Common Knowledge and Common Belief

Hans van Ditmarsch, Jan van Eijck, Rineke Verbrugge

Philosopher: Today, I suggest we discuss the important concepts of common knowledge and common belief. As far as I know, the first one to give a formal analysis of these concepts was the philosopher David Lewis, in his book *Convention* [26]. One of his examples is traffic conventions, about the role of common knowledge in how one behaves in road traffic. To explain this properly, I wonder if you would care to play a little game with me.

Cognitive Scientist: Sure.

Philosopher: Imagine yourself driving on a one-lane road. You have just come out of the Channel Tunnel on the British side and it is well-known that drivers who just went from France to England, or vice versa, tend to forget on which side of the road they have to drive, particularly if they find themselves on a quiet one-lane road where they are suddenly confronted with oncoming traffic. In case traffic comes towards you from the other direction, you will have to swerve a bit to the side to let it pass. In fact, you each have to swerve a bit.

Economist: Ah, this is beginning to sound familiar! If you swerve, you're a chicken. If not, and if you force the other to swerve, you're a tough guy. Unfortunately, when two tough guys come together, they will crash. There is interesting equilibrium behavior in examples like this. It's a standard setting for a two-person game in game theory [7].

Philosopher: Yes, you are right, but that is not what I wanted to explain. *(To the cognitive scientist again:)* Will you swerve left or right?

Cognitive Scientist: Well, if I remember that I am in England, where people have to drive on the left, I will swerve left. Otherwise, I will swerve right.

Philosopher: Yes, and how about the guy coming towards you? He and you

may both be cautious drivers, but if he will swerve right and you left, you will *still* crash. The point is that it is not enough for you and the on-comer both to know that you have to drive left. You would also like to know *that the other knows*. And this will affect your behavior. Wouldn't you agree that you will drive *more* cautiously—and swerve slightly more to the left—if you are uncertain whether the oncoming driver *also* knows that he has to drive on the left, than when you know that he knows to drive on the left?

Cognitive Scientist: Surely.

Philosopher: Then we are approaching common knowledge. Because surely you then also agree that this holds for the other driver as well. Now if you knew that the other driver did not know whether you knew to drive on the left, would that still affect your driving behavior?

Cognitive Scientist: It seems reasonable to be slightly *more* cautious when I do not know if he knows that I know, than when I know that he knows that I know, as his driving behavior will be slightly less predictable given his doubt about my knowledge—he might be tempted to avoid collision by a last-minute unexpected strong swerve to the right instead of to the left, if he were to think—incorrectly—that I am initiating that too.

Philosopher: Exactly. You are *very cautious* if you do not know, *slightly less cautious* if you know but not if the other knows, *even less cautious* if you know and also know that the other knows but not if he knows that, *and so on*: by repeating the above argument, you will all the time become slightly more confident about the other's road behavior but never entirely so. Driving on the left-hand side is what Lewis calls a *convention*, and this means that you know that I know that you know... up to any finite stack of knowledge operators.

Economist: As another instance of how relevant the concept of common knowledge is, you may care to mention that analyzing the properties of common belief is what earned the economist Robert Aumann the 2006 Nobel Prize for economics. In fact, independently from the logicians and philosophers, Aumann developed the concepts of common knowledge and common belief as ways to describe perfect rationality. Strategic choice assumes such common knowledge of each other's possible actions.

Computer Scientist: Ah, Aumann on agreeing to disagree. I have a surprise for you here. Recently at a very interesting workshop on new directions in

game theory in Amsterdam, the famous game theorist Dov Samet gave me a copy of an article by sociologist Morris Friedell, "On the structure of shared awareness", that already appeared in January 1969 in *Behavioral Science* [19]. This is based on a technical report from 1967, so it is even earlier than Lewis' much less technical book. Friedell's paper contains a proper definition of common knowledge, and it also has a wealth of fascinating examples of common knowledge in social situations. In fact, without knowing it, many later authors on common knowledge are just expanding on examples introduced by Friedell. So if anyone is the father of common knowledge, it is Friedell.

Economist: It is commonly believed among economists that Aumann was the first to give a formal analysis of common knowledge.

Philosopher: And it is commonly believed among philosophers that Lewis was the first.

Logician: But what Dov Samet was telling my colleague here shows that those common beliefs were wrong. A nice illustration of the fact that common beliefs may happen to be false.

Philosopher: Unlike cases of common knowledge. Maybe even something stronger was true: maybe it was commonly believed among economists that it was common knowledge that Aumann was the first to give this formal analysis. And *that* common belief was also false.

Computer Scientist: Friedell also has interesting things to say about ways of achieving common knowledge. Our dear Philosopher makes it sound like common knowledge is very hard to achieve. But that would be a mistake. Common knowledge is often easily achieved, by means of public announcement.

Cognitive Scientist: And what do you mean by public announcement, exactly?

Computer Scientist: Well, I suppose a public announcement is an event where something is being said aloud, while everybody is aware of who is present, and it is already common knowledge that all present are awake and aware, and that everybody hears the announcement, and that everybody is aware of the fact that everybody hears it, and ...

Cognitive Scientist: Ahem, an example may be clearer.

Computer Scientist: OK, at your service. It is already common knowledge among us that no one here has hearing difficulties and that everyone is wide

awake, right? *(In a loud solemn voice:)* I herewith announce to you all that the concert by Heleen Verleur and Renée Harp will take place on January 25. *(In a lower voice again:)* There you are. The date of the concert is now commonly known among the five of us.

Economist: Actually, this concert has already been announced by internal NIAS e-mail. But the thing is, several fellows don't read their e-mail, or only very irregularly. Should we still consider these e-mail notifications as proper public announcements?

Philosopher: I have a hard time remembering all those e-mails that I receive here.

Logician: There are various scenarios for which one can prove that it is impossible to achieve or increase a group's common knowledge [25; 28; 15].

Computer Scientist: I suppose the fact that fellows don't read their e-mails means that that channel is unreliable. Analysis of message passing through unreliable channels is old hat in computer science. We call it the *problem of the two generals*, or the *coordinated attack problem* Would anyone like me to elaborate?

Cognitive Scientist: Yes, please.

