Knowledge in strategic situations
A survey

Eric Pacuit*
Department of Computer Science
The Graduate Center of CUNY
365 5th Avenue, NYC, NY 10016
www.cs.gc.cuny.edu/~epacuit
epacuit@cs.gc.cuny.edu

Abstract

This paper surveys some recent papers that fit into a general project of developing formal tools for reasoning about social software. When developing such models, it is important to take into account both agents' utilities and agents' information. The two papers discussed in this paper primarily focus on the latter by making explicit the role of information in multi-agent situations. In the setting of deontic logic, [PPC] illustrates how knowledge effects an agent's obligation.

Choices that agents make, at least partially, depend on the agents' states of knowledge. In turn, the occurrence of an event changes the agents states of knowledge. There has been a lot of recent discussion in the literature about how to add such dynamic effects to otherwise static models of knowledge. [PP] looks at a formal model in which agents' knowledge changes as a result of communication. All results in this paper are taken from [PP] and [PPC].

1 Introduction

Social software, as discussed by Parikh in [Pa], generalizes the concept of computer software to a multi-agent social setting. The analysis of multi-agent procedures must take into account the utilities and information of the individual agents involved in the procedure. After all, when designing a piece of computer software, the programmer need not worry that a computer may simply decide that the next step of the algorithm is not worth its time. But in a multi-agent setting, such considerations must be taken into account when designing a social procedure. As a consequence, the study of social software must draw on results from game theory, mechanism design, artificial intelligence, philosophical logic and economics.

Just as one can prove that a certain implementation of a sorting algorithm is correct, perhaps one can prove that a certain piece of social software is correct. Before attempting such a correctness proof, an appropriate model of multi-agent knowledge is needed. Indeed, for may pieces of social software, in order to state correctness conditions, one must refer to the knowledge and beliefs of the agents in a particular situation.

Whether made explicit or implicit, knowledge theoretic properties such as common knowledge of rationality are important in understanding and modeling game-theoretic situations. Two questions are of interest when formalizing information in a strategic situation. First, it is natural to assume that the

*This survey is based on two papers: [PP] with Rohit Parikh, and [PPC] with Rohit Parikh and Eva Cogan.

agents' choices, at least partially, depend on their current states of information. This is particularly important when reasoning about what agents ought to do. Certainly an agent cannot be faulted for not performing an action whose need it did not know about. How should this dependency of actions on knowledge be modelled? In the context of deontic logic, this question is discussed in [PPC]. Second, another natural assumption is that the information an agent has changes by the actions that are performed. How should this dynamic feature be added to otherwise static logics of knowledge. In [PP], Pacuit and Parikh look at a multi-agent $\mathbf{S5}$ model with a "communication" modality.

The basic formal framework used is a history based model. The idea is that each agent has a set of possible actions, or choices, and at each moment in time the agents perform some action. Assume that only one agent performs an action at any moment. A sequence of events is called a (global) history. In computer science, history based models have been used to model computations in a distributed environment. See [FHMV] for a thorough discussion. In the game theory literature, the history based models are called extensive games. The reader is referred to the textbook [OR] for a discussion of extensive games. Halpern has a recent paper that demonstrates the usefulness of the history based model in game theory ([H03]). Various assumptions such as perfect recall, unique initial state, no learning, global clocks, etc. have been widely discussed in the computer science and game theoretic literature. Unless otherwise noted, we assume that each agent has access to the global clock and each agent has perfect recall. Intuitively, this means that as time passes the set of histories an agent considers possible decreases. The formal details of the model will be discussed in the next section.

This formal model provides a very natural setting to reason about agents in a social situation. In [PR], Parikh and Ramanajam use this model to provide a semantics of messages in which notions such as Gricean implicature can be represented. The first paper discussed in this survey ([PPC]), extends the basic model by attaching a real number to each global history, which is intended to represent the *social* utility of that history.

Starting with the intuition that agents cannot be expected to perform actions they are unaware of, in [PPC], Parikh, Pacuit and Cogan present a multi-agent logic of knowledge, action and obligation. The semantics used is a history based model described above. This model is applied to various deontic dilemmas that illustrates the dependency of an agent's obligation on knowledge. For instance a doctor cannot be expected to treat a patient unless she is aware of the fact that he is sick, and this creates a secondary obligation on the patient or someone else to inform the doctor of his situation. In other words, many obligations are situation dependent, and are only relevant in the presence of the relevant information. This creates the notion of *knowledge based obligation*.

Both the case of an absolute obligation (although dependent on information) as well as the notion of an obligation which may be over-ridden by more relevant information are considered. For instance a physician who is about to inject a patient with drug d may find out that the patient is allergic to d and that she should use d' instead. Dealing with the second kind of case requires a resort to non-monotonic reasoning and the notion of *weak knowledge* which is stronger than plain belief, but weaker than absolute knowledge in that it can be over-ridden.

The second paper discussed in this survey is [PP] in which Pacuit and Parikh present a logic of multi-agent knowledge with communication. The logic presented in in [PP], is a extension of *Topologic* to the case of many agents. *Topologic* is a bimodal logic of knowledge and "effort" first discussed by Moss and Parikh in [MP]. The basic intuition is that "effort" in a multi-agent setting is simply communication among some of the agents.

Agents are assumed to have some private information at the outset, but may refine their information by acquiring information possessed by other agents, possibly via yet other agents. Each agent's information is initially represented by a partition over a set of possible states. Agents are assumed to be connected by a *communication graph*. In the communication graph, an edge from agent i to agent j means that agent i can directly receive information from agent j. Agent i can then refine its information by learning information that j has, including information acquired by j from another agent, k. We introduce a multi-agent modal logic with knowledge modalities and a modality representing communication among agents. The validities of *Topologic* are shown to remain valid and the communication graph is completely determined by the validities of the resulting logic.

