Eating from the Tree of Ignorance

Jan van Eijck and Rineke Verbrugge

Jan and Rineke are having a discussion while making preparations for their farewell talk at NIAS. They have already sent out a title and an abstract.

Rineke: Have you seen the instructions for the NIAS lectures? They frighten me a bit, really. Listen to this: NIAS talks should combine scientific depth with general accessibility; they should be geared at the general NIAS audience, but they should definitely be more than just superficial overviews.

Jan: I suppose we should not only talk about ignorance but we should also presuppose ignorance. Ignorance about logic, that is. Can you show me our abstract again, please?

Rineke: Here it is.

Eating from the Tree of Ignorance

Jan van Eijck & Rineke Verbrugge

In this talk, we will first introduce some examples of contexts, such as negotiations, where participants need to reason about others' knowledge and ignorance. Based on these examples, we will then introduce a logical model that can be used to reason about knowledge and ignorance: epistemic logic. This logic turns out to be well-suited for modeling social types of knowledge, for example common knowledge, which forms the basis of conventions such as "everybody drives on the right". Limitations of the idealized logical point of view on knowledge will also be given.

Many social interaction protocols are designed to preserve certain kinds of ignorance. Anonymity and privacy boil down to guaranteed absence of knowledge. We will analyse a number of interaction protocols with tools from epistemic logic. Finally, we will argue that certain types of ignorance may be beneficial for the individual ignorant agents, for a group, or even for society at large.

The title is certainly intriguing enough. Strictly between you and me: What exactly did you have in mind when you proposed it?

Jan: I would like to explain the concepts of common knowledge and lack of common knowledge to our audience, and analyse some examples where ignorance is bliss. In particular, I would like to explain that if I manage to keep information about my personal life out of the public view, I am helping others to protect their own privacy as well. Together we can reap the fruits of this. Ignorance is bliss, for knowledge can be exploited. Ignorance can also be exploited, of course, but this is well-known.

Rineke: It is indeed often the case that your ignorance proves your innocence. This is because many obligations are knowledge based. A doctor who does not know that a patient is sick does not have a legal or moral obligation to treat that patient. This is all very nicely analysed in [6].

Jan: The other day, Rohit Parikh came up with an amusing example of the reverse. A case where to prove innocence would have involved showing that one did know. Some policewoman, posing as a teenaged girl, engaged a middle-aged man in erotic discussions over the web. The man was convicted under some kind of "protection of children" act and sent to prison for five years. But the policewoman was not in fact a minor, and had the man been able to prove that he knew he was chatting with an adult woman, he would not have been convicted. So it was his ignorance about the true identity of his date—the fact that he did *not* know she was not a child—which created an obligation not to chat about sex. The obligation would have vanished had he known about the true identity of the person he was chatting up.

Rineke: An intriguing case, certainly. It revolves around what others know about what we know. But some people may just not care about what others know.

Jan: If we want to analyse such cases for our public, we'll need to give them some background on reasoning about knowledge, and in particular epistemic logic.

Rineke: Okay, I'll do that by starting with some contexts that they're all familiar with, negotiations. It turns out that in the Camp David negotiations between Israel and Egypt in 1979, which were mediated by Carter, both Sadat and Carter made some strange mistakes that where based on their lack of reasoning about the knowledge of the others. It was Carter's role in the negotiation to devise proposals that were then separately critiqued by Begin and Sadat, after which Carter would present a new proposal, until a proposal would be accepted by both parties. Did you know that already on the second day Sadat, who trusted Carter as a friend, presented a letter to Carter? This epistle outlined Sadat's fallback position, detailing all his possible concessions [5]!

Jan: I hope that Carter did not misuse this knowledge later?