Computer Scientist: To immediately make the link with the topic at hand: it was proved by Halpern and Moses [21] that message exchange in a distributed environment, where there is no guarantee that messages get delivered, cannot create common knowledge. They use the example of two generals who are planning a coordinated attack on a city. The generals are on two hills on opposite sides of the city, each with their own army, and they know they can only succeed in capturing the city if their two armies attack at the same time. But the valley that separates the two hills is in enemy hands, and any messengers that are sent from one army base to the other run a severe risk to get captured. The generals have agreed on a joint attack, but they still have to settle the time.

Philosopher: So the generals start sending messengers. But they cannot be sure that the messengers succeed in delivering their message. And if they get through, there is no guarantee that the message of acknowledgement will get delivered. And so on.

Computer Scientist: You got the picture.

Philosopher: Suppose the general who sends the first messenger keeps sending messengers, all with the same story, until he gets an acknowledgement back, and then he keeps sending messengers to confirm the acknowledgement?

Computer Scientist: That procedure is known in computer science as the "alternating bit protocol" for sending bits over an unreliable channel. The sender repeats the transmission of a bit until an acknowledgement is received, then the sender acknowledges the receiver's acknowledgment until that is in turn acknowledged by the receiver, and only then the next bit is sent until that bit gets acknowledged, and so on.

Logician: The alternating bit protocol is also covered by Halpern and Moses' impossibility result. After the bit gets through, the receiver knows the value of the bit. After the acknowledgement gets back, the sender knows that the receiver knows the value of the bit. After the acknowledgement of the acknowledgement gets back, the receiver knows that the sender knows that the receiver knows the value of the bit, and if this gets confirmed, the sender knows that the receiver knows that the sender knows that the receiver knows the value of the bit. Still, this will not achieve common knowledge .

Philosopher: OK, you have made it quite plausible that message passing through unreliable channels cannot create common knowledge. And NIAS e-mail is perhaps not the proper medium for NIAS public announcements. But maybe we should turn it around: what are the properties of events that succeed in creating common knowledge? It seems to me that they all involve a shared awareness that a common experience takes place. It can involve various senses: hearing, seeing, maybe even touching or smelling.

Computer Scientist: This is getting sensual. Maybe intimate experiences such as eye-contact and touching are privileged in creating common knowledge? Friedell formulates the following obvious but important principle:

> If B sees A look at B, then A sees B look at A. From this and a few simpler properties one can demonstrate that eye contact leads to common knowledge of the presence of the interactants. It is no coincidence that eye contact is of considerable emotional and normative significance [19, page 34].

Cognitive Scientist: Would you stop looking me in the eye so intently, dear Computer Scientist? We already *have* common knowledge that we're both here... *(blushes)*

Computer Scientist: Indeed, here are those touchy situations that Friedell also analyzes, where some proposition is common knowledge, but the participants mutually pretend that the contrary proposition is the case [19]. If I'm not mistaken, such "open secret" situations will be extensively discussed during the NIAS lecture closing off our project (see page **??**).

Philosopher: There is a nice philosophy paper by Clark and Marshall about common knowledge as a background for mutual reference in discourse. They remark that common knowledge is often established by what they call "co-presence" [9].

Cognitive Scientist: Yes, but how does one know that an announcement has become common knowledge? I might have let my attention wander for a moment, or I might have misheard you. Actually, for smelling one would prefer some things *not* to be commonly observed. It is common practice in polite society to pretend one does not notice certain smells. This prevents what is generally known from becoming common knowledge.

Philosopher: I would put that differently. I would say it makes it possible to pretend of things that are in fact already common knowledge that they are not.

Computer Scientist: Seriously, whether you paid attention or not may not be the point. If an announcement is made, you were *supposed* to pay attention, and therefore the information can now be assumed common knowledge.

Philosopher: That is what happens in the public arena all the time. At the basis of legal relations between individuals and the state, or of the mutual legal relations between individuals, is the assumption that the law is common knowledge.

Cognitive Scientist: But this is a fiction. Professional lawyers have a full-time job to keep up with the law. Ordinary citizens can simply not be expected to cope.

Philosopher: You may call it a fiction. I prefer to say that it is a necessary presumption. Roman lawgivers found out long ago that if citizens within their jurisdiction could plead innocence because of being unaware of the law, no offender could ever get convicted. So they were quick to invent principles like *Ignorantia legis neminem excusat*, "ignorance of the law excuses no one".

Computer Scientist: And the counterpart of that is that the laws have to be

properly published and distributed. By being printed in a government gazette that every citizen has access to, for instance. Of course, the citizens are not supposed to read all that boring stuff. What matters is that they should be able to find out about it whenever they want. In this way, the publications in the government gazette amount to public announcements.

Cognitive Scientist: This connects to the conventions of driving that we started our discussion with. The traffic regulations are assumed to be common knowledge, although few people will be able to accurately reproduce all traffic rules. But if you are ignorant of the rules and cause a traffic accident, you are obviously still liable.

Philosopher: To prepare for this discussion I reread a classic publication from 1978 analyzing the concept of common knowledge, by Jane Heal [22]. Still a nice piece of philosophical exposition. The introduction is fabulous. Her work anticipates combining reasoning about knowledge and plausibility. If we're having dinner together and I drop a hot potato, it may be that

> I know that I have dropped that potato and so do you; but I hope and I believe that you do not know, and you hope that I do not know that you know" [22, p.116]

It also anticipates ways of linking knowledge to action for which, as far as I know, even now no good explanations can be given. Consider two agents, separated by a screen, who both repeatedly select one option from a set of many, simultaneously. When their selection is the same, a reward is given, and it is assumed to be common knowledge that they both get notified when that happens. But the notifications are private.

Computer Scientist: I don't think that is miraculous at all. This is a case where the private announcement "you get a reward for this choice" can achieve the effect of a public announcement, just *because* it is already commonly known that whenever one player gets the private announcement, the other player gets it as well. So what will happen is that after the first random common choice of C, the two players will keep choosing C to get rewarded again.

Philosopher: Ahem—exactly.