Similar logics have been studied starting with [Pl] and more recently in [BM, K, Vd, Ge]. In chapter 4 of [K], Kooi provides an excellent overview of the current state of affairs of these dynamic epistemic logics. We do not consider general epistemic updates as is common in the literature, but rather study a specific type of epistemic update and its connection with a communication graph.

This paper is organized as follows. The next section is brief overview of the formal models used in both papers. Section 3 presents some of the examples from [PPC]. Section 4 is a summary of [PP]. Finally the last section lists some ongoing research related to both papers.

2 Background: Two Logics

In this section I will review two different frameworks for reasoning about knowledge. In the interest of space, I will present only the basic ideas of each framework. The interested reader is referred to the relevant literature for more details.

2.1 History based models

History based models are often used to model distributed computing environments. The text [FHMV] (Chapter 4) has an extensive discussion about history based models. This section will present the framework of Parikh and Ramanjam found in [PR'85, PR].

Let E be a fixed set of events. Suppose that E^* is the set of finite strings over E and E^ω the set of infinite strings over E. Let $h, h', \ldots$ range over E^* and $H, H', \ldots$ range over $E^* \cup E^\omega$. Elements of $E^* \cup E^\omega$ will be called **histories**. Thus a history is just a string, or sequence of events.

Given two histories H' and H, write $H \preceq H'$ to mean H is a *finite* prefix of H'. Let hH denote the concatenation of finite history h with possibly infinite history H. Finally, let H_k denote the finite prefix of H of length k (given H is infinite or length $\geq k$). Given a set $\mathcal{H}$ of histories, define $\mathcal{P}(\mathcal{H}) = \{h \mid h \preceq H, H \in \mathcal{H}\}$. So $\mathcal{P}(\mathcal{H})$ is the set of finite prefixes of elements of $\mathcal{H}$.

A set $\mathcal{H} \subseteq E^* \cup E^\omega$ is called a **protocol**. Intuitively, the protocol is simply the set of possible histories that could arise in a particular situation. Following [PR'85, PR], no structure is placed on the set $\mathcal{H}$. I.e., the protocol can be *any* set of histories. Notice that no conditions are placed on $\mathcal{H}$. One natural condition that may be imposed is $\mathcal{P}(\mathcal{H}) \subseteq \mathcal{H}$, i.e., the protocol is closed under finite prefixes.

Let $\mathcal{A}$ be a finite set of agents. Given a particular finite global history H and an agent i, i will only "see" some of the events in H. This leads to a natural definition of agent uncertainty. For each $i \in \mathcal{A}$ define $\lambda_i : \mathcal{P}(E^\omega) \rightarrow E^*$ to be the **local view function** of agent i. $\lambda_i(H)$ is obtained by mapping each event "seen" by i into itself and each event not seen by i to the empty string[1]. Thus if h is the local history of agent i at some stage, and event e, which is visible to i, takes place next, then he will be the resulting history, otherwise the history remains h. Define for any subset $\mathcal{H} \subseteq E^\omega$,

$$\mathcal{H}_i := \{\lambda_i(h) \mid h \in \mathcal{P}(\mathcal{H})\}$$

[1] More formally, for each agent i, let E_i be the set of all events seen by agent i. Of course, for each i, $E_i \subseteq E$. Then $\lambda_i : \mathcal{P}(E^\omega) \rightarrow E_i^*$.

Thus $\mathcal{H}_i$ is the set of local histories of agent i.

Definition 2.1 *Let h, h' be finite global histories in some protocol $\mathcal{H}$. For each $i \in \mathcal{A}$, define $h \sim_i h'$ iff $\lambda_i(h) = \lambda_i(h')$.*

$\sim_i$ is clearly an equivalence relation. Given the dynamic nature of this model, it is natural to consider not only knowledge modalities, but also temporal modalities. Let Φ_0 be a countable set of propositional variables. A formula of multi-agent knowledge and time (denoted $\mathcal{L}_n^{KT}(\Phi_0)$) has the following syntactic form

$$\phi := p \mid \neg\phi \mid \phi \vee \phi \mid K_i\phi \mid \bigcirc \phi \mid \phi U \phi$$

where $p \in \Phi_0$ and $i \in \mathcal{A}$. As usual, $K_i\phi$ is intended to mean that i knows ϕ. $\bigcirc\phi$ is intended to mean ϕ is true after the next event and $\phi U \psi$ is intended to mean that ϕ is true until ψ becomes true.

Definition 2.1 *Suppose that $\mathcal{A} = \{1, \ldots, n\}$ is a set of agents. A* **history based model** *is a tuple $\mathcal{M}_H = \langle \mathcal{H}, \lambda_1, \ldots, \lambda_n, V \rangle$, where $V : \mathcal{P}(\mathcal{H}) \to 2^{\Phi_0}$, and each λ_i is a local view function of agent i.*

Truth is defined at finite histories. We assume the existence of a global discrete clock. Thus, for $H \in \mathcal{H}$, $H, k \models \phi$ is intended to mean that in history H at time k, ϕ is true. A primitive propositions p is true in history H at time k iff $p \in V(H_k)$. Truth of boolean connectives is as usual, and so we only give the definition of truth for the modalities:

1. $H, k \models \bigcirc\phi$ iff $H, k+1 \models \phi$
2. $H, k \models \phi U \psi$ iff there exists $m \geq k$ such that $H, m \models \psi$ and for all $l : k \leq l < m$, $H, l \models \phi$
3. $H, k \models K_i\phi$ iff for all $H' \in \mathcal{H}$ such that $H_k \sim_i H'_k$, $H', k \models \phi$.

In the above definition of truth of the K_i modality (item 3. above), it is assumed that the agents all share a global clock. This assumption is made in order to simplify the presentation. If the agents do not share a global clock, then item 3. should be replaced with the following definition:

3.' $H, k \models K_i\phi$ iff for all $m \geq 0$, for all $H' \in \mathcal{H}$ such that $H_k \sim_i H'_m$, $H'_m, m \models \phi$

It should be clear the K_i is an $S5$ operator for each i. A sound and complete axiomatization for knowledge and time can be found in [**?**], using a slightly different framework.