Rineke: Well, not intentionally, but accidentally he let slip to Begin that he had received such a letter from Sadat, even if he did not spill the beans about its precise contents to Begin. After that, Begin, who was a savvy negotiator, started to offer inconsequential concessions and to expect large concessions from Egypt, and Carter never caught on that Begin was pushing him to move in the direction of Sadat's fallback position. In addition, Begin took care to make it common knowledge that the Knesset would never accept an Israeli concession on Palestinian self-government on the West bank and the Gaza strip, and after that the whole issue was more or less left out of the negotiation.

Jan: I suppose that Sadat never caught on?

Rineke: No, he did not, and neither did the rest of the world, until quite recent analyses. The way that the Sinai issue was resolved by giving it back to Egypt while demilitarizing it, was even presented as a prime example of good 'win-win' negotiation in handbooks such as *Getting to Yes* by Ury and his Harvard colleagues. I could show our public some epistemic formulas summing up the situation: Carter knew that Sadat was prepared to make a concession on that issue ($K_C p$), and Begin knew that Carter knew ($K_B K_C p$), but Sadat didn't know that Begin knew that Carter knew ($\neg K_S K_B K_C p$).

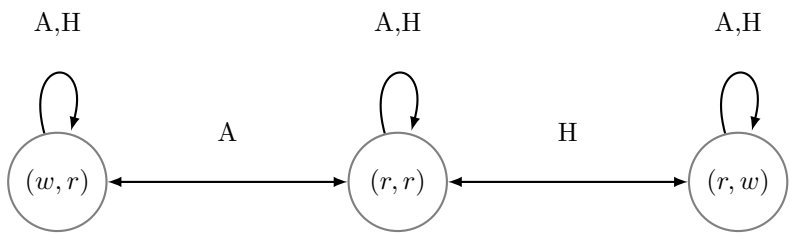
Jan: Will you also introduce possible worlds semantics in the lecture?

Rineke: Definitely, maybe first starting with a simpler example, such as the model with two states illustrating that our NIAS fellow Anne-Marie doesn't know whether it is raining right now in Damascus.

Jan: Then you could go on with a model of something more complicated, such as the wise persons puzzle.

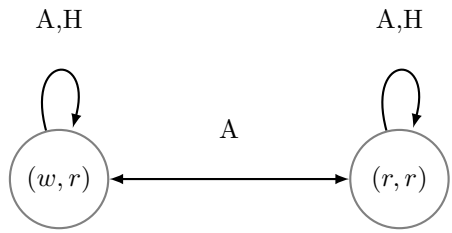
Rineke: That's a good idea. I'll tell them that there are two wise persons, Abelard (A) and Heloise (H). It is publicly known to everyone that there are three hats: two red ones and one white one. The king puts a hat on the head of each of the two wise persons, who cannot see their own hat but can see the other person's hat (and they both know this).

Jan: This is all easily captured by a Kripke model.



Rineke: That's right. Then I'll go on with the story: the king asks Abelard and Heloise sequentially if they know the color of the hat on their own head. The first person, Abelard, says that he does not know; the second person, Heloise, says that she knows.

I'll show the model after Abelard's admission of ignorance, and explain how cutting away accessibility arrows corresponds to eliminating ignorance:



Jan: I think by this point the public should be able to derive the color of Heloise's hat.

Rineke: I will do my best for that! After introducing the possible worlds semantics for common knowledge, I will tell the public something about people's cognitive limits in reasoning about other people's knowledge and ignorance, that I told you about before [4; 8].

Jan: Let us now go back to the issue of public and private information, and finally to ignorance as bliss.

Rineke: There is a saying "The innocent have nothing to fear", suggesting that only those with criminal intentions should worry about personal information getting public.

Jan: I don't know where you got that from, but I think it is very dangerous. The distinction between the public and the private sphere is fundamental in Western democracies. It is also in the Universal Declaration of Human Rights, in article 12. I looked it up.

> No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.

