Common Knowledge: A Survey

MS-CIS-91-14
LINC LAB 195

Marilyn A. Walker

Department of Computer and Information Science
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA 19104-6389

February 1991

COMMON KNOWLEDGE:A Survey

Marilyn A. Walker
Department of Computer and Information Science
University of Pennsylvania
lyn@linc.cis.upenn.edu

ABSTRACT. This paper discusses the motivation behind common knowledge. Common knowledge has been argued to be necessary for joint action in general and for language use as a particular kind of joint action. However, this term has been broadly interpreted. Two major issues must be addressed: (1) What mental state corresponds to common knowledge, ie. is knowledge, belief or supposition the appropriate mental attitude? (2) What inference process allows agents to achieve common knowledge?.

Most generally, common knowledge is used to describe the knowledge that is evidenced in reflexive reasoning. The term has also been used to refer to facts or objects which are mutually salient. One of the main problems for a theory of common knowledge is whether knowledge is the appropriate mental attitude. It seems as though probabilistic beliefs might approximate the cognitive phenomenon of common knowledge more closely than knowledge.

The main problem with a usable notion of common knowledge is that inference must play a critical role in what becomes common knowledge. I discuss the nature of conversational inference. It has a number of properties that distinguish it from other inferential systems, such as being apparently abductive and probabilistic, but a precise characterization of it is an unsolved problem. I suggest that in cases where ensuring common knowledge really matters, participants in dialogue accomplish this by exploiting opportunities for redundancy in conversation.

This paper was written to satisfy the requirement of the second written preliminary exam, whose focus was the following references:

References

[Bar88] Jon Barwise. *The situation in Logic-IV: On the Model Theory of Common Knowledge.* Technical Report No. 122, CSLI, 1988.

[CM81] Herbert H. Clark and Catherine R. Marshall. Definite reference and mutual knowledge. In Aravind K. Joshi, Bonnie Lynn Webber, and Ivan Sag, editors, *Elements of Discourse Understanding*, pages 10–63, Cambridge University Press, Cambridge, 1981.

[Lew69] David Lewis. *Convention.* Harvard University Press, 1969.

[Sch72] Stephen R. Schiffer. *Meaning.* Clarendon Press, 1972.

[SW82] Dan Sperber and Deidre Wilson. Mutual knowledge and relevance in theories of comprehension. In Neil V. Smith, editor, *Mutual Knowledge*, pages 61–87, Academic Press, New York, New York, 1982.

Contents

Common Knowledge

Marilyn A. Walker
lyn@linc.cis.upenn.edu

1 Introduction

1.1 Overview

What any fool would know, given a certain situation, has often been called COMMON KNOWLEDGE. It encompasses what is relevant, agreed upon, established by precedent, assumed, being attended to, salient, or in the conversational record[CM81, Pri81]. For instance if you and I are both normal humans sitting at a table with a candle on it, we may assume it is common knowledge that *There is a candle on the table.* If you told me, three utterances back in the conversation that your father was a college cha-cha champion, we may assume that it is common knowledge that *Your father was college cha-cha champion.* The convention of driving on the right in the U.S. and on the left in Great Britain is common knowledge established by precedent. If you and I have both agreed that *Blueberries should be consumed in great quantities when they are in season*, we may take it as common knowledge between us that *It is a good thing to buy blueberries when they are 2 for a dollar.* Our agreement provides evidence for the inference of the common knowledge fact. Thus knowledge as it has been used in this area often loosely refers to information that isn't actually **known** but rather assumed, supposed, or inferred[Sta78, CM81].

Common knowledge is strictly a generalization of the term mutual knowledge, where common knowledge is what is known within a group and mutual knowledge is limited to shared knowledge between two individuals, but the terms are often used interchangeably. Throughout this paper, I will use common knowledge, henceforth CK, even if the original author used the term mutual knowledge.

CK is more than just that two people know the same fact. In order for it to be CK, they must know they both know it, and know they know they both know it, etc. But how can agents distinguish between it being a coincidence that they both know a lot of the same facts, and that they have CK of a body of facts? Most definitions of CK include a number of conditions that must be met before CK is achieved. Two major issues must be addressed: (1) What mental state corresponds to CK, ie. is knowledge, belief or supposition the appropriate mental attitude?, (2) What inference process allows agents to achieve CK?.

CK is closely related to mutual belief, mutual expectations, and mutual intention[CC82, Lew69, Pow84]. The difference between CK and all these depends in part on distinguishing knowledge from action, as well as belief from knowledge. CK has been used to refer to two different kinds of beliefs, reflexive beliefs about what one another will do, and reflexive beliefs about what one another knows, supposes or believes. The characterization of CK depends on what inference processes are active, to what extent it is desirable to characterize the agents involved as being ideal reasoners,

the role of implicit vs. explicit knowledge and the effect of such vague notions as 'awareness' and 'attention'. Of particular interest throughout the paper is what is taken as good evidence of CK and what assumptions a particular CK fact is based on. There seem to be differences in degrees of CK; some facts are known, but others are known and salient.

Why should we be interested in CK? First it is claimed to be necessary for coordinated joint action between two or more people, or between two or more processors in a distributed system. Second, it plays a critical role in language both for the conventional meaning of utterances, and in conversational inference. This paper will primarily examine the second issue, the role of CK in language use[1].

1.2 The structure of this paper

First I want to justify the need for CK. CK is claimed to play a critical role in coordinated action[HM84, CC82]. I will briefly discuss its importance for action, then it's relation to conventional meaning. Both Lewis and Schiffer have claimed that common knowledge is an essential component of how it is that language can mean anything at all[Lew69, Sch72]. Lewis' idea is that language is a type of coordinated joint action. Schiffer's program is to show that Grice's definition of meaning needs to include a notion of CK in order to explain the role of the actual utterance in communication[Gri57, Sch72]. A particular subcase of conventional meaning is how referring expressions in language actually pick out the intended referent, and this has been studied in detail by Clark and Marshall[CM81]. The claim that CK is critical for reference is based on the fact that if I say to you *John is coming tonight*, you decide who the referent of *John* is, based on someone that you and I mutually know, and not just on all the possible people that John might designate. Clark and Marshall develop the relationship between definite reference and common knowledge as part of Clark's program of showing the role of collaboration in language use. I review the motivations of Lewis, Schiffer, and Clark and Marshall, in section 2, with respect to definite reference(section 2.2), convention(section 2.3) and meaning (section 2.4).