2.2 Topologic

In [MP], Moss and Parikh introduce a bimodal logic intended to formalize reasoning about points and sets. This new logic called *Topologic* can also be understood as an epistemic logic with an effort modality. Formally, the two modalities are: K and $\Diamond$. The intended interpretation of $K\phi$ is that ϕ is known; and the intended interpretation of $\Diamond\phi$ is that after some amount of effort ϕ becomes true. For example, the formula

$$\phi \to \Diamond K\phi$$

means that if ϕ is true, then after some "work", ϕ is known, i.e., if ϕ is true, then ϕ can be known with some effort. What exactly is meant by "effort" depends on the application. For example, we may think of effort as meaning taking a measurement, performing a calculation or observing a computation. In this paper we will think of effort as meaning consulting some agent's database of known formulas.

There is a temptation to think that the effort modality can be understood as (only) a temporal operator, reading $\Diamond\phi$ as "ϕ is true some time in the future". While there is a connection between the logics of knowledge and time and logics of knowledge and effort (see [H99, H00] and references therein for more on this topic), following [MP] it is assumed that such effort leaves the base facts about the world unchanged. In particular, in all topologics if ϕ does not contain any knowledge modalities, then $\phi \leftrightarrow \Box\phi$ is valid. Thus, effort will not change the base facts about the world – it can only change knowledge of these facts.

The family of logics introduced in [MP] and later studied by Dabrowski, Moss and Parikh, Georgatos, Heinemann, and Weiss ([DMP, G93, G94, G97, H99, WP]) has a semantics in which the acquisition of knowledge is explicitly represented. Familiar mathematical structures such as subset spaces, topologies, intersection spaces and complete lattices of subsets corresponding to natural notions of knowledge acquisition are attached to standard Kripke structures.

Given a set W, a subset space is a pair $\langle W, \mathcal{O}\rangle$, where $\mathcal{O}$ is a collection of subsets of W. A point $w \in W$ represents a complete observation about the world in which *all* facts are settled, whereas a set $U \in \mathcal{O}$ represents an observation. The pair (w, U), called a *neighborhood situation*, can be thought of as an actual situation together with an observation made about the situation. Formulas are interpreted at neighborhood situations. Thus the knowledge modality K represents movement within the current observation, while the effort modality $\Diamond$ represents a refining of the current observation.

Let Φ_0 be a finite or countable set of propositional variables, and $\mathcal{L}_0(\Phi_0)$ be the propositional language generated from Φ_0. A formula $\phi \in \mathcal{L}^{KE}(\Phi_0)$ has the following syntactic form

$$\phi ::= p \mid \phi \wedge \phi \mid \neg\phi \mid K\phi \mid \Diamond\phi$$

Definition 2.2 *An* **topologic model** *is a tuple $\langle W, \mathcal{O}, V\rangle$, where $\langle W, \mathcal{O}\rangle$ is a subset frame and $V : \Phi_0 \to 2^W$ is a valuation function.*

Formulas are interpreted at neighborhood situations.

1. $w, U \models p$ iff $w \in V(p)$
2. $w, U \models \neg\phi$ iff $w, U \not\models \phi$
3. $w, U \models \phi \wedge \psi$ iff $w, U \models \phi$ and $w, U \models \psi$
4. $w, U \models K\phi$ iff for all $v \in U$, $v, U \models \phi$
5. $w, U \models \Diamond\phi$ iff there is a $V \in \mathcal{O}$ such that $w \in V$ and $w, V \models \phi$

It is easy to see that K satisfies the axioms of $\mathbf{S5}$ and $\Diamond$ the axioms of $\mathbf{S4}$. To get a sound and complete axiomatization with respect to all subset spaces, the mix axiom needs to be assumed

$$\text{Mix} \qquad K\Box\phi \to \Box K\phi$$

[MP] provides a sound and complete axiomatization for all subset spaces. In [G93] and [G94], Georgatos provides a sound and complete axiomatization for subset spaces that are topological spaces and complete lattices. Dabrowski, Moss, and Parikh prove the same result using an embedding into $\mathbf{S4}$ ([DMP]). [G97] provides a sound and complete axiomatization for treelike spaces, and Weiss ([WP]) has provided a sound and complete axiomatization for intersection-spaces. Interestingly, it is shown in [WP] that an infinite number of axiom schemes are necessary for any complete axiomatization of intersection spaces. More recently, Heinemann [H99, H00] has looked at subset spaces and logics of knowledge and time, and the connection between hybrid logic and subset spaces [H02, H04].

3 Knowledge Based Obligations

In this section, a model of knowledge based obligation will be presented. The semantics is based on the history based models of the previous section. The examples and discussion are from [PPC].

a) Jill is a physician whose neighbour is ill. Jill does not know and has not been informed. Jill has no obligation (as yet) to treat the neighbour.

b) Jill is a physician whose neighbour Sam is ill. The neighbour's daughter Ann comes to Jill's house and tells her. Now Jill does have an obligation to treat Sam, or perhaps call in an ambulance or a specialist.

c) Mary is a patient in St. Gibson's hospital. Mary is having a heart attack. The caveat which applied in case a) does not apply here. The hospital has an obligation to be aware of Mary's condition at all times and to provide emergency treatment as appropriate.

d) Jill has a patient with a certain condition C who is in the St. Gibson hospital mentioned above. There are two drugs d and d' which can be used for C, but d has a better track record. Jill is about to inject the patient with d, but unknown to Jill, the patient is allergic to d and for this patient d' should be used. Nurse Rebecca is aware of the patient's allergy and also that Jill is about to administer d. It is then Rebecca's obligation to inform Jill and to suggest that drug d' be used in this case.

In all the cases we mentioned above, the issue of an obligation arises. This obligation is circumstantial in the sense that in other circumstances, the obligation might not apply. Moreover, the circumstances may not be fully known. In such a situation, there may still be enough information about the circumstances to decide on the proper course of action. If Sam is ill, Jill needs to know that he is ill, and the nature of the illness, but not where Sam went to school.