Computer Scientist: Michael Suk-Young Chwe's book *Rational Ritual* [8] also discusses such matters. Interestingly, Chwe pays attention to the *size* of groups for which common knowledge gets established. A brand name that

is common knowledge in a large group is worth a lot of money. Chwe analyzes the example of advertisements broadcasted during the American football Super Bowl. He compares the enormous cost of making something common knowledge by means of such advertisements to the obvious benefits. Part of the benefit is in the fact that the advertisements create common knowledge. An important consideration when deciding to buy a blu-ray media player, for example, is the knowledge that others are going to buy it too. The common knowledge created by an advertisement in the break of a nationwide TV-event gives the reassurance that lots of titles will soon become available in the new format.

Cognitive Scientist: I know that book. Actually, Chwe uses the example of the announcement of the new Apple Macintosh computer during a football Super Bowl, in 1984 I think. What I particularly like about the book is that it treats formal issues in a lucid not-technical way. But it assumes a firm grasp of technicalities, such as the distinction between general knowledge and common knowledge.

Logician: General knowledge among the members of a group of agents means that all individuals in the group know a certain fact, and *common* knowledge means: everybody knows that everybody knows, and so on [28; 15].

Computer Scientist: Let me propose a definition of common knowledge. A proposition φ is common knowledge if everybody knows that φ and everybody knows that φ is common knowledge.

Philosopher: That can hardly qualify as a definition. What you are saying is obviously circular. Besides, if I know that φ is common knowledge, then it logically follows that φ is common knowledge, for knowledge implies truth.

Computer Scientist: Yes, of course, but the definition states an equivalence. Truth does not in general imply knowledge, but in the case of common knowledge it does. If φ is common knowledge, then I know (and you know) that φ is common knowledge. And the circularity is not vicious.

Philosopher: I am of course familiar with recursive definitions, with a base case and a recursive case.

Computer Scientist: But this is an instance of what in computer science is known as a definition by co-recursion. Co-recursive definitions are like recursive definitions, but with the crucial difference that there is no base case. And they define infinite objects. Let me give you a simple example. An

infinite stream of zeros, call it *zeros*, can be defined as: *zeros* equals a zero followed by *zeros*. In lazy functional programming this is written as

```
zeros = 0 : zeros
```

If you execute this program in Haskell you will get an infinite stream of zeros flashing over your screen.

Philosopher: I suppose you mean an initial segment of an infinite list?

Computer Scientist: Yes, that is what I mean, of course. Even you are bound to get bored at some point, and break it off.

Philosopher: Ahem, nice example. Haskell is a programming language, I suppose?

Computer Scientist: Haskell is a language for functional programming, well suited for defining programs by co-recursion. As you can see from the example, Haskell uses colon for putting an element in front of a list. If you are interested, I can give you a reference to a textbook on Haskell programming with a whole chapter devoted to co-recursive definitions.[1] And I hope to have convinced you that my definition of common knowledge was as acceptable as my definition of the stream of zeros.

Philosopher: Yes, your recursive definition does make intuitive sense.

Computer Scientist: It is a co-recursive definition, not a recursive definition.

Philosopher: Thank you. I will try to keep the distinction in mind, at least while you are present. But let us move to the distinction between distributed knowledge and common knowledge. Am I right in saying that distributed knowledge is what a group would know if it had pooled their knowledge? If I know that p implies q, and you know p, then we have distributed knowledge of q.

Computer Scientist: Yes, that's right. Suppose Alice knows that p, Bob knows that p implies q, and Carol knows that q implies r. Then if they combine their resources they can figure out together that r is the case, so they have distributed knowledge that r. One obvious way to make r common knowledge is for Alice to shout p, for Bob to reply with announcing that p implies q, and therefore q, and for Carol to conclude by stating loudly that q implies r, and

[1] Doets and van Eijck [13].

therefore r. In short, they each make a public announcement of what they know, and their distributed knowledge turns into common knowledge.

Logician: Your example illustrates the difference quite nicely. Let us use Cp to express that p is common knowledge. If I know that p and you know that p implies q, these together do not imply Cq. But if Cp and $C(p \to q)$ then Cq. If p and $p \to q$ are common knowledge then the conclusion q is also common knowledge.

Computer Scientist: Indeed, that is all in accordance with the definition that I gave you.

Economist: (smiling) Well, it is common knowledge among economists that the analysis of common belief is crucial for understanding the way the stock-market functions. There may be rules of thumb for computing the value of stock like 'a share in company X should not cost more than twenty times the profit per share of company X', but these are not practical.

Philosopher: I suppose these days it is quite uncommon for companies to have an uninterrupted existence of twenty years. Without mergers or split-ups, I mean. Besides, nobody is willing to look that far ahead.

Economist: John Maynard Keynes, in his *General Theory of Employment, Interest and Money* [23] has something amusing to say about this:

> [..] professional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole; so that each competitor has to pick, not those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors, all of whom are looking at the problem from the same point of view. It is not a case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practise the fourth, fifth and higher degrees.

That is from Chapter Twelve, called "The State of Long-term Expectation".

Philosopher: You impress me. So you do really know your classics by heart?

Economist: Well, to be completely honest with you, I admit that I looked this one up for the occasion.

Philosopher: Your quote is interesting, for it talks about levels of mutual belief, and in the limit about common belief. The prize in the beauty contest goes not to the person who picks the prettiest girl, but to the person who picks the girl that is commonly believed to be the prettiest girl. If Keynes is right that the stock-market is about common belief, then the value of a share is what people believe it is. As long as a stock is commonly believed to be worth a lot, it does not matter if it is overvalued.

Economist: Until a stock-market crash occurs. Keynes himself was an avid speculator, and his friends had to bail him out during the crash that preceded the Great Depression.

Logician: That reminds me of the current credit crunch. I'm afraid that epistemic logic and the concept of common belief do not suffice to explain what's going on there. For example, imagine a rumor that a bank is going to go bankrupt. The rumor may be false, but it can start a chain reaction which results in the bank actually going bankrupt. If we want to be serious about social software, we need to be able to explain such a phenomenon, and possibly even to devise mechanisms to prevent them.

Economist: In fact, it does seem to me that epistemic game theory and behavioral game theory can already account for both epistemic and psychological aspects of the agents. In a recent paper by Bicchieri and Xiao [6], for example, the authors take on the challenge to investigate how social norms influence individual decision making. It turns out that what we expect others to *do* significantly predicts our own choices, much more than what we expect others to *think we ought to do*. Such findings are important if you want to design policies aimed at discouraging undesirable behavior [2].