I take it that an attack on privacy is an attempt at finding out facts about my private sphere that I choose to keep out of public view. Lines between the public and private spheres are drawn differently in different countries. Those who have lived in the UK know that even British quality newspapers publish stories about naughty vicars. Dirty stories about public figures do appear in the Dutch press as well, but usually not in the quality papers.

Rineke: In some countries individual privacy may conflict with freedom of speech laws. Such laws may require public disclosure of information which would be considered private in other countries and cultures.

Jan: Let's talk about procedures for preserving anonymity and privacy, then. Sending emails without cc:s (possibly with encryption). Or: Sending letters in closed envelopes (maybe with a seal). Such procedures are meant to keep the contents of the messages private. Third parties should remain ignorant of the message contents.

Rineke: Another example is discreetly finding out if you share a secret with someone, without disclosing the secret if you are not. The other party should not find out the secret if she does not already know it.

Jan: That's all rather run of the mill. A more sinister example would be organizing a secret society on a need-to-know basis. This is meant to prevent the membership list of the society from becoming common knowledge. Even members should remain (partly) ignorant about who are their fellow members.

Rineke: Here is an example that our NIAS audience is very familiar with. The reviewing process of scientific papers is meant to preserve anonymity. Authors should remain ignorant of the identity of the reviewer, in the interest of objectivity.

Jan: Anonymous reviewing has a long tradition. I would like to talk about the dangers of compromising the anonymity of reviewing. Here is my slide.

Dr A. "By the way, were you one of the reviewers of my paper?"

Dr B. "I am sorry, but I think we should not discuss this matter. Maybe I was, maybe I was not. I am not going to tell you, for I believe in anonymity of reviewing."

Dr C, who actually was not a reviewer "Well, if I had been I would of course not have been allowed to tell you. But in fact I was not."

Dr D, who actually was a reviewer "No, I did not review your paper."

Dr D to Dr B "In fact I did review that crap, but of course I couldn't tell poor Dr A."

Rineke: I guess the moral of the story would be that it is really hard to preserve anonymity. You shouldn't even tell your colleagues that you did *not* review their paper, as Dr C innocently does. I wish all of us were Dr. B's ...

Jan: But the need for privacy protection has increased enormously in the electronic communication age. It should be possible to anonymously cast an electronic vote. The software should be designed in such a way that others should not be able to detect your vote.

Rineke: Modern software designs can even do better: It is possible to anonymously cast a receipt-free vote. The idea is that not only should others not be able to detect your vote, you should not be able to prove your vote. The vote is kept private even when the voter wishes to reveal it. This property is required in a setting with vote-buyers or coercers, where the voter might be tempted or forced to reveal his vote.

Jan: Privacy protection is big business these days: there is no shortage of programs for hiding my identity when I surf on the internet: Anonymizer, IDecide, Disappearing, Hushmail, Zip Lip, Zero Knowledge, ... All these programs routinely use public key encryption, by the way.

Rineke: We should explain to the public how that works.

Jan: Yes, that is a nice challenge. Suppose I tell them that 40285327 is the

product of two primes, and challenge them to produce these primes. Let us say I allow them the use of a pocket calculator. I am not going to hand out calculators, of course, but it should not be difficult to convince them that this is a hard task.

Rineke: You should list some attempts at solutions, and explain why the outcomes of the trial attempts do not yield information that can be used to improve the guesses.

Jan: Yes, let us say I give them some calculation results:

```
40285327 / 7
5755046.71428571

40285327 / 509
79146.025540275

40285327 / 5333
7553.97093568348

40285327 / 5347
7534.19244436132
```

Rineke: Of course, you will have to explain to them that the four numbers $7, 509, 5333$, and 5347 are all primes. More generally, you have to tell them that there are reasonably efficient ways of finding out whether very large numbers are primes.

Jan: Yes, of course. And finally I tell them that 7879 and 5113 are primes, and I demonstrate to them how easy it is to calculate their product:

```
   7879
   5113   ×
-------
  23637
  78790
 787900
39395000   +
-------
40285327
```

Rineke: That should drive home the moral that multiplication of two large prime numbers is easy, but finding the prime factors of a large number is very difficult. No known method for finding the prime factors of a number is substantially better than trial and error.

