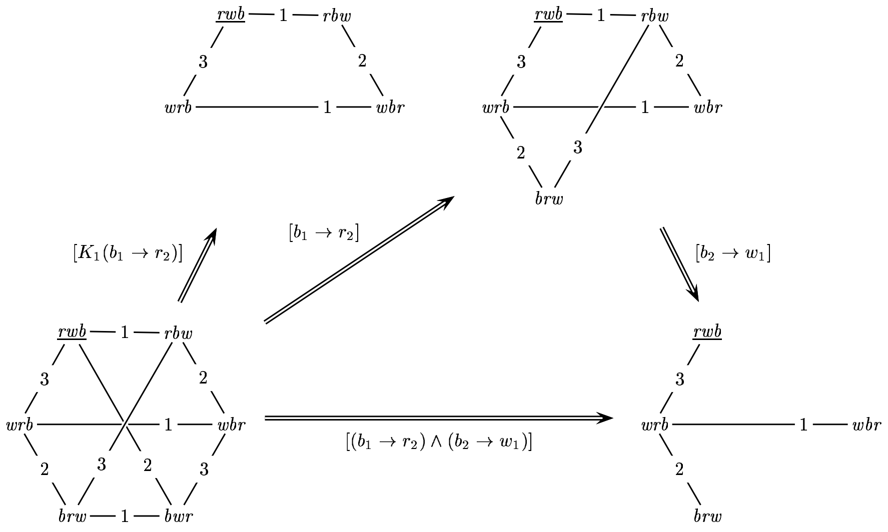

Figure 1: Examples of updates in the information state $(Hexa, rwb)$

We continue with defining the interpretation of propositions (statements). A fact p (such as ‘player 2 holds the white card’) holds in an information state, iff its point is in the subset $V(p)$ for that fact. Propositional logical connectives ($\neg, \wedge, \vee, \to$ for, respectively, ‘not’, ‘and’, ‘or’, and ‘implies’) take their standard interpretation. Proposition $K_n\varphi$ – ‘agent n *knows* that φ’ – holds in information state $(\langle S, \equiv, V\rangle, s)$, iff for all s' in S such that $s' \equiv_n s$, φ holds in $(\langle S, \equiv, V\rangle, s')$. Instead of *‘agent n does not know that not φ’* we also say *‘agent n can imagine that φ’*. The transitive closure of the union of all equivalence relations is also an equivalence relation: $\equiv_N := (\bigcup_{n\in N} \equiv_n)^*$. Two objects are equivalent in that sense if there is a finite chain of links (possibly empty) between them, whatever the labels. Proposition $C\varphi$ – ‘the agents *commonly know* φ’ – holds in $(\langle S, \equiv, V\rangle, s)$, iff for all s' in S such that $s' \equiv_N s$, φ holds in $(\langle S, \equiv, V\rangle, s')$. Both knowledge and common knowledge can be similarly defined for binary relations that are not equivalences, and the notion of common knowledge also extends to subgroups of the set N of agents. Finally, $[\psi]\varphi$ stands for *φ holds after update with ψ*. We also say, slightly abusing the language, that the update in this case is $[\psi]$. The relational interpretation of such updates is somewhat different from that of the other constructs, because we now have to refer to other information states. Proposition $[\psi]\varphi$ holds in information state (M, s), iff whenever ψ holds in (M, s), φ holds in the (I hope obvious) restriction of (M, s) to those states s' such that ψ holds in (M, s'). *Announcements* as in card problems are updates in this sense, because they are *public* and *truthful*, in other words, because they are spoken and everybody can hear them, and no lies are told.

We recommend the reader to check the following computations visually in Figure 1.

An atomic proposition or fact c_p describes that card c is held by player p. In information state $(Hexa, rwb)$ it holds that r_1, because player 1 has the red card in rwb. It holds that $K_1 r_1$ – 1 knows that he holds red –, because there is a 1-link between rwb and (only) rbw, and r_1 holds in both $(Hexa, rwb)$ and $(Hexa, rbw)$. It holds that $\neg K_1 \neg\neg K_2 \neg w_1$ – 1 can imagine that 2 can imagine that 1 holds white (even though 1 actually holds red) –, because rwb—1—rbw and rbw—2—wbr and in wbr 1 holds white. It holds that $C_{123}(K_1 r_1 \vee K_1 w_1 \vee K_2 b_1)$ – it is commonly known to all three players that player 1 knows his own card – because all deals of the information state are linked to rwb by some path, and because one of the three disjuncts holds for any deal: player 1 always knows his own card.