3.1 Actions and Values

We now sketch how to add a representation of social utilities and actions to the history based model. A complete discussion can be found in [PPC].

Assume a special set $Act \subseteq E$ of actions. Thus formally, the set of actions is just a distinguished subset of the set of all possible events. Let $a \in Act$ be an action and H be a finite global history. Now it may or may not be possible to perform a given the history H. Further, if a can be performed, then there is a set of possible extensions of H in which a has been performed. Formally, an action can be viewed as a partial function from the set of finite global histories to the set of all possible histories. Given a history H in which a can be performed, define

$$a(H) = \{H' \mid Ha \preceq H' \text{ and } H' \text{ a global history}\}$$

Given this definition it is straightforward to introduce a PDL style operator, $[a]\phi$ with the intended meaning 'after action a, ϕ is true, provided a *can* be performed.' Formally, given a history H in the domain of a,

$$H, t \models [a]\phi \text{ iff for all } H' \in a(H_t),\ H', t+1 \models \phi$$

Whereas the $\bigcirc$ modal operator is a linear time operator, i.e., it ranges over moments on a single global history, the dynamic modalities just defined are branching time operators.

We make the following assumptions about actions. These restrictions are not necessary, only made for the ease of exposition.

1. When an action is performed, it is performed at the next moment of time. We could weaken this assumption and assume that performing an action means performing that action eventually. In this case, $a(H)$ will be the set of global histories H' such that there is an $H_1 \in E^*$ and $HH_1a \preceq H'$. However, for now, we will use the above simpler definition of action performance.

2. The actions we consider are primitive, i.e., an action is just an element of the set of events E. One could develop a calculus of actions, where complicated actions are built from primitive actions using PDL (propositional dynamic logic) style operators. We refer the reader to Van Der Meyden [VM] for more on this topic. However, a large class of examples, including all the ones we consider in this paper, can be handled without adding the complications of a calculus of actions; and so we will not pursue this line of reasoning in this paper.

3. Each agent knows *when* it can perform an action. Thus if $H_t \sim_i H'_t$ and i can perform a at H_t then it can also perform a at H'_t.

4. Only one agent can perform some action at any moment. If no agents perform an action, then nature performs a 'clock tick'.

In order to define a notion a obligation, we define a "good" history. Formally, we assign a real number to each global history called its **value**. Under a natural assumption such as that the set of values is finite or compact, given any set of histories $\mathcal{H}$, there will be a subset of $\mathcal{H}$ which have the highest values. Given a history H, define $\mathcal{V}(H)$ to be the set of histories that extend H that have the highest value. A history $H' \in \mathcal{V}(H)$ is called an H-**good** history. Essentially, the models we consider are extensive games in which all agents are playing the same utility.

We can formalize the above discussion as follows. Let $\mathcal{H}$ be a set of global histories and $H \in \mathcal{H}$ a global history. For each $t \in \mathbf{N}$, let $\mathcal{F}(H_t) = \{H' \in \mathcal{H} \mid H_t \preceq H'\}$. That is, $\mathcal{F}(H_t)$ is the "fan" of global histories (in $\mathcal{H}$) that contain H_t as an initial segment. Recall that if $\mathcal{F}$ is any set of histories, $val[\mathcal{F}] = \{val(H) \mid H \in \mathcal{F}\}$. We require for each global history $H \in \mathcal{H}$,

1. For all $t \in \mathbf{N}$, $val[\mathcal{F}(H_t)]$ is a closed and bounded subset of $\mathbb{R}$.

2. $\cap_{t\in\mathbb{N}} val[\mathcal{F}(H_t)] = \{val(H)\}$

Condition 2 is a 'discounting' condition which ensures that values of histories depend only on what happens in a finite amount of time. If two histories agree for a long time then their values should be close.

Since $val[\mathcal{F}(H_t)]$ is closed and bounded for all t, there are maximal and minimal elements. Thus we define, $\mathcal{V}(H_t) = \{H' | H' \in \text{argmax}(val[\mathcal{F}(H_t)])\}$. Thus $\mathcal{V}(H_t)$ is the set of maximally good, (or just maximal) extensions of H_t.

Finally, we say that a *is good* to be performed at H, $\mathcal{G}(a, H)$, if $\mathcal{V}(H) \subseteq a(H)$, i.e., there are no H-good histories which do not involve the performing of a. And we say that a *may* be performed at H if $\mathcal{V}(H) \cap a(H)$ is non-empty.

We can now define knowledge based obligation.

Definition 3.1 (Knowledge based obligation) *Agent i is obliged to perform action a at global history H and time t iff a is an action which i (only) can perform, and i* knows *that it is good to perform a, i.e. $K_i(\mathcal{G}(a, H))$, or $(\forall H')(H \sim_{i,t} H' \wedge H' \in \mathcal{V}(H'_t) \rightarrow H' \in a(H'_t))$. I.e., putting this in terms of the agent's local history $h = \lambda_i(H_t)$, all maximal extensions of any H'_t with $\lambda_i(H'_t) = h$ belong to the range of the action a.*

We can formalize the above notion as follows. For each $a \in Act$, we define a primitive proposition $G(a)$. We say that $H, t \models G(a)$ iff all maximal global histories H' such that $H_t = H'_t$ are such that $H'_t a \preceq H'$. Then we say that i is obliged to perform action a at history H given time t if $K_i(G(a))$ is true at the pair H, t.

Suppose now that an agent acquires some knowledge. In that case, the set of global histories H such that $\lambda_i(H, t) = h$ will *decrease*, and the universal quantifier over all such histories will be more likely to become true. Thus before Jill was told of Sam's illness, the set of global histories compatible with her own local one included many where Sam was not ill. Receiving the information, however, deletes them, and in all global histories still compatible with her knowledge, she must act to help Sam. Similarly, in example b) Ann had an obligation to inform Jill, for before she tells Jill, in many of *Jill's* local histories compatible with Ann's, and in some global histories compatible with these latter, Ann's father is not ill and Jill cannot act. By informing Jill, Ann extends Jill's local history, and creates an obligation for Jill. Moreover, assuming that Ann knows that Jill does what she ought to, Ann herself has the obligation to inform Jill.

