Logics of communication and knowledge
Sietsma, F.A.G.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 1

Introduction

1.1 Motivation

Communication is all around us. We all communicate as soon as we are among others, which is usually the greatest part of our waking time. We communicate with our friends about the things we did last weekend or the movie we want to see next week. We communicate with our family about who will do the dishes or where we want to go for holidays. And when we go to work, we communicate with our colleagues in order to do our job.

Communication is a very important way for us to influence and interact with the things and people around us. If we were unable to communicate this would entirely change the way we behave and interact. In modern society, communication with our peers has become even more important than the ability to build something ourselves.

We can distinguish many different kinds of communication. There is one distinction that is particularly relevant here. On the one hand there is the live conversation which is a rapid exchange of short messages, usually single sentences. On the other hand there is communication with messages that are sent and received at separate times. These messages are usually longer. Often the first type of communication is spoken and the second is written, but there are exceptions to this rule. For example, instant messaging is a written form of communication of the first type, and recording messages on a voice mail machine is a spoken form of communication of the second type.

Communication and knowledge are closely related. Indeed, the goal of communication is to share information with other people. If this information is known to be truthful, we may call it knowledge. In any case, every successful act of communication creates the knowledge that a certain message is communicated.

Communication can be very simple. For example, when I call my flatmate and tell her that I did the groceries, she will know she will not need to pass by the supermarket on her way home. But there is more to be observed in this
situation: I also know that she knows I did the groceries. This can be quite important because when I realize later on that I forgot something, I will call her again to make sure she does pass by the supermarket. Furthermore, she knows that I know that she knows that I did the groceries. Therefore, if she later on realizes she needs something special from the supermarket, she might call me to say that she will be late for dinner because she is going to pass by the supermarket after all. This already shows that even in very simple acts of communication, there is a lot to be analyzed.

But there are also more complex forms of communication. A well known example of this is the Two Generals Problem, first published in \cite{Akkoyunlu1975} and described in the following form in \cite{Gray1978}. Suppose there are two generals, whose armies are situated on opposite hills. In the valley between them is their common enemy and they want to attack this enemy. If one of them attacks on his own he will certainly lose. On the other hand, if they attack together they will probably win. Therefore, they need to coordinate their actions to agree on a common date and time of attack.

They start communicating by sending each other messages. Each messenger will have to pass through the valley where the enemy is encamped, and risks his life by doing so. Therefore, the generals can never be sure that the messages they send out will reach the other hill. Luckily, the generals have their own personal seals which make it impossible for the enemy to fake a message and create false belief among the generals. Will the generals be able to coordinate their attacks?

It may come as a surprise that the answer to this question is “no”. To see why this is the case, suppose that the first general sends the following message: “I will attack on Friday morning at nine o’clock!”\footnote{Note: This is for illustrative purposes only.}. Now of course this message may not reach the other general, but let us give the generals the benefit of the doubt and suppose the message does reach its destination. Then the second general will know the date and time of attack. But on Friday morning, the first general will discuss with his officers and reason as follows: it could be that his message reached the second general and the second general knows he is supposed to attack today. But it could also be that the messenger was shot on his way, and then the second general will not attack today. Then if I attack now, I will be alone and I will certainly lose. That is a risk the first general is not willing to take.

Therefore the first general changes his message a bit. Instead of just sending the date and time of attack, he also asks the second general to send a messenger back in order to confirm their agreement. Supposing this first message reaches the second general, he will send a messenger back. Suppose this second messenger reaches the first general again. Then on Friday morning, the second general will reason as follows: “If the first general received my confirmation, he will attack with me. But if he did not receive it he will probably not attack and then I will lose!”\footnote{Note: This is for illustrative purposes only.}. Therefore, the second general will not attack.

The second general could extend the communication protocol even further by asking the first general to send a confirmation of the confirmation he received,
but this will only move the problem back to the first general, who would then not know whether the second general received his confirmation of the confirmation. This problem cannot be solved: whatever messages the generals send, the one who sent the last message will never know if the other one received it. Therefore they cannot coordinate their attacks without risking to attack on their own.

What the generals lack is exactly the type of knowledge that I shared with my flatmate in the first example. In that situation, she knows I did the groceries, I know she knows this, she knows I know she knows, I know she knows I know she knows it, etcetera ad infinitum. We call this kind of knowledge common knowledge: my flatmate and I have common knowledge of the fact that I did the groceries. In the example with the generals, after the first message the second general knows the date and time of attack. After the first general receives a confirmation back from him, he will know that the second general knows the date and time. If he also sends a confirmation back and this confirmation reaches the second general, the second general will know that the first general knows that the second general knows the date and time of attack. However, the first general does not know this because he does not know whether the last confirmation reached the second general. In other words, the generals cannot coordinate because they do not share common knowledge.