Jan: I could flash a slide with the RSA (Rivest, Shamir, Adleman) algorithm for public/private key generation [7]. *(Shows the slide.)*

1. Choose two large random prime numbers p and q,
2. Compute $n = pq$.
3. Compute the totient $\phi(n)$ of n.
   This is the number of positive integers i with $i \leq n$ and $\gcd(i, n) = 1$ (i co-prime to n).
   From p, q prime it follows that $\phi(n) = (p-1)(q-1)$.
   Example: $\phi(15) = 8$, for $1, 2, 4, 7, 8, 11, 13$, and 14 are co-prime to 15.
4. Choose an integer e with $1 < e < \phi(n)$ and e co-prime to $\phi(n)$.
   Release e as the public key exponent.
5. Compute d to satisfy $de = 1 + k\phi(n)$ for some integer k.
   I.e., $de = 1 \pmod{\phi(n)}$.
   Keep d as the private key exponent.

Rineke: It is not necessary to explain every detail. But the slide makes clear that as long as p and q remain secret, it does no harm to make n and e public.

Jan: I will just give an example of how this is used. Alice transmits her public key (n, e) to Bob and keeps the private key d secret. Bob then wishes to send message M to Alice. First he turns m into a number smaller than n. Next he computes cipher c given by $c = m^e \pmod{n}$ and transmits c to Alice. Alice can recover m from c by using her private key d, as follows: $m = c^d \pmod{n}$. From m, Alice can recover the original message M.

Rineke: The main point about public key cryptography, and the reason why it works so well, is that it is asymmetric.

Jan: There is a nice analogy to explain this: the simile of a padlock. Anyone can lock it, but only someone with the key can unlock it. If Bob has an open padlock and Bob knows that Alice is the only one with a key to it, Bob can send a secure message to Alice by putting the message into a box, locking the box with the padlock, and sending the locked box to Alice. For locking an open padlock you don't need the key, remember.

Rineke: That certainly explains why sending around public keys can do no harm. It is like sending about open padlocks, with a message on each padlock about who has the key to it. You should also explain how public key encryption can be used for authentication.

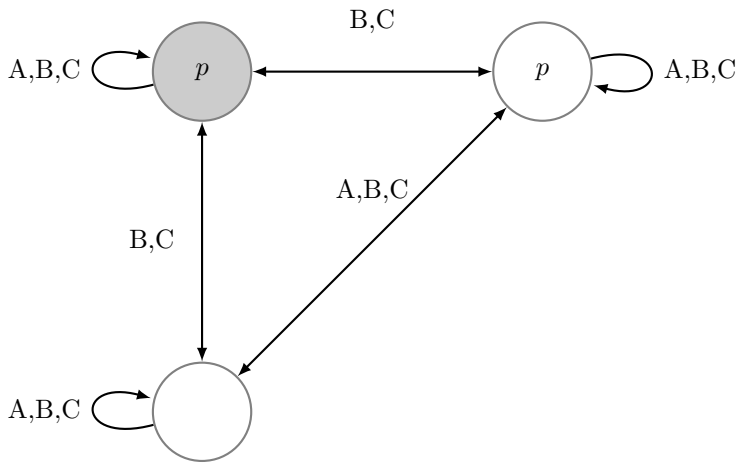
Jan: You mean, digital signatures? Unfortunately, this is where the padlock analogy breaks down. Digital signatures depend on the fact that the public key and the private key are inverses of each other. So a message encrypted with a private key can be decrypted with the corresponding public key. If you know that I am the only one who has the private key, then you can check that the encrypted message "This is my digital signature. Regards, Jan" really originates with me. Simply apply my public key to it and check if the expected plaintext comes out.