To see this more precisely we consider global histories consisting of four events, v, m, t, c where v stands for Sam vomiting, m stands for Ann telling Jill, t stands for Jill treating (or offering to treat) Sam and c is a clock tick which, unlike the other three, may occur more than once. Thus our global histories will consist of sequences in which events occur infinitely often, but v, m, t occur at most once. Moreover, since Ann is truthful, m never occurs without v occurring first. In those finite global histories in which v has occurred but not yet t, the best continuations are those in which t now occurs. And if v has not yet occurred then t (in the form of an offer to treat) may occur, but makes the history worse as the doctor is embarrassed by offering to treat a healthy man.

Thus we stipulate that all histories in which neither v nor t occurs have value 2, those in which t occurs without v have value 1 as do those in which v is followed by t. Finally those histories in which v occurs but not t have value 0 as they are the worst.

There are three agents, Sam, Ann, and the doctor, Jill. The event v is observed by Sam and Ann, m by Ann and Jill, and t, let us say, by all three. In a history in which v has occurred but not m, from Jill's point of view there are global histories in which v has not occurred which are compatible with her own local history. So she cannot know that it is good to treat Sam, although it is. She is not yet obligated to treat Sam. Once m occurs, she knows that v must have occurred, it is good to treat, and she knows it. So she is obligated.

Suppose again that v has occurred but not m yet. Then from Ann's point of view, Jill's local history is compatible with v not having occurred and in fact we will have $K_a(\neg K_j(V))$ (Ann knows that Jill does not know about the vomiting) where V denotes that vomiting has occurred. Since the vomiting *has* happened, all good histories now are those in which Sam has been treated, and those are included in the ones in which Ann has told Jill. So Ann ought to inform Jill about v, i.e. cause the event m, and then hope for t to take place. Ann has the obligation to tell Jill.

In a more complex scenario, with other agents, it could of course be that someone other than Ann had informed Jill of Sam's illness, but that Ann does not know this. We would say that Ann still has an obligation to inform Jill, and this can easily be expressed in our language.

Note that in our scenario, once the obligation to treat arises, it remains until treatment has taken place.

4 The Logic of Communication Graphs

The previous section's focus was how best to model the fact that an agent's choice of actions (especially obligatory actions) depends the agent's knowledge. In what follows, we shift the focus to the effect of communication on the agent's knowledge. In [PP] we begin by assuming that there is some constraints on who can talk to whom in a multi-agent setting. *Topologic* is then extended to the multi-agent setting, where a history based model is need to faithfully represent all the uncertainty the agents face in the situations considered.

5 Communication Graphs

Consider the current situation with Bush and Tenet[2]. If Bush wants some information from a particular CIA operative, say Bob, he must get this information through Tenet. Suppose that ϕ is a formula that representing the exact whereabouts of Bin Laden and that Bob is the CIA operative in charge of maintaining this information. In particular, $K_{\text{Bob}}\phi$, and suppose that at the moment, Bush does not know the exact whereabouts of Bin Laden ($\neg K_{\text{Bush}}\phi$). Obviously Bush can find out the exact whereabouts of Bin Laden ($\Diamond K_{\text{Bush}}\phi$) by going through the appropriate channels, but of course, *we* cannot find out such information ($\neg\Diamond K_{\text{e}}\phi \wedge \neg\Diamond K_{\text{r}}\phi$) since we do not have the appropriate security clearance. Presumably, going through the appropriate channels implies that as a *pre-requisite* for Bush learning ϕ, Tenet will also have come to know ϕ. We can represent this situation by the following formula:

$$\neg K_{\text{Bush}}\phi \wedge \Box(K_{\text{Bush}}\phi \rightarrow K_{\text{Tenet}}\phi)$$

where $\Box$ is the dual of diamond.

Let $\mathcal{A}$ be a set of agents. A **communication graph** is a directed graph $G_{\mathcal{A}} = (\mathcal{A}, E)$ where $E \subseteq \mathcal{A} \times \mathcal{A}$. We assume that E does not contain any reflexive arrows, i.e., for any communication graph $G_{\mathcal{A}} = (\mathcal{A}, E)$, for all $i \in \mathcal{A}$, $(i,i) \notin E$. Intuitively $(i,j) \in E$ means that i can directly receive information from agent j, *without* j knowing this fact. Thus an edge between i and j in the communication graph represents a one-sided relationship between i and j. Agent i has access to any piece of information that agent j knows. For example, during a lecture the students have access to the lecturer's information, but not vice versa. Another common situation that is helpful to keep in mind is accessing a website. When there is an edge between i and j we think of agent j as creating a website in which everything he *currently* knows is available, and agent i can access this website without j being aware that the site is being accessed. Of course, j may be able to access another agent's website and so update some of his information. Therefore, it is important to stress that when i accesses j's website, he is accessing j's current information.

The assumption that i can access all of j's information is a significant idealization from these common situations. This idealization rests on two assumptions: 1. all the agents share a common language, and 2. the agents make public all possible pieces of information. The fact that agents are assumed to share a common language is discussed in Section 5.2. For the second assumption, consider the tension between paparazzi and celebrities. This tension can be understood as the celebrities simply not wanting all of their current information made public. In other words, they want to remove, or

[2]Tenet was the current head of the CIA, but recently announced that he his stepping down sometime in July.

at least restrict, the connection in the communication graph from the paparazzi to themselves. This second assumption can be dealt with in our framework, but a more detailed discussion will be reserved for the full version.