Computer Scientist: So all this talk by the Dutch prime minister about norms and values will not influence the Dutch citizens' behavior one iota if we do not see the desired behavior around us.

Logician: That's what I always tell my spouse: It doesn't help to *tell* our children not to smoke or drink or lie: We should consistently set the right example. It's sure tiring to be a parent...

[2] These remarks about the 2008 credit crunch were inspired by contributions to an e-mail discussion by Rohit Parikh, Adam Brandenburger and Cristina Bicchieri.

Cognitive Scientist: Speaking about psychological aspects and children, common knowledge must also be relevant for what in cognitive science and psychology is known as 'theory of mind'. Around the age of four, children appear to develop a notion of another person's mind. They discover that what others think can be different from their own thoughts and that you can explain and predict other people's behavior in terms of their mental states. A well-known setting is the 'Sally-Anne' experiment[3] where a doll, Sally, puts a marble into her basket and then leaves the scene. While Sally is away and cannot see what happens, Anne takes the marble out of Sally's basket, and places it into her own box. Sally then returns and children have to answer the question where Sally will first look for her marble. Only from the age of four, children seeing the marble being moved will anticipate that Sally, who has *not* observed this move, will therefore later not know the new location of the marble.

Philosopher: Ah, that would explain why under-four-year-olds do not see the fun of performing magic tricks, for instance. The child knows that the coin is hidden beneath the sheet of paper, and the audience pretends to believe it has disappeared, and starts uttering sighs of amazement.

Logician: Yes, my five-year-old daughter loves that. Of course, the grown-ups have to play along by displaying their complete bafflement.

Logician: There may well be a relation between how conventions are formed in general and how a theory of mind develops in children. It seems only one step from whether you know that the other knows the location of a ball, to whether you know that the other knows on which side of the road to drive. But in such psychological experiments the higher-order setting never plays a role, as far as I know.

Cognitive Scientist: The standard setting of the Sally-Anne experiment does not test for higher-order aspects of knowledge: the child only needs to make a first-order false-belief attribution, that Sally believes that the marble is still in her own basket.

Logician: But recent investigations[4] pay special attention to just that higher-order aspect, and discuss experimental settings that corroborate the emergence of higher-order theory of mind, but only after the age of about six. It appears that even adults have some difficulty in applying third-order attributions such as "John doesn't know that Alice believes that he wrote a novel

[3] By Wimmer and Perner [35].
[4] See, e.g., [29; 34; 16].

under pseudonym". That is, of course, if they are not logicians.

Cognitive Scientist: Wow, if reasoning on three orders is already so hard for most of us, how can people ever draw correct conclusions about common knowledge, with all that complicated co-recursion it involves?

Economist: Indeed, it seems that in game settings people often just approximate common knowledge by a low stack of "we know that we know...", maybe only three or four levels [33].

Philosopher: Ah, now we are back on the English road where we started our discussion! As long as we know that we know that we know to drive on the left, we feel safe enough to proceed without swerving. At least I do.

Economist: I hope you don't mind if I get serious again. In "Agreeing to disagree" [1] Aumann introduces common knowledge as "everybody knows that everybody knows that ...". In the economics setting, instead of different *possible* situations—such as driving on the left, or on the right—the preferred model is that of different *probable* situations, and how events relate prior to posterior probabilities. Aumann shows that if agents have common knowledge of their posterior probabilities of an event, that these must then be the same. In other words, they can only agree to agree and they cannot agree to disagree. His presentation is elementary but it would still carry a bit too far to explain the details here.

Logician: What do you mean, carry us too far? Let *me* explain, then. The easiest way to explain what is behind Aumann's proof is this. It is not rational to agree to disagree, in an economic context at least, because this agreement would entail awareness of the fact that the disagreement can be exploited. What does it mean that you believe that the probability of an event is one half? Simply that if you are taking bets on this, then you will consider a bet with a return of two to one a fair bet. And if you believe that the probability is one in four and you are in a betting mood, then you will consider a bet with a return of four to one (including the stake) a fair bet.

Computer Scientist: Isn't that what bookies call an odds of three to one against? If the event happens you win three times your stake, otherwise you lose your stake.

Logician: That's right. Now consider what happens if I know that you believe that the probability of Barack Obama winning the presidential election is one fourth, and you know that I believe that this probability is one half. Then I

know that you are willing to take odds of three to one against, and you know that I am willing to take only equal bets. Then we should both be aware of the fact that someone can make money out of us, irrespective of how the election turns out.

Computer Scientist: Ah, I see. Assume Hillary Clinton places her bet of a thousand bucks with the guy who offers odds of three to one against Barack winning, and bets for two thousand bucks that Barack will lose with the guy who offers equal odds. If Barack wins, Hillary collects three thousand bucks from the first guy and loses her stake with the other, so she gains a thousand bucks. If Barack loses, Hillary loses her stake with the first guy but collects two thousands bucks from the other bloke, so again she pockets a profit of a thousand bucks.

Philosopher: Isn't that what gamblers call a Dutch book?

Logician: That's right. A Dutch book, a set of odds and bets which guarantees a profit, regardless of the outcome of the gamble, is what we have here. That's why agreeing to disagree is not rational for people who are willing to put their beliefs to the test by taking bets.

Philosopher: The explanation of degree of belief in terms of willingness to act, or to take bets, reminds me of Frank Ramsey's famous foundation of probability theory in terms of degrees of belief [32]. Ramsey remarks that the frequency account of probability does not explain what we mean by probability in cases of non-repeatable events. The election or non-election of Barack Obama is an example.

Logician: Actually the proof that Aumann gives does not involve betting or Dutch books. It is simply the observation that if φ is common knowledge between Alice and Bob, then φ has to hold in a set of members of the knowledge partition for Alice, and similarly for Bob.

Economist: There is also more recent stuff: in game theory, a lot of work is made of the analysis of strategic choice under assumptions of limited rationality. A case of opponent modeling where common knowledge is absent would be an example [14].