Rineke: But you have to be careful. This explanation may still confuse people. You have to make clear that what matters is that a message encoded with your private key can be decoded *only* with your public key. No other key will fit.

Jan: That's right. Your conclusion that the message must come from me depends on this.

Rineke: And on my assumption that your private key has not fallen into the wrong hands, of course. I suppose you could go on with explaining the effects of secret messages by showing the effects on Kripke models.

Jan: Yes, I have slides for this. Suppose p is a secret: Alice knows p, but Bob and Carol do not. They do not even suspect that Alice knows.

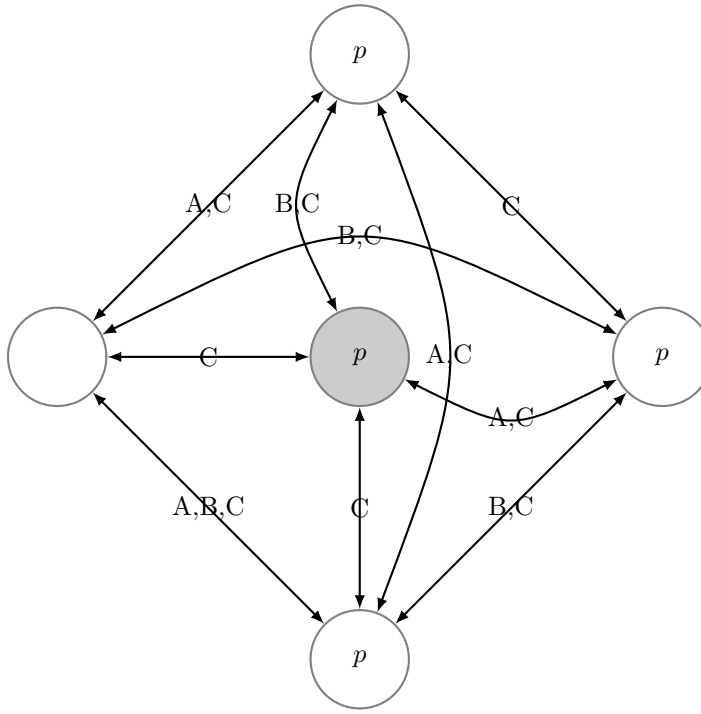


Rineke: Now the audience will certainly ask 'Why does this look so much more complicated?'

Jan: Yes, I will have to explain that the picture not only models the fact that Alice knows that p, but also that Bob and Carol do not know whether Alice knows or not. Bob and Carol cannot distinguish the actual situation (pictured in grey) from a situation where p is true but where Alice does not know this (the situation on the right in the picture) or from a situation where p is false but Alice does not know this (the situation at the bottom in the picture).

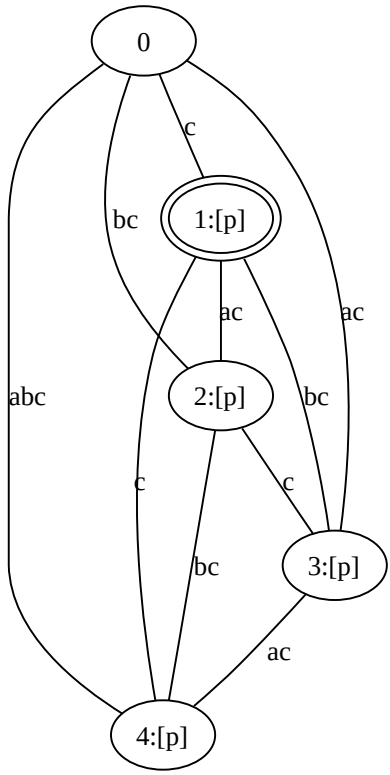
Rineke: And I suspect that the situation gets even more complicated when Alice tells Bob the secret?

Jan: Indeed, it does. For now Carol is the only one who does not know p, but Carol still does not know that the other two know. Here is a picture. This time I have left out the loop arrows at the individual nodes.