Second, before we can discuss in any detail the various formulations of CK, I need to give the definitions. There are three main definitions in the literature, the ITERATE approach, the FIXED-POINT approach, and the SHARED-ENVIRONMENT approach, and these are given section 3.

Third, there are two known problems with these definitions. The first problem is that under some definitions, CK is both mathematically suspect and psychologically implausible(section 4.1). The second problem is that it is unclear exactly what mental state corresponds to CK, and indeed whether knowledge is the right description of the phenomenon(section 4.2).

[1]There is a great deal published on Common Knowledge in the Distributed Systems literature. This community uses both the iterated definition and the fixed point definition. This work focusses on issues such as the effect of asynchronous communication, fault free vs. faulty channels, and message delivery times on the achievement of CK. Many of these issues have parallels in human communication, ie. Cohen [Coh84] and Oviatt [OC89] have show that asynchronous dialogues have much more elaborate referring expressions than synchronous ones, and Krauss and Bricker [KB67] show that delays of even .25-1.8 of a second can disrupt human dialogue and decrease referential efficiency. However it is beyond the scope of this paper to review that literature. Nevertheless I would like to note some significant theorems. In a distributed system, with asynchronous messaging, it can be shown that it is impossible to achieve CK if communication is not guaranteed. Furthermore, even if communication is guaranteed, if two processes don't share the same clock, CK cannot be achieved in finite time. CK can only be obtained in the general case if one is willing to allow the system to be in an inconsistent state for short periods of time. This is achieved by using an eager protocol, which broadcasts that p is CK. This of course is not strictly true at the time of the broadcast, but will be true within the maximum message delivery time[HM84].

Fourth, I will present Barwise's model theory for CK. The reason I wait until now to do it, is that Barwise addresses some of the known problems that I discuss in the previous section. First, he tackles the problem of some of the definitions being nonwellfounded by arguing that reality is not wellfounded anyway. He uses Aczel's theory of nonwellfounded sets to give a mathematical basis to the problematic definitions[Acz88]. Secondly, Barwise compares the three definitions given in section 3. Up to this point, I have not addressed the question of whether in fact the three definitions might be equivalent. Barwise argues that they are not equivalent, and commits himself to the idea that the fixed point representation is the 'right' one(section 5).

Fifth, with the mathematics out of the way, the various bases for CK that have been proposed are described. The issues I am concerned with in this section are whether CK is structured so as to facilitate memory access and its use in reasoning processes for comprehension of language. Is CK structured by the evidence which supports it, as Clark and Marshall argue? Are some CK facts held with more certainty than others? What happens when an agent receives new information?

Sixth, I discuss Sperber and Wilson's theory of Relevance. The reason for delaying the discussion to this point is that Sperber and Wilson reject the notion of CK altogether and develop an approach to communication based on a single PRINCIPLE OF RELEVANCE. They say:

> ...the only cases where a genuine effort is made to establish CK of the meaning, reference and implications of texts are legal documents and treatises, where the risk involved in misunderstanding is so great that the cost of reducing it is acceptable. .... the formal argument that CK is a necessary condition for comprehension applies only to perfect comprehension and not to the imperfect form which is felt to be quite sufficient in daily life.

Their approach is described separately because in the main, it is not comparable with the other approaches. Not only do they feel that CK is unnecessary, but they argue that language is a one-sided communicative process. They emphasize the inferential nature of language, an aspect only lightly touched on in the other papers(section 7). I give arguments against the relevance of Relevance in section 8.

Finally, in section 9, the problems with the establishment and maintenance of common knowledge in an ongoing conversation are explored. I will examine proposals about what distinguishes conversational inference from other inferential processes[Lev85]. I will briefly talk about how the common ground changes from one utterance to the next and some problems with this[Sta78, Pri78]. Then I suggest two factors that may play a role in establishing CK in conversation.

2 Motivation for Common Knowledge

2.1 Role in Coordinated Action

Coordinated action is taken to be any type of action that requires the participation and cooperation of multiple agents. Coordination games are one such type of action, but are typified by the inability of the players to communicate and reach an agreement. The natural interaction of two people normally allows communication freely, but it can be shown that their joint actions still require either CK or belief[CC82, Pow84]. These actions can only achieve the goal to which they are directed when two or more people intentionally coordinate their individual actions. Shaking hands

is a good example, and so are commonplace activities such as rowing a boat, dancing, or lifting a piano. Conversation is arguably a joint act as well, since my goal to communicate can't be satisfied by my actions alone; I need your understanding to achieve my goal. Why CK is proposed to play a crucial role in all these types of action is the subject of this section. The reader who is already familiar with this motivation can skip this section.

CK was first proposed as an important component of the strategic decisions that a 'player' might make when participating in a Coordination Game[Sch60]. Coordination games are at the opposite end of the scale from games of conflict. In conflict situations one player wins only if the other loses. In a coordination game, the players win by cooperating with one another. Coordination games demonstrate the power of reflexive reasoning. They are situations of interdependent decision by two or more agents whose self interest coincides with each other's, so that relative to some classification of actions, the agents have a common interest in all doing the same one of several alternative actions([Lew69], p. 24).

As an example, consider two people who parachute unexpectedly into the area shown in Figure 1(From Schelling [Sch60]). They both have maps and know that the other has one, but neither knows where the other has dropped and they have no means to communicate directly. Their goal is to meet as soon as possible.

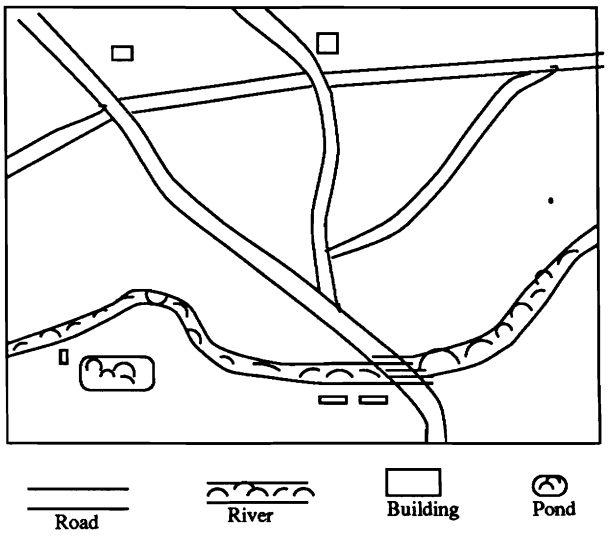

Figure 1: Parachutists Map

Another such puzzle is to tell a group of people to put a check mark in one of the sixteen squares in Figure 2. You win if you all put your marks in the same square.