Example: Suppose there are three agents $\mathcal{A} = \{1, 2, 3, 4\}$ and suppose that the communication graph $\mathcal{G}$ is the tree rooted at 1, where 1 has two children: 2 and 4, and 2 has only 3 as a child. Suppose that the initial partitions of the agents is given by the vector $\vec{\mathcal{P}} = (\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3, \mathcal{P}_4)$. Since neither 3 nor 4 are connected to any other agent, their initial information cannot change, i.e., in this situation they cannot learn any new information.

Since agent 1 is connected to all of the other agents, it is possible by asking enough questions, agent 1 can learn everything the other agents' know. However, since the only connection between agent 1 and agent 3 is through agent 2, any information agent 1 learns from agent 3, must first be learned by agent 2.

5.1 Language and Semantics

This section is a sketch of the semantics found in [PP]. The reader is referred to [PP] for a more detailed discussion.

We first define three languages. Let Φ_0 be a finite set of propositional variables. Let $\mathcal{L}_0$ be the set of base formulas, i.e. $\mathcal{L}_0$ is just a basic propositional language. Let $\mathcal{L}$ be the full language, i.e., $\phi \in \mathcal{L}$ iff it has the following syntactic form

$$\phi := p \mid \neg\psi \mid \phi \wedge \psi \mid K_i\phi \mid \Diamond\phi$$

Note that $\mathcal{L}$ is the multi-agent version of $\mathcal{L}_n^{KE}$. Finally, $\mathcal{L}_1$ is a sublanguage of $\mathcal{L}$ in which there are no embedded K_i operators.

Given a particular communication graph G, the uncertainty faced by the agents can be summarized as follows.

1. Agents are uncertain about the actual state of the world.
2. Agents are uncertain about which communication has taken place. This leads to uncertainty about the structure used to represent the information possessed by the agents, i.e., uncertainty about the Kripke structure. This uncertainty can be caused in two ways:
   (a) Agents are uncertain about the communication among agents not reachable via the communication graph.
   (b) Suppose that agent i learns some information P from agent j's website. Agent i is uncertain as to *how* agent j came to know P, i.e., what questions did j ask and to whom.

Define an **event** as a tuple (ϕ, i, j) to mean that i learns information ϕ from j, where ϕ is a ground formula (an element of $\mathcal{L}_0(\Phi_0)$). Of course there must be an edge between i and j in the communication graph. Formally given a communication graph $\mathcal{G} = (\mathcal{A}, E_{\mathcal{G}})$,

$$E = \{(\phi, i, j) \mid \phi \in \mathcal{L}_0, (i, j) \in E_{\mathcal{G}}\}$$

Given the set of events, we can then define a history based semantics as usual. A **history** is a finite sequence of events. I.e., $H \in E^*$ Given two histories H, H', say $H \preceq H'$ iff $H' = HH''$ for some history H'', i.e., H is an initial segment of H'. Given a history H, let $\lambda_i(H)$ be i's local history. I.e., this is a sequence of events that i can "see". Formally λ_i maps each event of the form (ϕ, i, j) to itself and other events to the null string. Then define

$$H \sim_i H' \text{ iff } \lambda_i(H) = \lambda_i(H')$$

Let W be a set of states. Assume that initially, nature informs each agent of the truth value of a particular set of propositional variables. This generates an initial vector of partitions (over the set of states W), say $\vec{\mathcal{P}}^0$.

Let H be a global history. We will say that w and H are compatible if H is a possible global history given w. That is given the assumption that agents are truthful implies that agent i can only learn ϕ from agent j if ϕ is actually true. Thus any global history H will be compatible with some subset of states.

Formulas will be interpreted at pairs (w, H) in where H is compatible with w. In the topologic setting, the neighborhood situation (w, U) is intended to represent a state and an observation. In this framework, an observation about a multi-agent social situation is a sequence of communications that have taken place. As usual, we assume that the extension of formulas in $\mathcal{L}_0$ is independent of the history.

The semantics of $\Diamond\phi$ is straightforward, i.e., $\Diamond\phi$ will be true at a pair (w, H) if there is an extension of H that makes ϕ true. Defining truth of the knowledge modality takes more work. What is needed is a map that takes a state and a history and returns a set of states (for each agent) that the agent considers possible given the uncertainty described above. Formally, $P(H, w)$ is a vector of sets where $P(H, w)_i$ is the set of states that i considers possible at state w given the compatible history H.

Given a state w, define $P(H, w) = (P_1, \ldots, P_n)$ recursively as follows:

1. $P(\epsilon, w) = (\vec{\mathcal{P}}^0_1(w), \ldots, \vec{\mathcal{P}}^0_n(w))$
2. Suppose that $H = He$, where e is the event (ϕ, i, j). Then $P(He, w) = (P(H, w)_1, \ldots, Q_2, \ldots, P(H, w)_n)$, where Q_2 is defined as follows:

   Let $Q_1 = P(H, w)_i$. Then Q_2 is the set of $v \in Q_1$ such that there is a H' i-compatible with H such that $P(H', v)_j \subseteq \hat{\phi}$

where $\hat{\phi} \subseteq W$ is the set of states in which ϕ is true. Note that since $\phi \in \mathcal{L}_0$, the extension of ϕ is independent of the particular history.

Truth will be defined at a state w and a history H such that H is a possible global history given w. Suppose that $V : \Phi_0 \to 2^W$ is a valuation function.

- $w, H \models p$ iff $w \in V(p)$
- $w, H \models \neg\phi$ iff $w, H \not\models \phi$
- $w, H \models \Diamond\phi$ iff $\exists H',\ H \preceq H'$ and $w, H' \models \phi$
- $w, H \models K_i\phi$ iff $\forall v \in P(H, w)_i,\ v, H \models \phi$

We abbreviate $\neg K_i \neg\phi$ and $\neg\Box\neg\phi$ for $L_i\phi$ and $\Diamond\phi$ respectively. We say ϕ is **valid** in $\mathcal{M}$ if for all $(w, \mathcal{P}),\ w, \mathcal{P} \models \phi$, denoted by $\models_{\mathcal{M}} \phi$.