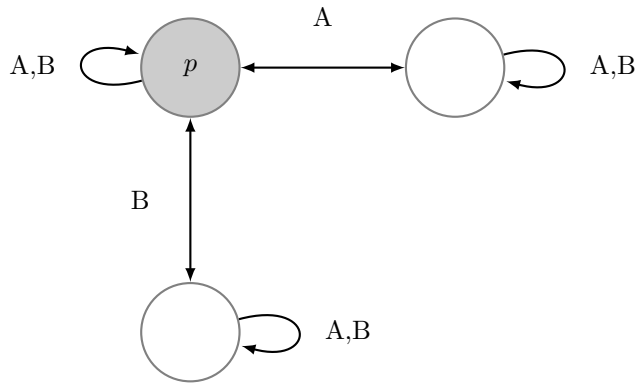
Rineke: Wow. How did you manage to construct that?

Jan: Well, it is a redrawing of a picture that was generated by my DEMO epistemic modelling tool [3]. Here is the original picture. This will give the audience a chance to apply what they learnt from you.



Rineke: There is also the well-known puzzle about a situation where neither A nor B knows whether p. Then they meet; B asks whether p, and A truthfully answers, 'Yes, I know'.

Jan: You mean the case where A has the additional piece of information that if p is not the case, then B knows that not p? The case of the chair of the programme committee who has been told by his secretary that all authors of rejected papers have been notified. When Doctor B meets Professor A, then B's question 'Has my paper been accepted?' reveals to A that the answer must be 'Yes', for A reasons that otherwise B would have known. Here is a picture:



Rineke: Another example you should certainly mention is the protocol of the dining cryptographers [1].

Jan: Chaum's famous protocol to protect privacy. Yes, I am sure they will love it. Three cryptographers are eating out. At the end of the dinner, they are informed that the bill has been paid, either by one of them, or by NSA (the National Security Agency). They want to find out whether NSA paid or not. They also want to respect each other's rights to privacy: In case one of them has paid the bill, her identity should not be revealed to the two others.

Rineke: Why is it important not to use a trusted outsider or a ballot box?

Jan: Methodological reasons. How can one convince oneself that a protocol is secure? The computer science approach is to make very strong assumptions about the presence of bad guys and about their capabilities. The use of a trusted outsider makes a protocol vulnerable. Traffic over an electronic network can reveal information, so it has to be assumed that all conversations can be overheard. Chaum's proposal demonstrates that the thing still can be done.

Rineke: I remember that each cryptographer tosses a coin with each of his neighbours, with the result of the toss remaining hidden from the third person. You can do a demonstration for the audience, with real coins hidden behind menus.

Jan: That should make it easy to explain why each cryptographer has a choice between two public announcements: That the coins that she has ob-

served agree or that they disagree. This is a public statement about private information: the others both hear it, but they cannot check the truth of the statement. And then the protocol is simply this:

- If she has not paid the bill she will say that they agree if the coins are the same and that they disagree otherwise;
- if she has paid the bill she will say the opposite: she will say that they agree if in fact they are different and she will say that they disagree if in fact they are the same.

Rineke: But why does this solve the problem?

Jan: We should let the audience find out, really. But that may be a bit tough on them. OK, note that as far as coin agreement is concerned there are just two possible situations. Either all coins agree (all heads or all tails), or two of the coins agree and the third one is different. Now assume nobody is lying about the agreement of the two coins she can see. Then if all coins are the same there will be no statements of disagreement, and if one of the coins is different there will be two statements of disagreement. So, if no one has picked up the bill there is an even number of disagreement statements. If one of the three is lying about what she observes, one of the statements will change. So if one of the three paid the bill, there will be an odd number of disagreement statements.

Rineke: It is clear why this calls for an epistemic analysis. To show that the procedure is secure one should show that the identity of the payer really is kept secret.