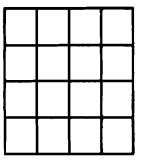

Figure 2: Sixteen Squares

A player reasons that *I want to go wherever the other player(s) expects me to go, but he will go*

whereever he expects me to go, which will of course depend on where I think he will go, ... The other player(s) reasons reciprocally. This reflexive reasoning process typifies a coordination game and is paradigmatic of CK.

Schelling tried a number of puzzles like this on a sample of respondents and found that people really do coordinate and are much better at these games than what random probability would predict. Schelling claims that the reason coordination is possible is that most situations provide some focal point for a concerted choice, ie. there is some notion of salience that participants recognize. These clues for coordination depend on some rational for each person's expectation of what the other expects him to expect to be expected to do. He suggests that finding a key may depend on imagination, analogy, precedent, accidental arrangement, symmetry, aesthetic or geometric configuration, ethical reasoning, and who the parties are and what they know about each other ([Sch60], p. 55)[2].

One thing that unifies all the coordination game examples is that it is not possible for the players to communicate with one another and thus agree on what would be the most beneficial course of action. This doesn't mean that the only interesting cases of coordination operate when communication isn't possible. As the following examples will show, even when communication is possible, the achievement of CK is not a trivial problem.

Consider Clark and Carlson's example of a duet between Itshak Perlman and Pinchas Zuckerman. The whole duet is an example of an adjustable joint act, but as Clark and Carlson note, the joint act of initiating the first note is complicated enough.

> Imagine that Perlman has been practicing his gesture to start the first note and now he wants it to be taken for real. When he gestures this time he must believe that Zuckerman will take the gesture for real, otherwise Zuckerman won't play the first note and their joint act will fail. But Zuckerman won't play if he believes that Perlman believes himself to be still practicing since in that case Perlman won't play, and Perlman won't play unless he believes that Zuckerman believes that he Perlman, believes that Zuckerman believes that this time he is gesturing for real.

In principle, this process of reflexive reasoning could continue ad infinitum.

Another case of the effect of CK on the reasoning processes of two interacting agents is an example of what is known as the Conway paradox. Consider two players, Ellen and Tom, in a game of poker. Suppose that each of them gets an ace. Thus each of them knows:

> Either Ellen or Tom has an ace.

If someone were to come along and ask each of them whether they knew if the other had an ace, they would of course reply *no*. If someone were to say to them *At least one of you has an ace. Do you know whether the other has an ace?*, they would still answer *no*. But then, further reasoning is possible in a way that it was not possible before, even though the information *At least one of you has an ace* was already known to both of them. Upon hearing Tom say *no*, Ellen may reason

[2]In Schelling's sample, 7 out of 8 of the parachutists managed to meet at the bridge. In the squares problem the upper left corner received 24 votes out of 41, and all but three of the remainder were distributed in the same diagonal line.

that if Tom doesn't know whether I have an ace, after hearing that one of us does, it must be because he has an ace. And Tom can reason reciprocally. So what was said must have added some information. This added bit of information must be the COMMON KNOWLEDGE that one or the other has an ace.

2.2 Definite Reference and Common Knowledge

Clark and Marshall develop the relationship between definite reference and common knowledge[3]. Consider their scenario:

> Version 1: On Wednesday morning Ann reads the early edition of the newspaper which says that *Monkey Business* is playing that night at the Roxy. Later she sees Bob and asks, *Have you ever seen the movie showing at the Roxy tonight?*

The question they ask is what facts does Ann have to assure herself of in order to make felicitous use of the definite referring term t, *the movie showing at the Roxy tonight*, to refer to the real world referent R, the movie MONKEY BUSINESS. Obviously Ann herself must know that t describes a unique referent, that there aren't two movies showing at the Roxy tonight. That is Ann must be certain that on uttering her reference the following condition will be true:

> Condition(1):Ann knows that t is R.

But what if Ann didn't think that Bob knew what movie was playing tonight? Then she couldn't refer to *Monkey Business* by *the movie showing at the Roxy tonight.* Thus Ann must engage in a certain amount of reflexive reasoning in order to be sure of:

> Condition(2):Ann knows that Bob knows that t is R.

This might seem like it should be enough but consider another version of the scenario.

> Version3: On Wednesday morning Ann and Bob read the early version of the newspaper and they discuss the fact that *A Day at the Races* is showing that night at the Roxy. When the late edition of the paper comes out, Bob reads the movie section and notes that the film has been corrected to *Monkey Business*, and circles it with his red pen. Later Ann picks up the late edition, notes the correction and recognizes Bob's circle around it. She also realizes that Bob has no way of knowing that she has seen the late edition. Later that day Ann sees Bob and asks *Have you ever seen the movie showing at the Roxy tonight?*

This scenario satisfies conditions (1) and (2) but Bob is very likely to take Ann's reference R, to be *A Day at the Races.* Thus Ann must reason about Bob's reflexive reasoning and thus must satisfy:

> Condition(3):Ann knows that Bob knows that Ann knows that t is R.

[3]Clark and Marshall do not consider attributive uses of definite referring expressions.

This is called shared$_3$ knowledge, due to the 3 levels of nesting in the knowledge statements. Clark and Marshall develop this hierarchy of facts further and show that in principle ever and ever more complicated scenarios could be devised that would lead to more conditions, hence to an infinity of such conditions, ie. to shared$_\infty$ knowledge. But shared$_\infty$ knowledge is exactly what CK is. The claim is that what Ann and Bob need to be sure of is their CK of the fact that t is R[4].

This leads Clark and Marshall to formulate the DIRECT DEFINITE REFERENCE CONVENTION(p 26). The speaker sincerely intends to refer, by using a term t, to (1) the totality of objects or mass within a set of objects such that (2) the speaker has good reason to believe, (3) that on this occasion the listener can readily infer (4) uniquely (5) common knowledge of the identity of that set (6) such that the intended objects or mass in the set fit the descriptive predicates in t, or if t is a rigid designator, are designated by t.

Thus felicitous reference depends on the speaker and hearer establishing certain kinds of CK; in principle no finite level of shared knowledge is enough. This isn't surprising since definite reference is, according to Clark and Marshall, a perfect example of something that speakers and listeners achieve through coordination(cf. Clark and Wilkes-Gibbes[CW86]).