Philosopher: I am still wondering about this funny kind of definition that you call co-recursion. It seems like some kind of infinitary process is going on. How can we make sure it ever stops? I mean, imagine sending a romantic email, with 'I adore you' or that sort of thing. You get a reply "I am so glad to know

that you adore me", you send a reply back "Now I am delighted, for I know that you know that I adore you", only to get an exciting response: "How sweet for me to know that you know that I know that you adore me." Obviously, this nonsense could go on forever, and never achieve common knowledge of the basic romantic fact.

Logician: That's brilliant. For it *does* never stop if you do it like this. But if the two lovebirds get together, they may still go through the whole exchange that you mentioned, but only for the fun of it. For the first "I adore you" creates common knowledge.

Economist: Of course. And there are lots of everyday examples where the creation of common knowledge is crucial. Indeed, certain rituals are designed for it, and it is unwise not to observe them. Take the old-fashioned ritual that takes place when you withdraw a large amount of money from your bank account and have it paid out to you in cash by the cashier. The cashier will look at you earnestly to make sure she has your full attention, and then she will slowly count out the banknotes for you: one thousand (counting ten notes), two thousand (counting another ten notes), three thousand (ten notes again), and four thousand (another ten notes). This ritual creates common knowledge that forty banknotes of a hundred dollars were paid out to you.

Philosopher: Such rituals are important, indeed. Suppose you have four thousand bucks in an envelope, and you hand it over to a friend who is going to do a carpentry job at your home, say. Then what if this friend calls you later with dismay in his voice, and the message that there were just thirty-five banknotes in the envelope?

Economist: Then you are in trouble indeed, for you have failed to create common knowledge that the forty notes were there when you handed over the envelope. You failed to observe an important ritual, and this failure may result in the end of a friendship.

Logician: Maybe you only got what you deserved. Why pay for a carpentry job in cash unless one of you wants to fool the tax office?

Philosopher: Let us move on to the logic of common knowledge. How do we know that the concept of common knowledge is well-defined? And how do we know that common knowledge can be achieved in a finite number of steps?

Logician: The answer to the first question lies in a famous theorem by Tarski and Knaster. Let F be the operation of mapping a set of situations X to the

set of situations where X is general knowledge and where $F(X)$ is also general knowledge. Then this operation is monotonic. This means that it preserves the ordering on situations. If X is less informative than Y then $F(X)$ will also be less informative then $F(Y)$. Then F is guaranteed to have a fixpoint.

Philosopher: What do you mean by "less informative"?

Economist: And what is a fixpoint?

Logician: What 'less informative' means depends on the context. For sets of situations this will be reverse inclusion. If you can exclude more situations, you know more. Anyhow, Tarski and Knaster [24] prove that all monotonic functions have fixpoints. A fixpoint or fixed-point of a function F is a value X for which $F(X) = X$ [5].

Computer Scientist: Here is an easy example. The Dutch mathematician and philosopher of mathematics Brouwer proved a famous theorem stating that every continuous function from a compact convex set into itself has a fixpoint. Each map of the town of Wassenaar, where we are located here at NIAS, can be seen as the image of a continuous function that maps the real town onto its representation on the map. It follows from Brouwer's theorem that the map of Wassenaar that I have in front of me has the property that one point on the map coincides precisely with its pre-image.

Logician: Yes, of course, but you have to look really closely to see it. The fixpoint is the location of NIAS on the map. There is also a more procedural analogy for fixpoints. This is perhaps more illuminating in the context of common knowledge. Suppose you are painting your walls and you would like to mix exactly the same kind of beige as the small amount you have still left in your tin, which you now dub your "reference tin". Then you take a large new tin of white paint, and you keep adding small drops of brown and mixing, until you think you've almost attained the intended beige. At that moment you add a drop from the reference tin to the new mixture, without mixing, and look closely whether the reference drop is still darker than the new mixture. If it is, you go on adding drops of brown to the new tin and mixing, taking care to check at regular intervals. If you're careful, this procedure is bound to lead to the fixpoint. This works much better than trying out your new paint on the wall next to the old beige!

Philosopher: I like this. Let us move on.

[5] A lucid account of this material is in Davey and Priestley's textbook [10].

Logician: As you all know, an agent a is said to know φ in a state s if the proposition φ holds in all states that a cannot distinguish from s. These are called the "accessible" situations. Intuitively, "accessible from s" means "consistent with a's information in state s". You can picture this as a link with a label for the agent. If a state s where p is true is linked for agent a to a state s' where p is not true, this represents the fact that a does not know whether p is the case.

Philosopher: So when you talk about what agents know about what other agents know, this corresponds to more than one such step.

Logician: That's right. Let's take the case of two agents, Alice and Bob, who want to achieve common knowledge on who is going to collect the kids from daycare. Common knowledge is important here, for it is not enough that Alice knows that Bob knows that it is his turn today. Bob should also know that she knows. And so it goes on. *(writes on the whiteboard:)*

$$1 \text{ — Alice — } 2 \text{ — Bob — } 3 \text{ — Alice } 4 \text{ — Bob — } 5 \text{ — } \cdots$$

So there is a path, with a link from state 1 to state 2 for Alice because Alice cannot distinguish these states, followed by a link from 2 to 3 for Bob, for Bob cannot tell 2 and 3 apart, and so on. Something is common knowledge for Alice and Bob if it is true in all situations that are on such a path.

Philosopher: Ah, now I see how fixpoints come in. For common knowledge you have to compute the transitive closure of the union of the accessibility relations for Alice and Bob.

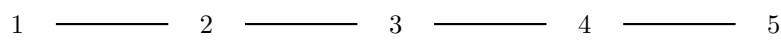
Logician: Exactly.

Computer Scientist: Let me elaborate. The fixpoint procedure for making a relation transitive goes like this:

1. Check if all two-step transitions can be done in a single step. If so, the relation is transitive, and done.

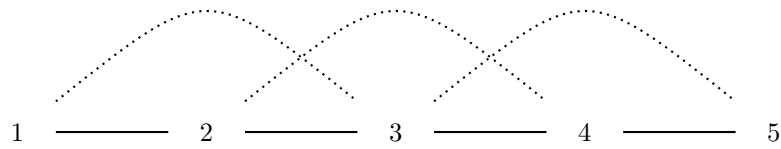
2. If not, add all two-step transitions as new links, and go back to 1.