Jan: I prepared a slide for that. I analysed the situation with the DEMO model checker [3], as follows. I started out from the assumption that no one knew anything, and where this ignorance was common knowledge. Next, I updated with the public announcement of 'at most one cryptographer paid', so that this also became common knowledge.

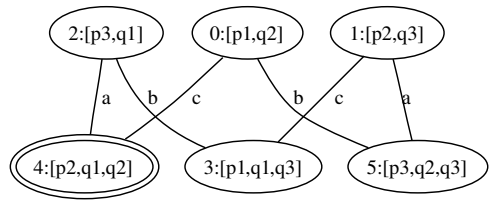
Rineke: Yes, for public announcements of factual information always generate common knowledge. We discussed these matters before (page **??**).

Jan: Next, I updated with the information that every participant knew whether she had paid or not. This is obvious, but it has to be spelled out. Next, update with the results of the coin tosses, update with appropriate

group announcements of the results of the coin tosses, and update with appropriate public announcements about coin (dis)agreement, and Bob's your uncle.

Rineke: Then you show the picture.

Jan: Here is my slide. Of course, I will have to explain that q_i means that coin i shows heads, and p_j that participant j has paid the bill.



- Cryptographer 2 paid, coins $1, 2$ show heads.
- $\neg K_a p_2, \neg K_c p_2$
- $C_{\{a,b,c\}}(\neg K_a p_2 \wedge \neg K_c p_2)$.
- $K_b(q_1 \wedge q_2 \wedge \neg q_3)$

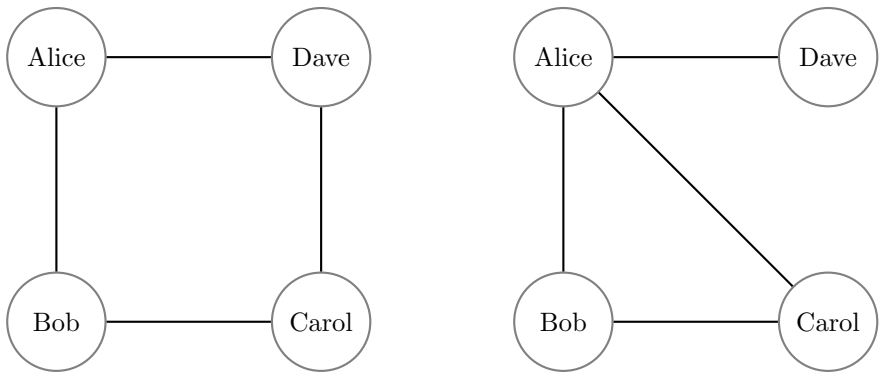
When I discussed the protocol some weeks ago, in a talk for an ILLC workshop in Amsterdam, Valentin Goranko made an instant proposal for a version of the protocol that could be used to check how many out of N dining cryptographers have made contributions to the bill, without revealing their identities. Let someone start by whispering a number M larger than N into the ear of her lefthand neighbour. The neighbour then whispers a number to his lefthand neighbour, and so on. The ones who did not pay pass on the same number they heard, but the ones who paid increase the number by one. After one round, the initiator of the protocol hears the number K, and she knows that $K-M$ people contributed to the bill. In the course of a second round everyone finds out, by comparing the number they heard the first time with the number they heard the second time.

Rineke: Brilliant.

Jan: But not quite as good as Chaum's original proposal. Assuming that every conversation can be overheard this is an insecure procedure.

Rineke: This audience will certainly be very interested in social networks and coordinated action. We should make sure we have enough time to discuss that as well.

Jan: I have prepared an example based on a nice description from [2]. Here are two social networks, where the individual members think the same and have the same intentions, but where the results are radically different.



The links picture the communication channels. In each picture, everyone thinks: "If I know for sure that at least two other people are going to take action, I will join in." In each picture, everyone communicates this intention to their neighbours.