Clark and Marshall develop a taxonomy of the different ways that someone might come to know something, and then show that this taxonomy has an effect on the different kinds of referring expressions that can be produced as well as how speakers can repair failed references by increasing the strength of the description. These increases in strength depend on reducing the number of assumptions that have to be made in order to achieve CK, so that their definite reference statement (2) would be that the speaker has an even better reason to believe that the listener can pick out the referent. I will develop this perspective in more detail in section 6, after I have provided the definition for CK that Clark and Marshall's analysis is based on.

2.3 Convention

Lewis suggests that language arises out of CONVENTION. Convention is defined as a regularity in the behavior of members of a population that they maintain because they **mutually know** that they have maintained it in the past and that it has solved for them a recurring kind of coordination problem.

Conventions are dependent on a reflexive reasoning process which derives MUTUAL EXPECTATIONS. First he defines:

> HIGHER ORDER EXPECTATIONS: A first order expectation about something is an ordinary expectation about it. An (n +1)th-order expectation about something is an ordinary expectation about someone else's nth order expectation about it.

An ordinary expectation would be something like *You expect that I will go there.* A second order expectation would be *I expect that you expect that I will go there.* The replication of another's

[4]Webber shows that in fact the listener need not really believe or know the description holds and Perrault and Cohen show that the speaker need not at some small finite level. What matters is the participants beliefs about what the other believes, not their own beliefs. Rather than providing evidence against common knowledge, this actually supports the notion of common knowledge, since the common knowledge facts past that finite level must hold[Web78, PC81].

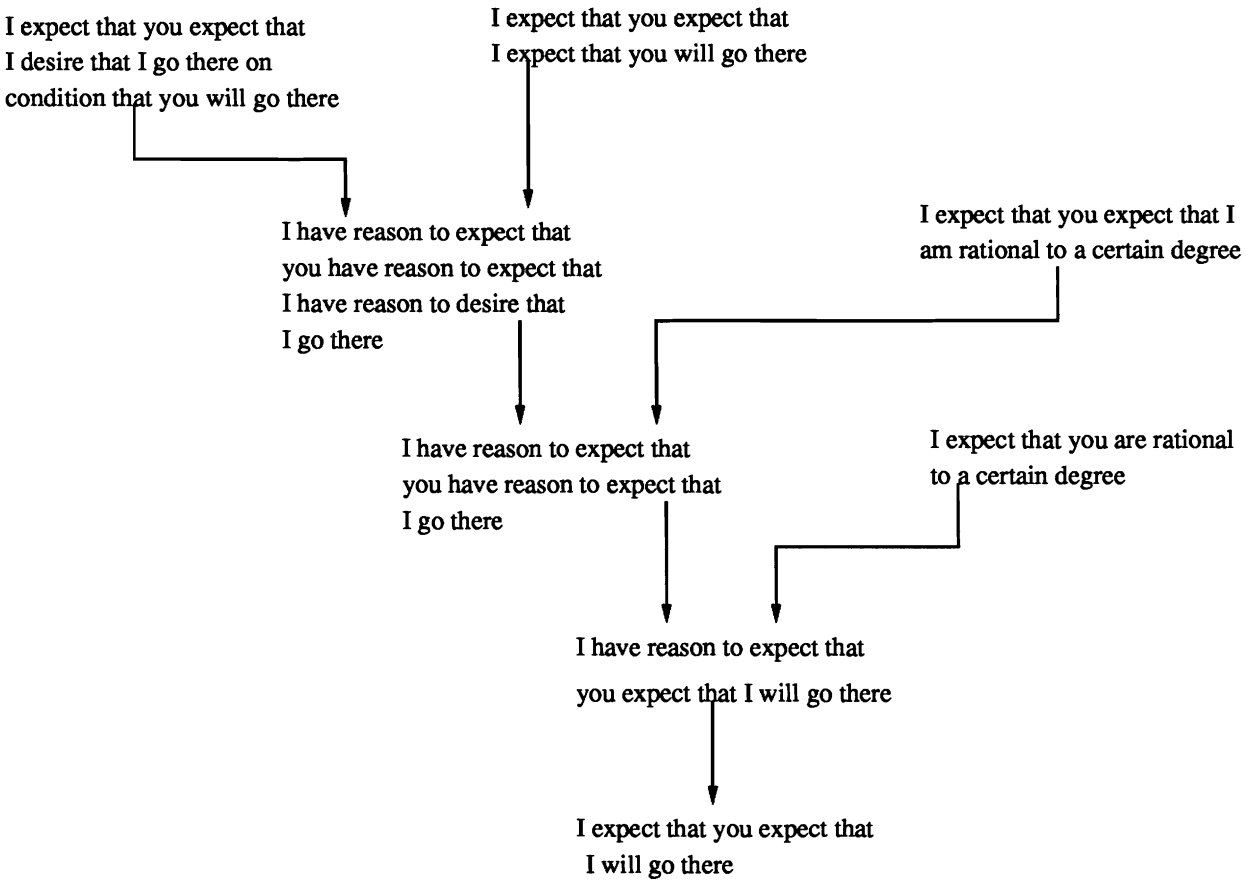

Figure 3: Mutual Expectations and Rationality

practical reasoning, along with second-order expectations about matters of fact, and first-order expectations about their preferences and rationality, justify the formation of a first-order expectation about their action. In the case of problems of interdependent decision, e.g. coordination problems, some of the necessary second-order expectations must be about the reasoner's own actions(See Figure 3 from Lewis[Lew69]).

Sometimes conventions can come about by way of some kind of natural meaning. Suppose that one time at a party, a husband wiggles his ears and his wife thereby infers that he is bored. If she tells him *I could tell you were bored because you were wiggling your ears*, then after that, he can actually signal *I am bored* by wiggling his ears. Thus ear wiggling has become a convention between the two of them, meaning *I am bored.* What is important to note here is the arbitrariness of the signal with respect to its meaning. The agreement on meaning or CK is what matters.

Lewis distinguishes between a particular language that a population actually uses and what are possible languages, claiming that languages that are actually used become so by convention. There are many possible languages and our choice of a particular one only depends on all of us using it so that we may communicate by it. Thus a particularly important kind of convention is one by which certain actions come to serve as signals. To explain how language is developed, Lewis defines a two-sided signalling problem(p 130).

An example of a two-sided signalling problem is that which held between the sexton of Old North Church and Paul Revere. The sexton's contingency plan is that:

- If the redcoats stay at home, hang no lantern in the belfry.
- If the redcoats go by land, hang one lantern in the belfry.
- If the redcoats go by sea, hang two lanterns in the belfry.