Now suppose that in information state $(Hexa, rwb)$ an insider says: "If 1 has blue then 2 has red". This is update $[b_1 \rightarrow r_2]$. This formula holds in all information states $(Hexa, s)$ except $(Hexa, bwr)$. Therefore update of $(Hexa, bwr)$ with $b_1 \rightarrow r_2$ results in the restriction of $(Hexa, bwr)$ to five deals. See Figure 1. If, instead, player 1 had said to player 2: "If I have blue then you have red", the update is $[K_1(b_1 \rightarrow r_2)]$ instead. Now both deal bwr and brw must be removed: $K_1(b_1 \rightarrow r_2)$ doesn't hold in $(Hexa, brw)$, because brw—1—bwr and $b_1 \rightarrow r_2$ doesn't hold in $(Hexa, bwr)$. In the spirit of example *iii*: player 1 can only know that 'If I have blue then you have red' is true, if she doesn't hold blue herself.

Update $[(b_1 \rightarrow r_2) \wedge (b_2 \rightarrow w_1)]$ results in an information state that illustrates why common knowledge is required for a solution of the cards problem: in the resulting information state (see Figure 1), 1 knows 2's card (a), 2 knows 1's card (b), and 3 doesn't know 1's or 2's card (c). The last holds, because rwb—3—wrb: 3 cannot distinguish betweens deals where 1 and 2 hold different cards. Also 3 doesn't know that 1 knows 2's card. Now the last is the same as saying that 3 can imagine that 1 doesn't know 2's card. And that holds because of rwb—3—wrb—1—wbr: 3 can imagine that the deal is wrb, and if it had been wrb 1 doesn't know 2's card, because 1 cannot distinguish wrb from wbr where 2 holds a different card. In other words, even though 1 knows 2's card, this is not common knowledge. In this information state, single update $[(b_1 \rightarrow r_2) \wedge (b_2 \rightarrow w_1)]$ gives the same result as the sequence $[b_1 \rightarrow r_2][b_2 \rightarrow w_1]$ of two updates, see Figure 1.

If a player p says φ in the cards game, this corresponds to the update $[K_p\varphi \wedge [K_p\varphi]Cc]$, where c is the description in this logic of 'C doesn't know any of A's or B's cards'. We have observed that $[K_p\varphi \wedge [K_p\varphi]Cc]Cc$ always holds (example *iv* provided an information state where $[K_p\varphi \wedge [K_p\varphi]c]\neg c$ holds). A solution of the cards problem consists of a (finite) sequence π of such updates, such that afterwards $C(a \wedge b \wedge c)$ holds (Cc holds anyway). This may also be expressed as the validity of a 'correctness statement' $\psi \rightarrow [\pi]C(a \wedge b \wedge c)$, where ψ is a description in epistemic logic of the initial information state. In an information state for card games, a proposition $C\varphi$ holds iff φ holds for all deals in its domain (iff for all $d \in S$, φ holds in $(\langle S, \equiv, V\rangle, d)$). This is the precision of 'all deals that are still consistent with the information given so far'.

We finish with a historical note. We have seen that statements about knowledge can be formalized in an epistemic logic, a 'logic of knowledge'. Epistemic logic is a modal logic. Modal logic is as ancient as Aristotle. Relational semantics for modal logic originates with (Kripke, 1959). Epistemic logic is traced to (Hintikka, 1962), and its extension with common knowledge to (Lewis, 1969) and later (Aumann, 1976), with various seminal contributions of different authors from (Fagin et al., 1995). Dynamic epistemic logic, i.e. extensions with updates and other dynamic features, is of more recent date: (Plaza, 1989; Gerbrandy, 1999; Baltag, 2002; van Ditmarsch, 2000). Note however that dynamic issues have an entirely separate history, see (van Benthem, 1996; Harel et al., 2000).

Solutions and generalizations

Again, we assume w.l.o.g. that A holds $\{0, 1, 2\}$, B holds $\{3, 4, 5\}$ and C holds 6. All solutions obviously satisfy the required constraints, we will not go into that any further.

NUM Call the solution in (Makarychev and Makarychev, 2001) ('each of the players A and B declares the sum modulo 7 of her/his three cards') NUM. There are various other solutions.

NUMb Alternatively, only A declares the sum modulo 7 of his three cards, and B merely announces C's card: "C has card 6".

The formulation of the problem in (Makarychev and Makarychev, 2001) is actually biased towards the NUM and NUMb solutions. If it is merely stated that all cards are different, say have names d, e, f, g, h, i, j, more symbolic solutions come to the fore. For simplicity, we keep their current names. For all solutions, B may either repeat A's protocol (or *any* other protocol!), or simply announce C's card, so we need not mention B's part any further. The protocols underlying the following solutions are left implicit.

NUM12 This is a variant of NUM. Player A *openly* renames all cards $1, 2, \cdots, 7$ so that his cards add up to 12. So she says: "I rename 0 to 3, and 1 remains 1, and ..., and the sum of my cards is now 12." As there are five different ways to add three of $1, 2, \cdots, 7$ up to 12 (and *only* for sum 12), this is also a solution.