The difference between the two situations is that when I talk to my flatmate on the phone, I am sure she can hear me. We instantly acquire common knowledge of the content of our conversation. On the other hand, the generals are never sure their message reaches the other side and therefore they cannot create common knowledge. This is an example of unreliable communication. In this work we will mostly assume that messages that are sent by one party are also received by the other party, and that this fact is common knowledge. An exception to this rule is Chapter 6 where we will distinguish between potential and definitive knowledge. In the example of the generals, the generals have potential knowledge of a message if it is sent to them and they have definitive knowledge of it if they also sent a confirmation of receiving it. Common knowledge is a very important concept which is extensively discussed in this work, especially in Chapter 5.

A fairly new form of communication is email communication. For example, instead of calling my flatmate to tell her I did the groceries, I could send her an email with this information. Email communication can also be more complex: I could include more people as Carbon Copy (CC) recipients in order to start a group conversation over email, or I could even include some Blind Carbon Copy (BCC) recipients who would receive the email without the other recipients being aware of this. In the first case, upon reading the email my flatmate would know that I did the groceries. In the second case, she would also know that the CC-recipients also know this, if they received the message. In the third case, her knowledge would be the same as in the second case because she cannot see the fact that there were BCC-recipients. However, if she takes the time to reflect on all possibilities she might realise that it is possible I included some BCC-
recipients. So in all cases she will consider it possible that more people than that she is aware of received my email.

All these considerations depend greatly on whether we assume that other people read their email. This may be a very reasonable assumption. For example, there are companies where it is required of the employees to check their email daily and read everything of importance. In private communication this is less strict, but even then there are people who can be counted upon to read their email at least daily and sometimes even hourly. On the other hand, there are also people who forget to check their email, or simply do not read all emails they receive. And even if we read our email thoroughly, there are spam filters that may accidentally remove email from our inbox or network errors that may result in emails being lost.

In the example with me and my flatmate, the situation can be analyzed easily by hand. It is not hard to figure out who knows what by just looking at the email I sent her. However, when the number of emails and recipients grows this analysis becomes a daunting task and infeasible for humans. For example, in large companies tens of thousands of emails are sent and received every day. When some secret piece of information was leaked via email, the complexity of finding out who was the source of this information leak, or who else received this secret information, is overwhelming. In situations like this, it would be very helpful if the analysis of people’s knowledge during an email conversation could be automated. Due to the intricacies involved when studying knowledge, knowledge about knowledge and common knowledge, logic is a very suitable tool for such an analysis. This explains the extensive reliance on logic in this thesis.

1.2 Overview of the dissertation

The general set-up of this dissertation is as follows. I first give some preliminary definitions in Chapter 2. In Chapters 3, 4, 5 and 6, I present four different models of how knowledge evolves during communication. Each of these models depends on different assumptions and is therefore suitable for different situations. Chapter 3 focuses on a situation where all possible messages are known by the agents, for example during a game or during the execution of some protocol. The model presented in Chapter 4 is a very general model which can be used to model many types of communication. It is not tailored towards one single situation, but can be adapted as desired. Chapter 5 and 6 focus specifically on email communication. The model presented in Chapter 5 is of a more theoretical nature and focuses on modeling common knowledge. Also, it rests on the assumption that all emails that are sent are also received and read. On the other hand, the model presented in Chapter 6 distinguishes two kinds of knowledge in order to make a distinction between when an email is sent and when it is also read. In Chapter 7, I take a closer look at the models that are used in Chapter 3. These are the so-called
action models that are used in epistemic logic to model communicative actions. In Chapter 8 I show how these models can be used in communication about beliefs and belief revision and finally, in Chapter 9 I study a situation in which the agents that communicate are not necessarily truthful, which leads to a study of the effect of lying. I also present a case study of a game of Liar’s Dice.

The contents of each chapter is briefly sketched below.

**Chapter 2** This is an introductory chapter explaining some basic concepts from epistemic logic.

**Chapter 3** In this chapter I propose a framework for modeling message passing situations that combines the best properties of dynamic epistemic semantics and history-based approaches. I assume that all communication is truthful and reliable. I also assume there is a dynamic set of messages that may be sent, which is known by all agents. The framework consists of Kripke models with records of sent messages in their valuations. I introduce an update operation for message sending. With this update I can study the exact epistemic consequences of sending a message. I define a class of models that is generated from initial Kripke models by means of message updates, and axiomatize a logic for this class of models. Next, I add an update modality and sketch a procedure for defining it by means of equivalence axioms. This chapter is based on joint work with Jan van Eijck [Sietsma and van Eijck, 2011].