Wait, let me draw a picture.

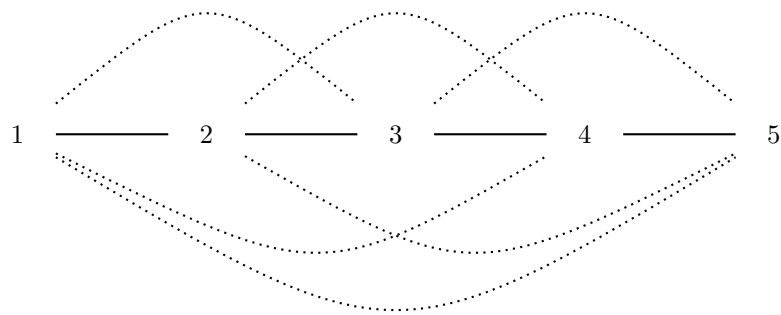


Philosopher: I suppose we can think of the link from 1 to 2 as a link for Alice, and the link from 2 to 3 as a link for Bob, and so on?

Computer Scientist: That's right, but I have blurred the distinction by taking the union of Alice's and Bob's links. Anyway, our check reveals that not all two step transitions can be done in single leaps, so the relation is not transitive. In the first step, we add all two-step links as new links:



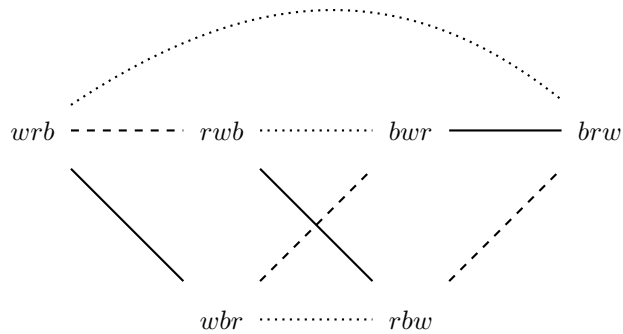
Now we check again. No, this is not yet transitive. So we add all two-step links in this new picture as extra links:



Philosopher: I can see that this is an example of a fixpoint procedure. You are changing the relation step by step, until it has the required property. After your final step the relation has indeed become transitive: all states are now

connected by direct links. So a proposition is common knowledge between Alice and Bob if it is true in all those states.

Computer Scientist: I have another nice example for this, a simple card game situation [12]. Consider the situation where Alice, Bob and Carol each receive a card from the set *red*, *white* and *blue*. They can all see their own card, but not those of the others. I will draw a possible worlds model of this situation. *(draws on the whiteboard:)*



Each world represents a card distribution in alphabetical order of the agents, with obvious color abbreviations. For example, *wbr* represents the state in which Alice has white, Bob has blue and Carol has red. The solid lines are for Alice. If she has white, she can see that she has white, but she cannot distinguish *wbr* from *wrb*. And similarly for the cases where she holds blue, and for the cases where she holds red.

Philosopher: Let me see. Then the dotted arrows must represent Bob's knowledge relation, and the dashed arrows Carol's. So now one can say things like "Alice holds white" by means of propositional atoms such as w_{Alice}.

Computer Scientist: That's right. "Alice holds white" is true in *wbr* and *wrb* but not in the other four worlds. Also, in both of these worlds "Alice knows that she holds white" is true, for the knowledge relation for Alice links *wbr* and *wrb*, and links no other worlds to these two, and in both of these w_{Alice} is true.

Philosopher: So in situation *wbr* it is common knowledge among Alice, Bob

and Carol that Alice doesn't know that Bob has blue? *(writes on the white-board:)*

$$wbr \models C_{\{Alice,Bob,Carol\}} \neg K_{Alice} b_{Bob}$$

Computer Scientist: That's right. This is because all six worlds can be reached from *wbr* in one or more steps by accessibility relations for agents in the group, and it is clear that $\neg K_{Alice} b_{Bob}$ in all worlds, for it holds everywhere that Alice can access at least one world in which Bob doesn't have blue.

Logician: There is a slight further subtlety. In the literature one finds both the transitive closure, and the reflexive transitive closure as definitions of the accessibility relation for common knowledge. The first is common among philosophers, and the second among computer scientists. Even standard text-books take different stances on this issue.[6] When modeling knowledge and not belief, both definitions amount to the same, because of the assumed property that *known* propositions are true. Nobody is so presumptuous as to claim the opposite implication that truths are always known. Computer scientists are more interested in knowledge. But for beliefs there is a real difference, and the most natural interpretation of common belief uses transitive closure only.

Philosopher: If you do not require that individual beliefs are true but then all of a sudden require that common beliefs are true, you get a rather confusing mix. So I suppose the philosophers were right in proposing transitive closure for both common knowledge and common belief.

Computer Scientist: When reading about societal problems like climate change, the confusing thing is the disagreement about what is common knowledge and what is common belief, which sometimes amounts to a common illusion.

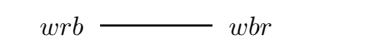
Logician: Yes, and what makes it worse is that there are certain think-tanks involved in the Republican War on Science [30] who are trying to create a common illusion *that* the greenhouse effect is a common illusion. But maybe we should save these matters for another discussion and stay with our present topic now. (See page **??**.)

Cognitive Scientist: I am still puzzled about some aspects of this common knowledge. In developmental psychology, even though we at some stage think

[6] Meyer and van der Hoek [28] take the reflexive transitive closure; Fagin et al. [15] the transitive closure.

to discover the *presence* of a theory of mind, we find it very hard to explain how such knowledge of others' knowledge is *formed*. Are we to think of this as some kind of category shift? Something that is not there initially, and then appears all of a sudden? If I understand this fixpoint process right, it must be very hard to achieve common knowledge in real-life situations. Can anyone say more about how this is possible?

Logician: I see this picture of computing transitive closure by means of a gradual process of adding links to a relation has confused you, and I am sorry. You should not think about common knowledge as a new relation that gets computed in stages, but as something that can be achieved in one go. Common knowledge of φ can be seen as the result of removing all non-φ situations from the picture. This can be done very easily, by means of a public announcement. Think of the card situation again. Suppose Alice suddenly says aloud: "I am holding white". Then the picture simplifies to this:



Now it has become common knowledge that Alice holds white. And it is common knowledge that the only uncertainty that remains is Alice's uncertainty about the cards of Bob and Carol.