Paul Revere acts according to a contingency plan that:

- If there is no lantern in the belfry, go home.
- If there is one lantern, warn the countryside that the redcoats are coming by land.
- If there are two lanterns, warn the countryside that the redcoats are coming by sea.

More formally, a two-sided signalling problem is a situation S involving an agent called the communicator and one or more other agents called the audience, such that it is true that, and it is CK that

- One of several situations s_1, ...s_m holds. The communicator but not the audience is in a good position to tell which one it is.
- Audience members have a choice of responses r_1,...r_m and there is a one to one function from s_i onto r_j such that everyone prefers that each member of the audience do $F(s_i)$ on each condition that s_i holds for each s_i.
- The communicator can give any one of a set σ_k of signals and the audience can tell which one he gives.

A communicators contingency plan F_c, is any way in which his signal may depend on s_i. If F_c is one to one, then it is admissible. Similarly an audience's contingency plan F_a, is any way in which their response depends on s_i. If the range of F_a coincides with the range of F, the preference function, then F_a is admissible. Lewis proves that signalling systems must be composed out of admissible contingency plans, and thus that every signaling system is a solution to a particular coordination problem.

However, although this captures a nice intuition about the core meaning of language, the sexton can only signal one of a certain preset number of signals and Paul Revere can only respond with one of a certain number of preset responses. This conventional signalling was presumably set up by agreement. The sexton has no way to signal to Revere if it should happen that the British split up into two groups, some going by sea and some by land. What could Revere take him to mean if he alternately put up two lanterns, then one, then two, then one?

Lewis acknowledges that of course the idea of language as a convention of two-sided signalling is very rudimentary, since by the definitions given, any possible language $\mathcal{L}$ is limited by such facts as:

- There is only a closed finite set of sentences of $\mathcal{L}$.
- Sentences are reserved for use in a particular activity.
- Users of $\mathcal{L}$ have no choice about **how** to say something.

- There is no ambiguity or indexicality.

So it is obvious that a two-sided signalling system doesn't actually characterize natural language. The main problem is that there is a one to one mapping of language and situations. In order to extend the idea of a language to remedy these faults, the definition of interpretation must be changed so that it considers the occasion of use as a parameter of meaning (p 163). The occasion of use provides a context in which to interpret the utterance, and it is at this point that a second level of CK becomes important for determining the meaning of an utterance. I will discuss Lewis's ideas on context in section 6.

2.4 Meaning

Schiffer is concerned with ruling out certain counter examples to Grice's definition of SPEAKER MEANING. First, consider Schiffer's version of Grice's definition(unmodified except that Schiffer believes he has made Grice more precise):

> S meant something by or in x iff S uttered x intending
>
> 1. that x has certain features f
> 2. that a certain audience A recognize (think) that x is f
> 3. that A infer at least in part from the fact that x is f that S uttered x intending (4)
> 4. that S's utterance of x produce a certain response r in A
> 5. that A's recognition of S's intention (4) shall function as at least part of A's reason for his response r.

Note the circular quality of the intentions here in (4) and (5) . S produces an utterance intending that A understand what he means by recognizing his intentions. Since one can know that S meant something without knowing what he meant, Grice suggests "that to ask what [S] meant is to ask for a specification of the intended effect"(p 385). But as Schiffer points out this can't be exactly right. Grice misses the whole notion of linguistic conventions. I cannot utter *The flamingoes are flying south this year* with the intention of getting you to pass the wine, unless that is something that has been set up as a particular signal between the two of us.

Schiffer's argument concerns showing that Grice's definition will only work, if the speaker and audience mutually know, among other things, the effects particular utterances are intended to produce. In addition, Schiffer's counterexamples show that it is always possible in principle to devise problematic scenarios for any finite level of shared knowledge, just as Clark and Marshall did. Schiffer's solution is to incorporate the notion of CK directly into the definition of speaker meaning.

> S meant that p by (or in) uttering x if S uttered x intending thereby to realize a certain state of affairs $\mathcal{E}$ which is (intended by S to be) such that the obtainment of $\mathcal{E}$ is sufficient for S and a certain audience A **mutually knowing* (or believing*)** that $\mathcal{E}$ obtains and that $\mathcal{E}$ is conclusive (very good or good) evidence that S uttered x with the primary intention

1. that there be some ρ such that S's utterance of x causes in A the activated belief that $p/\rho(t)$[5] ;
2. satisfaction of (1) to be achieved at least in part by virtues of A's belief that x is related in a certain way R to the belief that p;
3. to realize $\mathcal{E}$.

By Schiffer's definition, utterances realize states of affairs that provide evidence for the active inferences of an audience. This evidence has a conventional basis in CK. Schiffer includes the notion of an activated belief, rather than just any belief in order to account for cases of reminding, pointing out, etc, in which a speaker presumably tells the hearer something that they already know, and the actual saying of the utterance is the only thing required for the satisfaction of the speaker's intention. Thus Schiffer distinguishes between known facts and their accessibility or salience in terms of activation. For instance, reminders achieve their goal without any action by the hearer. For example:

(1) A: Now what was that girl's name?

B: Rose.

Schiffer also provides a parallel definition for action rather than belief. With these modifications, he claims that the definition of speaker meaning now captures the demand that meaning and communication be rational in a certain way. I say something a certain way, because I know by saying it that way that you are likely to understand me. Secondly, by including the truth supporting reasons as part of your reason for your activated belief, it reflects the fact that communication in general aims at the production of knowledge and not merely belief(p. 58, 63).

2.5 Summary

Clark and Marshall claimed that CK was necessary for understanding because a certain kind of reflexive reasoning is involved in the production of felicitous referring expressions. Lewis and Schiffer claim that this same reflexive reasoning process is necessary for the production of rational utterances in general. Clark and Marshall are using the notion of CK in a slightly broader sense than Lewis however, because they consider the role of context and the linguistic context in particular in determining CK, but both Lewis and Schiffer speak of utterances providing evidence or a basis for inferring CK. I will not consider Lewis's or Schiffer's proposals on meaning in any detail after this. However, I discuss the aspects of their proposals that import on how CK is maintained and used in conversation. In the next section, I present the formal definitions of CK.

3 Definitions of Common Knowledge

There are three main definitions of CK. I first present the definitions, then discuss some common problems with these definitions. In section 5, I present Barwise's model theory for each definition[Bar88]. Lewis's definition is called the SHARED-ENVIRONMENT approach and this is used

[5] p is the response with reason(s) ρ, and that these are truth supporting reasons is denoted by $\rho(t)$

by Clark and Marshall. Schiffer's main definition is called the ITERATE approach. An alternative approach, offered by Harman, is called the FIXED POINT definition[Har77].