Thus the formula $\Diamond K_i\phi$ is interpreted as "There is a sequence of communications that results in agent i knowing ϕ."

Axioms

1. All propositional tautologies
2. $(p \rightarrow \Box p) \wedge (\neg p \rightarrow \Box \neg p)$, for $p \in \Phi_0$.
3. $\Box$ satisfies the **S4** axioms and rules.
4. K_i satisfies the **S5** axioms and rules.
5. $K_i \Box \phi \rightarrow \Box K_i \phi$

These axioms and rules are shown to be sound and complete with respect to the set of all subset spaces ([MP]). Thus, they represent the core set of axioms and rules for any topologic. Soundness of axioms and the rules are easy to verify.

5.2 Connection with Communication Graphs

Let ϕ be a formula in our language $\mathcal{L}$, and consider the formula $K_i \phi \rightarrow \Diamond K_j \phi$. Intuitively, this formula says that if i knows ϕ then it is possible for agent j to know ϕ. One would expect that this formula will always be true provided that j is directly or indirectly connected to i in the communication graph. However, this does not quite work for *any* formula ϕ. For example, let ϕ be the formula $p \wedge \neg K_j p$, where $p \in \Phi_0$. Suppose that j is connected to i in some communication graph $\mathcal{G}$. It is easy to construct a model in which $K_i(p \wedge \neg K_j p)$ is true at some pair $(w, \vec{\mathcal{P}})$. However, no pair $(w, \vec{\mathcal{P}}')$ with $\vec{\mathcal{P}}' \preceq_{\mathcal{G}} \vec{\mathcal{P}}$ can satisfy the formula $K_j(p \wedge \neg K_j p)$, since K_j is an **S5** modal operator.

Nonetheless, there is a certain class of formulas for which the above statement will hold. Suppose that $\mathcal{M}$ is the multi-agent model described above.

Definition 5.1 *ϕ is* **stable in** *$\mathcal{M}$ iff $\phi \rightarrow \Box \phi$ is valid in $\mathcal{M}$.*

We say that ϕ is stable if ϕ is stable in all models. If ϕ is a ground formula, i.e., $\phi \in \mathcal{L}_0$, then ϕ is stable. This is easy to see, since using axiom 1 and 2, one can show that if $\phi \in \mathcal{L}_0$, then $\vdash \phi \leftrightarrow \Box \phi$.

The following results show that communication graphs are completely characterized by the set of valid formulas in models based on the particular communication graph. The proofs of these theorems can be found in [PP].

Lemma 5.1 *Let $\mathcal{G}$ be a communication graph and $\mathcal{M}$ a model generated by $\mathcal{G}$. If ϕ is stable in $\mathcal{M}$ and there is a path from j to i in the communication graph, then $K_i \phi \rightarrow \Diamond K_j \phi$ is valid in $\mathcal{M}$.*

In fact we can show something stronger, that the communication graph is characterized by formulas valid in models based on the graph.

Theorem 5.2 *Let $\mathcal{G} = (\mathcal{A}, E)$ be a communication graph. Then $(i, j) \in E$ if and only if, for all $l \in \mathcal{A}$ such that $l \neq i$ and $l \neq j$ and all stable ϕ, the scheme*

$$K_j \phi \wedge \neg K_l \phi \rightarrow \Diamond (K_i \phi \wedge \neg K_l \phi)$$

is valid in all models generated by $\mathcal{G}$.

6 Conclusions and Further Work

This paper surveys some recent papers that fit into a general project of developing formal tools for reasoning about social software. The frameworks sketched in this paper present the first steps in developing formal models appropriate for reasoning about social software. There is much more that can be done. The following lists some of the ongoing research.

1. Notice that in the knowledge based obligation framework, acquiring knowledge may create an obligation, but it cannot erase (a persistent) one. The existence of an obligation is a universally quantified formula whose truth value can only go from *false* to *true* as the domain shrinks. Dealing with this case will require a resort to the notion of a default history.

   To deal with such cases [PPC] introduces the notion of *weak knowledge* or more prosaically, *justified belief*. For each H_t, divide its extensions into two (there could be more than two) parts, the normal extensions (of which there must be some) and the unusual extensions. Now we say that ϕ is justifiably believed by i at H, t iff for all normal extensions H' of some H'_t which are i, t-equivalent to H_t, $H', t \models \phi$. Justified belief no longer implies truth as H itself might not be one of these normal extensions. More details can be found in [PPC].

2. Given a set of histories and values assigned to each history, we can ask, "Is it possible to program the agents in such a way that *if the agents do what the know they ought to do*, then one of the best histories is produced?

   We first must decide on how much computational power we will ascribe to the agents. Assuming that agents have perfect recall requires that they have unbounded memory, and we will need to model them as Turing machines whereas assuming that agents are finite automata means that agents have bounded memory.

3. In [PP], the type of communication studied is very specific. Agents are assumed to have a database of known facts which they can make public. The other agents can then access this database and learn any fact which is currently available in that database. It would be interesting to see if the update procedure can be extended to conscious communication, i.e., communication in which *both* agents are aware of the update. Along the same line, we could consider different types of updates, such as lying, updating to subgroups and so on.

4. We remark that the logic of communication graphs can be seen as a demonstration for the need for cryptographic protocols. Two issues are important here. This first is that an agent may only want part of its knowledge base to be accessible by the public. This may be modeled in our framework by attaching to each agent a set of formulas that are in the public domain, and so when i is directly connected to j, i can only update by sets definable in the publicly accessible language. The second issue is that we may not know the exact structure of the communication graph. For example, if Ann accesses some information from Bob's website, but unknown to Ann, Charles is listening in, then the communication graph does not have an edge between Ann and Bob, but only a path from Ann to Bob going through Charles. Then clearly as a condition for Ann learning some information from Bob, Charles must become informed of that same piece of information. Thus cryptographic protocols essentially ensure that there are direct edges between agents in the communication graph.