Rineke: And then the point is that it makes a difference whether I am in touch with the neighbours of my neighbours or not.

Jan: Indeed. Because knowing that my neighbour will join if at least two other people join is not enough *for me* to be sure that he will join in. He can be sure about *me*. But how about his *other* neighbours?

Rineke: So in fact, in the situation on the left no one joins in, and in the situation on the right, where three people have links to the neighbours of their neighbours, these three people join in. Shouldn't we also mention examples where we don't yet have a full analysis? Otherwise, they might think that we have exhausted our subject at the end of our project.

Jan: How about Olmert's nuclear slip-up? This is really puzzling to me. Here is the relevant quote from the Internet:

> Ehud Olmert, the Israeli Prime Minister, faced calls for his resignation today after admitting — in an apparent slip of the tongue — that Israel has got nuclear weapons.
>
> But Israeli officials tried to push the cat back into the bag, denying that Mr Olmert had made any such admission and falling back on the Jewish state's policy of "nuclear ambiguity".
>
> Widely considered the Middle East's sole nuclear power, Israel has for decades refused to confirm or deny whether it possesses the atomic bomb. Mr Olmert appeared to break that taboo in an interview with a German television station as he began a visit to Berlin.
>
> TimesOnline, Dec 12, 2006

Clearly, even if everyone has individual knowledge that ϕ, public announcement of ϕ still has an epistemic effect. But that was not quite the case here. Twenty years ago, an Israeli dissident, Mordechai Vanunu, gave full disclosure of the Israeli nuclear program. This then became public knowledge, we must assume. The pictures he took were in the *London Sunday Times*, in 1986. In Vananu's own words, he was sentenced to 18 years in jail for "revealing something that everyone knew already" - or in our parlance, for turning general knowledge, "everyone knows", into common knowledge. Still, this seems to be not quite the same as an official public statement by the Israeli prime minister that 'Israel is a nuclear power'.

Rineke: And what is even more puzzling is how they think they can wiggle out of it again, by denying that Olmert had said what he said. Can public exposure be undone? By erasing the information from the minds of those who heard the interview?

Jan: That's a puzzle for sure.

Rineke: I am curious what our colleagues from the humanities and social sciences will have to say about all those intriguing dilemmas that we discussed today.

Jan: Me too. You know, I've begun to look forward to our farewell lecture and the discussions with our NIAS fellows.

Rineke: Only a pity that it is really meant as a farewell, and that our project will be over soon. There's still so much we could do! *(Looks wistfully.)*

References

[1] D. Chaum. The dining cryptographers problem: unconditional sender and receiver untraceability. *Journal of Cryptology*, 1:65–75, 1988.

[2] Michael Suk-Young Chwe. *Rational Ritual*. Princeton University Press, Princeton and Oxford, 2001.

[3] Jan van Eijck. DEMO — a demo of epistemic modelling. In Johan van Benthem, Dov Gabbay, and Benedikt Löwe, editors, *Interactive Logic — Proceedings of the 7th Augustus de Morgan Workshop*, number 1 in Texts in Logic and Games, pages 305–363. Amsterdam University Press, 2007.

[4] L. Flobbe, R. Verbrugge, P. Hendriks, and I. Krämer. Children's application of theory of mind in reasoning and language. *Journal of Logic, Language and Information*, 17:417–442, 2008. Special issue on formal models for real people, edited by M. Counihan.

[5] J. Oakman. The Camp David Accords: A case study on international negotiation. Technical report, Princeton University, Woodrow Wilson School of Public and International Affairs, 2002.

[6] Eric Pacuit, Rohit Parikh, and Eva Cogan. The logic of knowledge based obligation. *Synthese*, 31:311–341, 2006.

[7] R. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, February 1978.

[8] R. Verbrugge and L. Mol. Learning to apply theory of mind. *Journal of Logic, Language and Information*, 17:489–511, 2008. Special issue on formal models for real people, edited by M. Counihan.