I will use some simple facts in Barwise's notation in the definitions that follow. Barwise's prototypical situation is playing stud poker. In five card stud, each player receives one 'down' card and four 'up' cards. A player's down card cannot be seen by the other players, but everyone can see all the up cards of each player. The operator S in $(p_i S \phi)$ indicates that p_i sees that ϕ. The other relation Barwise uses is the relation H, which stands for a player having a card, e.g. $(p_2 H 3\clubsuit)$ if p_2 has the $3\clubsuit$. The S (sees) relation stands in for knowledge, and reflects the fact that physical presence is a particularly strong basis for knowledge

The comparison depends on a way to model facts and situations. For the moment just think of a situation as a certain state of affairs in the world, from which a set of facts can be derived. Fact is used in a neutral way, a fact need not be true in all situations. The facts and situations constructed out of the two relations, S and H, are the only ones considered. These situations represent the shared information of some situation s_u, e.g. the 'up' cards in a game of poker. For the purposes of this paper, Barwise represents situations by sets of facts, so one can interpret the constituent relation $\in$ between facts and situations as set membership without loss of understanding. Barwise also use $s \models \theta$ to denote θ is a fact of s. I will develop Barwise's approach more technically in section 5.

3.1 Iterate Definition

Schiffer was concerned with Grice's definition of speaker meaning. One aspect he was concerned with was the regressive series of intentions that Grice seemed to require speakers to have. Schiffer proposed replacing this with certain requirements of CK of utterance meaning[6]. Although the ITERATE definition below is the one that is commonly attributed to Schiffer, Schiffer developed another definition that is much closer to that of Lewis. I will discuss this in section 3.3.

In Schiffer's definition of CK he enumerates the conditions necessary for common knowledge between two agents p_1 and p_2:

$$\{(p_1 S s_u), (p_2 S s_u), (p_1 S (p_2 S s_u)), (p_2 S (p_1 S s_u)), ...$$

Here we use Barwise's S relation for 'sees', the shared situation is given by s_u and p_1 and p_2 are agents. The fact that p_1 and p_2 have CK of the 'up' cards in s_u is represented by an infinite number of distinct well-founded facts[7].

3.2 Fixed Point Definition

Harman suggested that CK might be explained as knowledge of a self-referential fact:

> A group of people have CK of q if each knows q and WE KNOW THIS

[6] Schiffer actually provides different definitions for CK and mutual knowledge. However his definition for CK allows a proposition to be mutually believed within a group without it being mutually believed between two members of the group([CC82], footnote 6.)

[7] McCarthy showed that it doesn't follow from the iterate definition that if it is CK amongst the members of G that p, then it is CK amongst the members of G that it is CK amongst the members of G that p[MSHI78].

The THIS refers to the whole fact known([Har77],p. 422). Although Harman seems to suggest that this is just a succinct way of representing Schiffer's iterated definition, it turns out to have different mathematical properties. Among these differences is one that Harman notes himself, which is that his formulation has the property that where there is common knowledge, it is known that there is CK[Bar88].

In Barwise's notation this would be represented by the shared situation s_c, where again s_u represents the shared information:

$$s_c = \{(p_1 S(s_u \bigcup s_c)), (p_2 S(s_u \bigcup s_c))\}$$

In contrast with Schiffer's iterate definition which contains an infinite number of facts, the fixed point account contains just two facts, but note that it is circular and hence not wellfounded.

3.3 Shared Environment Definition

Lewis's definition of CK relies on mutual expectations such as those in coordination problems(See section 2.3).

> It is common knowledge in a population P that Ψ if and only if some state of affairs $\mathcal{A}$ holds such that:
>
> 1. Everyone in P has reason to believe that $\mathcal{A}$ holds.
> 2. $\mathcal{A}$ indicates to everyone in P that everyone in P has reason to believe that $\mathcal{A}$ holds
> 3. $\mathcal{A}$ indicates to everyone in P that Ψ.

The state of affairs $\mathcal{A}$ is a BASIS for CK in a population P that Ψ. This state of affairs, $\mathcal{A}$, INDICATES to someone x that Ψ if and only if, if x had reason to believe that $\mathcal{A}$ held, x would thereby have reason to believe that Ψ. What $\mathcal{A}$ indicates to x will depend, therefore, on x's inductive standards and background information(p 52). $\mathcal{A}$ provides the members of P with part of what they need to form expectations of arbitrarily high order, with respect to other members of the same population, that Ψ. The part it gives them is the part peculiar to the content Ψ. The rest of what they need to form any higher order expectations is mutual ascription of some common inductive standards and background information, rationality, and mutual ascription of rationality. The inductive standards and rationality assumptions will be discussed further in section 4.2. Clark and Marshall use Lewis's definition in their work, calling it the COMMON KNOWLEDGE INDUCTION SCHEMA.

Using Barwise's notation and situation theory, the state of affairs $\mathcal{A}$ is a situation and common knowledge becomes knowledge between two agents. The supports relation $\models$ is used for INDICATES. Lewis' shared environment approach of CK is recast as:

- **Shared Situation**
  - $\mathcal{A} \models \Psi$
  - $\mathcal{A} \models p \ sees \ \mathcal{A}$
  - $\mathcal{A} \models q \ sees \ \mathcal{A}$

However, Lewis' definition cannot just be replaced with Barwise's. Barwise's focus on physical situations and conventionally signalled facts, such as a card being the Q♠, and being 'up', loses some of the generality of Lewis's definition. That is situations, as modeled, consist of sets of facts. A fact is either an element of a situation or not. There is no room in this analysis for a situation providing grounds, ie. good evidence for, or giving someone a reason to believe a particular fact. There is also no room in this analysis for the auxiliary assumptions that are characteristic of Lewis's approach, such as the inductive standards and background information which he makes use of in the definition of INDICATES.

As well as his iterate definition, Schiffer also developed a similar definition to Lewis's. His claim is that it is a truth about knowledge in general that, for any property H such as being a normal sighted person sitting at the table, and any proposition Ψ such as *There is a candle on the table*, if one knows that whoever is H knows that Ψ, then if one knows of any particular person that he is H, then one knows that he knows that Ψ. He notes that in each case of mutual knowledge there is a **finitely** describable situation such that in virtue of certain general features of the situation, it follows that two people have an infinite amount of knowledge about each other. He concludes that there will always be a set of conditions which are such that S and A will know that Ψ, just in case these conditions are satisfied.