CNF This is a nondeterministic protocol. CNF stands for 'conjunctive normal form'. This is its logical form. One of its twelve different executions has A-part:

> *A says: "I have one of the cards* $0, 1, 2$*, and one of* $0, 3, 4$*, and one of* $0, 5, 6$*, and one of* $1, 3, 6$*, and one of* $1, 4, 5$*, and one of* $2, 4, 6$*, and one of* $2, 3, 5$*."*

Here, 'one' means 'at least one'. After A's announcement B can deduce A's cards: "Because I have 3 and 4 and because A must have one of $0, 3, 4$, A has 0. Because I have 4 and 5 and A has one of $1, 4, 5$, A has 1. Because I have 3 and 5 and A has one of $2, 3, 5$, A has 2. So A's hand is $0, 1, 2$." Player 3 cannot derive any card of player 1 (or 2), because the solution is symmetrical in all cards.

DNF This is also a nondeterministic protocol. DNF stands for for 'disjunctive normal form'. One of its twelve different executions is:

> *A says: "I hold one of the following seven hands:* $\{0, 1, 2\}$*,* $\{0, 3, 4\}$*,* $\{0, 5, 6\}$*,* $\{1, 3, 5\}$*,* $\{1, 4, 6\}$*,* $\{2, 3, 6\}$*,* $\{2, 4, 5\}$*."*

The reader may determine for him/herself why this is also a good solution. To each execution of DNF corresponds a logically equivalent execution of CNF. The two shown are not equivalent.

In the NUM-like solutions, the use of numbers is not essential: A might have simply declared that he has one of five hands, instead of, in DNF, one of seven. So with DNF, less information is exchanged. Are there less informative solutions than DNF?

Players A and B can also communicate their hands under the given conditions if A holds four, B holds two, and C holds one card, and if A holds four, B holds seven, and C holds two cards. And they *cannot* if A and B hold two, and C holds one card. That is much harder to prove. What are the general criteria, for any number of players and cards, under which hands can be openly communicated in a finite number of communications? It is decidable: there is a crude algorithm, using model checking for finite multiagent information states. If it can be done, can it always be done in two communications? I don't know yet. More general results on making implicit knowledge explicit, suggest so. And what is the least informative way? I don't know either. All this seems relevant to cryptology.

Finally, back to (Makarychev and Makarychev, 2001). I initially contacted the authors because I thought I had found a counterexample to their requirement (1) for protocols. However, it was pointed out to me by Alexander Shen that in (1), instead of 'there should exist another configuration [that gives every card to a different player]', the correct translation from the Russian original should have been 'there should exist another configuration *where C holds the same card* [that gives every *other* card to a different player]' (see (Makarychev, 2001)). Ah, well.

References

Aumann, R. (1976). Agreeing to disagree. *Annals of Statistics*, 4(6):1236–1239.

Baltag, A. (2002). A logic for suspicious players: Epistemic actions and belief updates in games. *Bulletin of Economic Research*, 54(1):1–45.

Fagin, R., Halpern, J., Moses, Y., and Vardi, M. (1995). *Reasoning about Knowledge.* MIT Press, Cambridge MA.

Gerbrandy, J. (1999). *Bisimulations on Planet Kripke.* PhD thesis, University of Amsterdam. ILLC Dissertation Series DS-1999-01.

Harel, D., Kozen, D., and Tiuryn, J. (2000). *Dynamic Logic.* MIT Press, Cambridge MA. Foundations of Computing Series.

Hintikka, J. (1962). *Knowledge and Belief.* Cornell University Press, Ithaca, NY.

Kripke, S. (1959). A completeness theorem in modal logic. *Journal of Symbolic Logic*, 24:1–14.

Lewis, D. (1969). *Convention, a Philosophical Study.* Harvard University Press, Cambridge MA.

Makarychev, K. (2001). Logicheskie voprosy peredachi informacii (logical issues of information transmission). Master's thesis, Moscow State University. Diplomnaja rabota, part 1.

Makarychev, K. and Makarychev, Y. (2001). The importance of being formal. *Mathematical Intelligencer*, 23(1):41–42.

Plaza, J. (1989). Logics of public communications. In Emrich, M., Pfeifer, M., Hadzikadic, M., and Ras, Z., editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, pages 201–216.

van Benthem, J. (1996). *Exploring logical dynamics.* CSLI Publications.

van Ditmarsch, H. (2000). *Knowledge games.* PhD thesis, University of Groningen. ILLC Dissertation Series DS-2000-06.

van Ditmarsch, H. (2001). Killing cluedo. *Natuur & Techniek*, 69(11):32–40.