5. Another interesting application of the logic of communication graphs is the epistemic foundations of solution concepts. We would like to apply the communication graph framework to

backwards induction and other solution concepts. Thus illustrating the relationship between strategic choices in a game-theoretic situation and communication among the agents.

References

[BM] Baltag, A. and Moss, L., Logics for Epistemic Programs, to appear in the *Knowledge, Rationality, and Action* section of *Synthese*.

[BPX] Belnap, N., Perloff, M., and Xu, M., *Facing the Future*, Oxford 2001.

[DMP] Dabrowski, A, Moss, L, and Parikh, R. Topolgical reasoning and the logic of knowledge. *Annals of Pure and Applied Logic*, **78**, (1996), pp. 73 - 110.

[G93] Georgatos, K, Modal Logics for Topological Spaces. PhD Dissertation. Graduate School and University Center. City University of New York, 1993.

[G94] Georgatos, K, Knowledge Theoretic Properties of Topological Spaces. In *Knowledge Representation and Uncertainty*. M. Masuch and L. Polos, Eds. Lecture Notes in Artificial Intelligence, vol. 808, pages 147-159, Springer-Verlag, 1994.

[G97] Georgatos, K, Knowledge on Treelike Spaces. *Studia Logica*, **59**, (1997), pp. 271 - 231.

[Ge] Gerbrandy, J., *Bisimulations on Planet Kripke*, Ph.D. dissertation, University of Amsterdam, 1999.

[H03] Halpern, J., A computer scientist looks at game theory, *Games and Economic Behavior*, **45:1**, 2003, pp. 114-131.

[FHMV] Fagin, Halpern, Moses, and Vardi, *Reasoning about knowledge*, MIT Press 1995.

[HV] Halpern, J., and M. Vardi, The complexity of reasoning about knowledge and time, in *J. Computer and System Sciences*, **38** 195-237, 1989.

[H99] Heinemann, B., Temporal Aspects of the Modal Logic of Subset Spaces, *Theoretical Computer Science*, **224(1-2)**:135-155, 1999.

[H00] Heinemann, B., Extending Topological Nexttime Logic. In S. D. Goodwin, A. Trudel, editors, *Temporal Representation and Reasoning*, TIME-00, Cape Breton, Nova Scotia, Canada, pages 87-94, IEEE Computer Society Press, Los Alamitos, CA, 2000.

[H02] Heinemann, B., A Hybrid Treatment of Evolutionary Sets. In C. A. Coello Coello, A. de Albornoz, L. E. Sucar, O. Cair Battistutti, editors, MICAI'2002: *Advances in Artificial Intelligence*, Mrida, Yucatn, Mexico. Volume 2313 of Lecture Notes in Artificial Intelligence, pages 204-213, Springer, Berlin, 2002.

[Hi] Hilpinen, R., Deontic Logic, in *Blackwell guide to philosophical Logic*, Ed. Lou Goble, Blackwell 2001, 159-182.

[Ho'01] Horty, J., *Agency and Deontic Logic*, Oxford 2001.

[HMV] Halpern, J., R. van der Meyden, and M. Vardi, Complete axiomatizations for reasoning about knowledge and time, *SIAM journal of computing*.

[K] Kooi, B., *Knowledge, Chance, and Change*, Ph.D. dissertation, University of Groningen, 2003.

[H04] Heinemann, B., A Hybrid Logic of Knowledge Supporting Topological Reasoning. In *Algebraic Methodology and Software Technology*, AMAST 2004, Stirling, United Kingdom. Lecture Notes in Computer Science, Springer, Berlin, 2004. *To appear*.

[LS] Lomuscio, A., and M. Sergot, Deontic interpreted systems, *Studia Logica*, **75** (2003) 63-92.

[MP] Moss, L. and Parikh, Topological Reasoning and the Logic of Knowledge, *TARK IV*, Ed. Y. Moses, Morgan Kaufmann, 1992.

[OR] Osborne, M. and Rubinstein, A. *A course in game theory*, MIT Press, 1999.

[P95] Parikh, R., Knowledge based computation (Extended abstract), in *Proceedings of AMAST-95*, Montreal, July 1995, Edited by Alagar and Nivat, Lecture Notes ib Computer Science no. 936, 127-42.

[P03] Parikh, R., Levels of knowledge, games, and group action, in *Research in Economics*, **57**, (2003) 267-281.

[PP] Pacuit, E., and R. Parikh, A Logic for communication graphs, to be presented at the *Association for Symbolic Logic* annual meeting in Pittsburgh, May 2004.

[PR'85] Parikh, R., and R. Ramanujam, Distributed processes and the logic of knowledge, in *Logic of Programs*, LNCS #193, Springer 1985, pp. 256-268.

[PR'03] Parikh, R., and R. Ramanujam, A Knowledge based Semantics of Messages, in *J. Logic, Language and Information*, **12**, (2003) 453-467.

[Pa] Parikh, R., Social Software, *Synthese*, **132: 3**, Sep 2002, pp. 187-211.

[PPC] Parikh, R., Pacuit, E. and Cogan, E., The logic of knowledge based obligation, to be presented at the *Society for Exact Philosophers* meeting in the University of Maryland, May 2004.

[PR] Parikh, R. and Ramanujam, R., A knowledge based semantics of messages, in *J. Logic, Language, and Information*, **12**, pp. 453 - 467, 2003.

[Pl] Plaza, J., Logics of public communications, *Proceedings, 4th International Symposium on Methodologies for Intelligent Systems*, 1989.

[Vd] van Ditmarsch, H., *Knowledge Games*, Ph.D. dissertation, University of Groningen, 2000.

[VM] van der Meyden, R. The Dynamic Logic of Permission, *Journal of Logic and Computation*, Vol 6, No. 3 pp. 465-479, 1996.

[V] Vickers, S. *Topology Via Logic*, Cambridge University Press. 1989.

[WP] Weiss, M. A. and Parikh, R., "Completeness of Certain Bimodal Logics of Subset Spaces", *Studia Logica*, **71:1**, pp. 1 - 30, 2002.