Thus he redefines mutual knowledge as:

> S a speaker, and A an audience, mutually know* that Ψ iff there are properties F and G such that
>
> 1. S is F
> 2. A is G
> 3. both being F and being G are sufficient for knowing that Ψ, that S if F and that A is G
> 4. for any proposition σ, if both being F and being G are sufficient for knowing that σ, then both being F and being G are sufficient for knowing that both being F and being G are sufficient for knowing that σ.

Thus both Lewis and Schiffer abstract away from the proposition Ψ that is known, to the conditions under which someone would come to know a proposition. These conditions require certain assumptions regarding attention, structural homology, and capabilities for reflexive reasoning, plus properties that may depend on what particular proposition is being taken to be common knowledge.

I've offered three definitions for CK. I will look at their formal properties in section 5. First though, I will go on to look at the problems with these definitions and to discuss the precise characterization of CK. Is knowledge the right attitude or should it perhaps be belief or supposition?

4 Two issues for Theories of Common Knowledge

4.1 Infinity of conditions

Both Lewis and Harman foreshadow a debate about CK and its potential usefulness. Harman notes that:

If self-referential facts are disallowed, a technical problem arises – how to avoid saying that each person must know each of the following infinite regress of facts:

1. σ
2. we know (1)
3. we know (2)

   .....

On this interpretation, CK presents the psychologically implausible claim that humans evaluate an infinite set of facts. Clark and Marshall call this problem the COMMON KNOWLEDGE PARADOX, which may be briefly motivated by the fact that according to the definite reference convention in section 2.2, Ann must simultaneously check an infinity of conditions, like (1), (2) and (3), each check taking a finite amount of time, and yet apparently produce a referring expression in just a few seconds. Sperber and Wilson completely reject the notion of CK, with this problem as one of their main concerns([SW82] p 62). They point out that this infinite set of conditions should cause problems with comprehension, which don't actually seem to occur. They suggest that, in real life, if any such unnaturally complex situation arose as that found in Clark and Marshall's movie at the Roxy examples, either the hearer would ask for clarification or as likely as not, misunderstanding would occur (see section 7).

Lewis proposes that although in principle his schema will generate an infinite number of common knowledge statements, the constraining factor is actually our beliefs about one another's rationality (p 55-6). By Lewis's schema, one has reasons to have arbitrarily high order expectations, since each term in the sequence acts as a reason for the next higher term. Lewis also suggests that anyone who has reason to believe something will come to believe it, provided he has a sufficient degree of rationality. However he claims that the degrees of rationality we are required to have, to have reason to ascribe to others, etc. need complex levels of inferencing once they go beyond the first few orders of expectations. The result is that expectations of only the first few orders are actually formed. In fact he commits himself further to the notion of some kind of truncation heuristic by saying that the more orders we have the better, but we rarely do have higher order expectations than, say fourth(p 32).

Another approach to this problem is to propose that mutual absence of doubt is a better characterization of the actual phenomena. Rather then verifying indefinitely many beliefs, we check whether any doubts are present([Rev87], Davies p. 717, [Gri82]). Nadathur and Joshi proposed that both mutual belief and mutual knowledge are too strong for dialogue[NJ83]. *There is a notion weaker than mutual knowledge or mutual belief, that is operative in practice, and it is this that a system which reasons about knowledge and belief must give expression to.* They suggest that mutual beliefs are more accurately characterized as conversational conjectures based on absence of doubt.

The problem has led to claims that in fact there is no infinite hierarchy, that reflexive reasoning processes, or nested beliefs are only actually used up to a certain level, for instance that shared_4, or some earlier level is enough[BH79]. Both Schiffer and Clark and Marshall consider this proposal in some detail and reject it. Schiffer rejects it because whatever level is chosen as the topmost one, "it is always possible to imagine two people a little bit smarter and a little bit subtler". Clark and Marshall suggest two versions of this strategy, the PROGRESSIVE CHECKING strategy, in which Ann progressively checks conditions (1), (2), (3) etc., and the SELECTIVE CHECKING strategy, in which Ann chooses a higher level condition such as (3) to check. They reject both of these because conditions with nested knowledge of level 4 or greater seem to be unlikely mental objects for humans

to assess. Furthermore, they suggest a better solution. The paradox they are concerned with comes about because of their definite reference condition and the two implicit assumptions:

Assumption I: Ann ordinarily tries to make definite references that are felicitous.

Assumption II: To make a felicitous definite reference, Ann must assure herself of each of the infinity of statements (1), (2), (3), and so on.

Clark and Marshall are loath to abandon Assumption I. However their adoption of Lewis's definition of CK, as the COMMON KNOWLEDGE INDUCTION SCHEMA, allows them to abandon Assumption II, and thus solve the common knowledge paradox(p. 33). The point of the schema is that Ann and Bob don't have to confirm the infinity of conditions in the iterate definition of CK. They need only be confident that they have a proper basis $\mathcal{A}$, grounds that satisfy all three requirements in Lewis's definition. With these grounds, Ann and Bob tacitly realize that they could confirm the infinity of conditions as far down the list as they wanted to go, but they need not actually do so. CK then can be treated as a single mental entity instead of an infinitely long list of ever more complex mental entities. I will return to this suggestion in section 5.

4.2 Knowledge, beliefs, assumptions

Another question of concern with respect to CK is whether knowledge is indeed the right notion, in contrast with for instance, beliefs, assumptions or suppositions. Any model of CK must depend on an underlying model of information, belief or knowledge [Kon85, Moo85, MSHI78]. An inference mechanism over the given set of facts also needs to be spelled out. But the discussion of these aspects gets very little attention from Lewis, Schiffer, Clark and Marshall and Barwise. Sperber and Wilson argue for the rejection of CK based in part on the claim that knowledge is much too strong.

Although a review of the various logics of knowledge and belief that have been proposed is beyond the scope of this paper, in this section I will review some claims of what properties a logic for CK must have. (See Halpern and Moses [HM85].)

We might be interested in knowledge, beliefs, suppositions or assumptions. However even if we restrict ourselves to knowledge, there is no agreement on what axioms are appropriate. Do you know what you know? Do you know what you don't know? None of the authors I have considered have proposed cases of CK based on lack of evidence. But autoepistemic reasoning for a single agent would have her reason that *If I had eaten duck's feet, I would know it, therefore I have never eaten duck's feet.* Lack of evidence for any joint action would seem to provide a basis for CK, e.g. *If we had ever danced together, we would know it, so we must never have danced together.* Do you know only true things, or can you know something that is false? Some would hold that all knowledge is true, and only beliefs can be false. Do you know all the logical entailments of your knowledge? Then you must know all tautologies as well.