Cognitive Scientist: So the result of publicly announcing φ is that φ will become common knowledge.

Logician: Well, not quite. Suppose instead of "I am holding white", Alice would have announced "I am holding white, but you guys don't know it yet." Then the second part of this becomes false as an effect of the announcement.

Philosopher: Alice is using a variation on the famous Moore sentence [31, p.543]: "I went to the pictures last Tuesday, but I don't believe that I did."

Logician: Yes, the effect can be truly destructive. "Your wife is cheating you, but you don't know it yet." After that announcement the addressee *does* know, so the statement has made itself false.

Cognitive Scientist: Moore sentences have the property that you cannot truthfully repeat them. So indeed, not all φ can be made common knowledge by publicly announcing them. I see that now.

Computer Scientist: By the way, the application of fixpoints to the logic of knowledge may originate with John McCarthy. In a small note in the early 1970s that at the stage he did not even consider important enough to publish[7] McCarthy formalizes two logical puzzles, one called the "Wise Men" puzzle (this is also known as "Muddy Children"), and the other a puzzle about numbers, called the "Sum and Product"-riddle. In the course of solving those riddles he almost off-handedly introduces the reflexive transitive closure of accessibility relations, and he uses this to account for what agents learn from the announcements made in those riddles. He also promises a further analysis in terms of a knowledge function, and handling time and learning, but I don't think that follow-up paper ever appeared.

Cognitive Scientist: So it seems we have another pioneer of the logic of common knowledge.

Logician: A lucid account of the interaction of public announcement and common knowledge can be found in a short note by Johan van Benthem from 2000, available on internet [5]. The crucial logical operation here is relativization. Imagine an information state involving several agents, with several worlds connected by agent accessibilities. Then the effect of a public announcement A is that all non-A worlds get eliminated from the picture. Van Benthem's key observation is that this *semantic* process of elimination of non-A worlds has as its *syntactic* counterpart the well-known logical operation of relativization of a formula to A. In the model that results from updating with the public announcement A a formula φ is true if and only if the relativization of φ to A is true in the original model. In the note Van Benthem then introduces the concept of relativized common knowledge, and conjectures that relativized common knowledge cannot be expressed in terms of plain common knowledge.

Computer Scientist: That squares well with an observation in Baltag et al.'s [2]. There it is shown that there is no sentence of the language of epistemic logic extended with a common knowledge operator that expresses "after public announcement of φ it is common knowledge that ψ".

Philosopher: I suppose relativized common knowledge is common knowledge relativized to an announcement? So it expresses what has to be true in a model *before the public announcement* in order to create common knowledge *after the announcement*?

[7] It was only later included in an overview of previously unpublished notes [27].

Logician: More precisely, after a φ announcement it is plain common knowledge that ψ if and only if it is φ-relativized common knowledge that after a φ announcement ψ holds. Later on, Van Benthem showed together with Van Eijck and Kooi [4] that if you take propositional dynamic logic as your epistemic language then the effect of *any* update that can be represented as a so-called finite action model is expressible in the epistemic language.

Philosopher: I thought propositional dynamic logic was designed for reasoning about the correctness of computer programs.

Computer Scientist: That's right. Propositional dynamic logic, or PDL for short, is an extension of Hoare logic.

Logician: But the beauty of formal systems is that they can be reinterpreted and reused. For instance, PDL has a construct for program composition: first execute program P, next execute program Q. We can reinterpret this to express the epistemic relation of what Alice knows about Bob's knowledge. Similarly, PDL has a construction for non-deterministic choice between two programs P and Q. We can reinterpret this as the relation of what Alice and Bob both know. Finally, PDL can express reflexive transitive closure, for executing a program P an arbitrary finite number of times. We reinterpret that as the reflexive transitive closure of a knowledge relation.

Philosopher: And taken together these PDL constructs can express common knowledge?

Logician: Yes, common knowledge between Alice and Bob that φ is expressed as follows. *(writes on the white-board)*

$$[(a \cup b)^*]\varphi$$

This is true if in every world that is reachable via the reflexive transitive closure of the union of the accessibility relations of Alice and Bob it holds that φ.

Philosopher: That is indeed what common knowledge amounts to. Now I suppose that PDL also has a construct that can be used to express relativized common knowledge?

Logician: Right again. For that you need PDL-tests, formulas that check that a condition holds somewhere in a program. The familiar programming construct of 'if φ then P else Q' is expressed in PDL by: *(writes on the*

white-board again)

$$(?\varphi; P) \cup (?\neg\varphi; Q).$$

What you need for relativized common knowledge is test for a property along a path, to express that in every world that is reachable via a sequence of φ worlds along the reflexive transitive closure of the union of the accessibility relations of Alice and Bob, it holds that ψ. Here is the formula: *(writes on the white-board)*

$$[(?\varphi; (a \cup b))^*]\psi.$$

Philosopher: Beautiful. Let me guess now. The principle that expresses the effect of public announcements on common knowledge will state that after public announcement of φ it has become common knowledge for Alice and Bob that ψ if and only if it is already φ-relativized common knowledge for Alice and Bob that ψ. Is that right?

Logician: Almost right. Let me use $!\varphi$ for a public announcement. Then this is what we get: *(writes on the white-board)*

$$[!\varphi][(a \cup b)^*]\psi \leftrightarrow [(?\varphi; (a \cup b))^*][!\varphi]\psi.$$

This has the shape of a reduction axiom: note that the public announcement $[!\varphi]$ occurs on both sides in the equivalence, but on the right-hand side the formula it has scope over has lower complexity. This means that the axiom can be used to define a translation from the language of PDL plus public announcement operators to the language of PDL without public announcement operators. And in [4] it is shown that this trick not only works for public announcements, but that something similar can be done for *any* update action.

Cognitive Scientist: This bit on relativized common knowledge went over my head, I am afraid. But I can appreciate the logical puzzles that have to do with common knowledge, such as the Wise Men puzzle and this Sum and Product riddle.