**Chapter 4** In this chapter, I develop a very general framework based on epistemic logic that can be adapted to the needs of a great number of different situations. The network over which the agents communicate is explicitly specified in this framework, and therefore it can be used to model a situation where not all agents are able to communicate with each other. By combining ideas from Dynamic Epistemic Logic and Interpreted Systems, the semantics offers a natural and neat way of modeling multi-agent communication scenarios with different assumptions about the observational power of agents. I relate the logic to the standard DEL and IS approaches and demonstrate its use by studying a telephone call communication scenario. This chapter is based on joint work with Yanjing Wang and Jan van Eijck [Wang et al., 2010].

**Chapter 5** Here, I focus on email communication specifically. I consider a framework in which a group of agents communicates by means of emails, with the possibility of replies, forwards and BCC. I study the epistemic consequences of such email exchanges by introducing an appropriate epistemic language and semantics. This allows me to find out what agents exactly learn from the emails they receive. Common knowledge plays a big role in this framework and I show how to determine when a group of agents
Chapter 1. Introduction

acquires common knowledge of the fact that an email was sent. I also give an analysis of BCC and I look at email communication from the perspective of distributed systems. This chapter is based on joint work with Krzysztof Apt [Sietsma and Apt, 2012].

Chapter 6 In this chapter I also analyze email communication, but now I focus on the difference between sending an email and knowing its content has been read. This is not the same thing, especially when one considers the existence of network errors, spam filters and people who simply do not read all the emails they receive. Such an analysis is interesting in many situations. One example is when someone’s knowledge about some email at a particular moment may be relevant in a court case. I distinguish two kinds of knowledge: potential knowledge, which is acquired at the moment an email is sent to someone, and definitive knowledge, which is acquired when that person also shows his knowledge of the email by replying to it or forwarding it. I incorporate both kinds of knowledge in my logic. I present a semantics for this logic that can be decided quite easily and is therefore applicable in practice. I also show that from the epistemic point of view, the BCC feature of email systems cannot be simulated using messages without BCC recipients. This chapter is based on an unpublished manuscript that I finished in 2012.

Chapter 7 In this chapter I take a closer look at the models I use in Chapters 3 and 9. These are Kripke models, used to model knowledge in a static situation, and action models, used to model communicative actions that change this knowledge. The appropriate notion for structural equivalence between modal structures such as Kripke models is bisimulation: Kripke models that are bisimilar are modally equivalent. I would like to find a structural relation that can play the same role for the action models that are of great importance in information updating. Two action models are equivalent if they yield the same results when updating Kripke models. More precisely, two action models are equivalent if it holds for all Kripke models that the result of updating with one action model is bisimilar to the result of updating with the other action model. In this chapter I propose a notion of action emulation that characterizes the structural equivalence of the important class of canonical action models. Since every action model has an equivalent canonical action model, this gives a method to decide the equivalence of any pair of action models. I also give a partial result that holds for the class of all action models. This chapter is based on joint work with Jan van Eijck [Sietsma and van Eijck, 2012].

Chapter 8 This chapter focuses on the interplay between knowledge and belief. Models of knowledge change into models of belief when one drops the
1.2. Overview of the dissertation

assumption that all communication is truthful. This corresponds to the assumptions that all relations in the Kripke models are equivalence relations. In this chapter, the only constraint I impose on these relations is that they are linked. Linkedness is a new extension of the notion of local connectedness for multiple agents. It assures that if there are three alternatives, one agent prefers the second over the first, and the other agent the third over the first, that both agents make up their mind about whether they prefer the second or the third alternative. This is important in consensus-seeking procedures like Dutch meetings, where the participants vote on different subjects according to a set agenda. I show how my framework can be used to model such procedures, and use it to analyze the discursive dilemma, a well known problem in judgement aggregation [List and Pettit, 2005]. This chapter is based on joint work with Jan van Eijck [Sietsma and van Eijck, 2008].

Chapter 9 This chapter has a more philosophical flavor as compared to the other, more technical, chapters. I model lying as a communicative act changing the beliefs of the agents in a multi-agent system. Following St. Augustine, I see lying as an utterance believed to be false by the speaker and uttered with the intent to deceive the addressee. The deceit is successful if the lie is believed by the addressee. I provide a logical sketch of what goes on when a lie is communicated. I present a complete logic of manipulative updating, to analyze the effects of lying in public discourse. Next, I turn to the study of lying in games, in particular the game of Liar’s Dice. First, a game-theoretical analysis explains how the possibility of lying makes such games interesting, and how lying is put to use in optimal strategies for playing the game. I also give a matching logical analysis for the games perspective, and implement that in the model checker DEMO. There is a difference between lying in games and the logical manipulative update: instead of taking each utterance to be truthful, in a game the players are aware of the fact that the other players may lie. This chapter is based on joint work with Hans van Ditmarsch, Jan van Eijck and Yanjing Wang [van Ditmarsch et al., 2012].