Computer Scientist: Then it may interest you that both of these riddles have old roots. The wise men riddle occurs in a puzzle book by Gamow & Stern from 1958 [20], but a friend of mine claims having seen this in Russian puzzle books from the first half of the twentieth century. The 'Sum and Product' riddle almost certainly originates with the Dutch topologist Hans Freudenthal. He stated it in the Dutch-language mathematics journal *Nieuw Archief voor Wiskunde* (New Archive for Mathematics) in 1969 [17] and presented its

solution in the next issue [18]. McCarthy only later became aware of that source of the riddle.[8]

Logician: In any case, it is clear that McCarthy's promise of follow-up was eventually fulfilled by the development of dynamic epistemic logic over the past 25 years or so![9]

[8]Details on the dissemination are in [11].
[9]Overviews of that development can be found in [15; 3; 12].

References

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

[2] A. Baltag, L.S. Moss, and S. Solecki. The logic of public announcements, common knowledge, and private suspicions. In I. Bilboa, editor, *Proceedings of TARK'98*, pages 43–56, 1998.

[3] A. Baltag, H.P. van Ditmarsch, and L.S. Moss. Epistemic logic and information update. In J.F.A.K. van Benthem and P. Adriaans, editors, *Handbook on the Philosophy of Information*, Amsterdam, 2008. Elsevier. To appear.

[4] J. van Benthem, J. van Eijck, and B. Kooi. Logics of communication and change. *Information and Computation*, 204(11):1620–1662, 2006.

[5] Johan van Benthem. Information update as relativization. Available from `http://staff.science.uva.nl/~johan/Upd=Rel.pdf`, 2000.

[6] C. Bicchieri and E. Xiao. Do the right thing: But only if others do so. *Journal of Behavioral Decision Making*, 21:1–18, 2008.

[7] K. Binmore. *Fun and Games*. D.C. Heath, Lexington MA, 1992.

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

[9] H. H. Clark and C. Marshall. Definite reference and mutual knowledge. In A. Joshi, B. Webber, and I. Sag, editors, *Elements of Discourse Understanding*, pages 10–63. Cambridge University Press, 1981.

[10] B.A. Davey and H.A. Priestley. *Introduction to Lattices and Order (Second Edition)*. Cambridge University Press, Cambridge, 2002. First edition: 1990.

[11] H.P. van Ditmarsch, J. Ruan, and R. Verbrugge. Sum and product in dynamic epistemic logic. *Journal of Logic and Computation*, 18:563–588, 2008.

[12] H.P. van Ditmarsch, W. van der Hoek, and B.P. Kooi. *Dynamic Epistemic Logic*, volume 337 of *Synthese Library*. Springer, 2007.

[13] K. Doets and J. van Eijck. *The Haskell Road to Logic, Maths and Programming*, volume 4 of *Texts in Computing*. King's College Publications, London, 2004.

[14] H.H.L.M. Donkers, J.W.H.M. Uiterwijk, and H.J. van den Herik. Selecting evaluation functions in opponent-model search. *Theoretical Computer Science*, 349(2):245–267, 2005.

[15] R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.

[16] 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.

[17] H. Freudenthal. (formulation of the sum-and-product problem). *Nieuw Archief voor Wiskunde*, 3(17):152, 1969.

[18] H. Freudenthal. (solution of the sum-and-product problem). *Nieuw Archief voor Wiskunde*, 3(18):102–106, 1970.

[19] M.F. Friedell. On the structure of shared awareness. *Behavioral Science*, 14(1):28–39, 1969.

[20] G. Gamow and M. Stern. *Puzzle-Math*. Macmillan, London, 1958.

[21] J.Y. Halpern and Y. Moses. Knowledge and common knowledge in a distributed environment. In *Proceedings of the 3rd ACM Symposium on Principles of Distributed Computing (PODS)*, pages 50–61, 1984. A newer version appeared in the *Journal of the ACM*, vol. 37:3, 1990, pp. 549–587.

[22] J. Heal. Common knowledge. *The Philosophical Quarterly*, 28(111):116–131, 1978.

[23] John Maynard Keynes. *The General Theory of Employment, Interest and Money*. Macmillan and Cambridge University Press, 1936. Full text available on the internet at `http://www.marxists.org/reference/subject/economics/keynes/general-theo%ry/`.

[24] B. Knaster. Un théorème sur les fonctions d'ensembles. *Ann. Soc. Polon. Math*, 6:133–134, 1928.

[25] Leslie Lamport, Robert Shostak, and Marshall Pease. The Byzantine generals problem. *ACM Trans. Program. Lang. Syst.*, 4(3):382–401, 1982.

[26] D.K. Lewis. *Convention: A Philosophical Study.* Harvard University Press, Cambridge (MA), 1969.

[27] J. McCarthy. Formalization of two puzzles involving knowledge. In Vladimir Lifschitz, editor, *Formalizing Common Sense : Papers by John McCarthy*, Ablex Series in Artificial Intelligence. Ablex Publishing Corporation, Norwood, N.J., 1990. original manuscript dated 1978–1981.

[28] J.-J.Ch. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science.* Cambridge Tracts in Theoretical Computer Science 41. Cambridge University Press, Cambridge, 1995.

[29] L. Mol, N. Taatgen, R. Verbrugge, and P. Hendriks. Reflective cognition as secondary task. In B.G. Bara, L. Barsalou, and M. Bucciarelli., editors, *Proceedings of Twenty-seventh Annual Meeting of the Cognitive Science Society*, pages 1925–1930, Mahwah (NJ), 2005. Erlbaum.

[30] Chris Mooney. *The Republican War on Science.* Perseus Books, New York, 2005.

[31] G.E. Moore. A reply to my critics. In P.A. Schilpp, editor, *The Philosophy of G.E. Moore*, pages 535–677. Northwestern University, Evanston IL, 1942. The Library of Living Philosophers (volume 4).

[32] F.P. Ramsey. Truth and probability. In R.B. Braithwaite, editor, *The Foundations of Mathematics and other Logical Essays*, pages 156–198. Kegan, Paul, Trench, Trubner & Co, London, 1931.

[33] D. O. Stahl and P. W. Wilson. On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10:218–254, 1995.

[34] 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.

[35] H. Wimmer and J. Perner. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13:103–128, 1983.