Axioms for standard modal logics are typically taken from the following schemata:

M1: P, where P is a tautology.
M2: $\Box(P \supset Q) \supset (\Box P \supset \Box Q)$
M3: $\Box P \supset P$
M4: $\Box P \supset \Box\Box P$
M5: $not\Box P \supset \Box not\Box P$

The modal logic that includes only axioms M1 and M2 is called K. Adding M3 gives the modal logic T. If the modal operator $\Box$ is interpreted as KNOW, then adding M3 enforces that whatever is known must be true. This is clearly wrong if the modal operator is interpreted as BELIEVE. Adding M4, positive introspection, gives us the logic called S4. With M4 an agent can reason that if she knows P, she knows that she knows P. Adding M5, negative introspection, yields S5, so that if an agent does not know P, she knows that she doesn't know it. M5 is often considered inappropriate for a psychologically valid account of knowledge. Standard approaches to distinguishing knowledge from belief just eliminate M3 from S4 or S5, but the possible worlds model that the modal logics are based on still enforces that all agents are logically omniscient. Even Axiom M1 seems an unlikely axiom for belief. It often requires a great deal of inference to determine whether a statement is a tautology or not.

Barwise distinguishes between having information and having knowledge. Your information is the facts that you have at your disposal, whereas your knowledge depends on which inferences you do in fact make as a result of these facts([Bar88, HM85]). Barwise views his discussion of CK as being about shared information. He gives a motivating example. Consider the situation:

$$s = \{(H, Tom, Q\spadesuit), (K, Ellen, s), (K, David, s), (K, Tom, s)\}$$

where H is the relation used for 'having a card' and K is used for 'knowing'. It is clear that the fact:

$$\theta = \{(H, Tom, Q\spadesuit) \bigwedge (K, Ellen, \theta) \bigwedge (K, David, \theta) \bigwedge (K, Tom, \theta)\}$$

holds in this situation. But it is questionable whether it is a fact that Tom knows that David knows that Ellen knows that he, Tom, has the Q♠. Some sort of inference is required to get each iteration and the players might not make the inference. Even if they do, the other players may have doubts about whether they did. Once one player has doubts about some players making the relevant inference, the iterated knowledge facts break down. Both Lewis's and Schiffer's formulations of CK depend on being able to ascribe a certain level of rationality to others, in order to get the next fact in the hierarchy of CK facts. Clark and Marshall adopt Lewis's rationality assumptions. But Barwise notes that there may be many reasons why one would not ascribe the necessary rationality to another.

Then there is the role of suppositions. Clark and Marshall take Stalnaker's view of presupposition as CK. 'A proposition is presupposed if the speaker is disposed to act as if he assumes or believes that the proposition is true, and as if he assumes or believes that his audience assumes or believes that it is true as well.... The propositions presupposed in the intended sense need not really be common or mutual knowledge: the speaker need not even believe them. He may presuppose any proposition that he finds it convenient to assume for the purpose of the conversation, provided he is prepared to assume that his audience will assume it along with him'([Sta78] p 321). Clark and Marshall claim that to refer felicitously Ann must know that t is R, but they note that often all Ann will be able to check is her belief or assumption or supposition instead of her **knowledge** that t is R. They state further that the appropriate propositional attitude, be it knowledge, belief, assumption, supposition, or even some other term, depends on the evidence Ann possesses and

other facts, but that **know** can be used as a general term, and that they could replace their usage of **know** with belief or certain other terms without affecting their argument(p 12).

They offer no direct support for this claim, but it has implications for the formalism that we choose to represent CK, since the commonly used axioms for belief are not the same as those for knowledge. The most prevalent formal model of belief is based on possible world semantics. However this approach does not allow one to distinguish between two logically equivalent sentences, requiring the beliefs of an agent to be closed under logical consequence. The problem is that humans are not ideal reasoners. Humans are clearly resource bounded; they do not infer all the logical consequences of their beliefs, either because they don't have time to draw the inferences or because they might not have a necessary inference rule[Kon85, Lev84]. Also it may be important to distinguish between the beliefs of the different agents involved. Can we ever assume identical beliefs? Do we separately maintain your beliefs, and my beliefs, and my assumptions about your beliefs?

Clark and Marshall postulate a tradeoff between evidence and assumptions(p 34):

> Evidence + Assumptions + Induction Schema = CK.

Since the induction schema is fixed, weak evidence implies that strong assumptions must be made in order to satisfy the induction schema and infer CK(see section 6). Clark and Carlson follow this with the claim that objections to mutual belief rest on the false assumption that mutual beliefs cannot vary in strength([CC82], p 6). Imagine in the duet example, that as the evening wears on that Perlman notices that Zuckerman is becoming absent-minded. By the induction schema, Perlman has reason to believe the agreement holds, but not very good reason. His grounds are weaker than earlier in the evening, and his belief that they mutually believe that the next gesture is for real is correspondingly weaker. That is mutual beliefs range from weak to strong in line with the grounds on which they are based. The stronger the grounds the stronger the mutual beliefs.

Prince explicitly rejects the term CK, preferring to refer to assumptions[Pri78, Pri81]. As she describes it, on entering into a conversation with someone we use particularized knowledge or stereotypical knowledge about them. This brings in a certain number of TACIT ASSUMPTIONS, henceforth TA's. A conversation among n participants will involve n sets of TA's. Each person sets aside a subset of TA's for each of the other participants, which are more or less accurate. But according to Prince, this is as close as one gets to CK, unless the participants are clairvoyant.

Sperber and Wilson seem to base many of their arguments against the notion of CK, specifically against the use of the term knowledge in the definition([SW82], p. 68-9). (See section 7). They point out that

> ..in fact we all take risks whenever we engage in verbal communication... what this suggests is that the formal argument is irrelevant to actual comprehension. It leaves out a simple fact: we don't need to be sure that a remark is say, in English but only to have sufficient ground for assuming that it is.

In fact, in more recent work, they use a notion called MUTUAL MANIFESTNESS, which looks like a paraphrase of Lewis's definition of CK. But they claim that being mutually manifest is not the same thing as being mutually known[SW86].

It seems clear that most discussions of CK have in fact used the term knowledge, where it would not be appropriate if knowledge were to refer to only true, certain facts as given by axiom M3

above. It also seems clear that in order to deal with representing statements that might not be true, but just more or less strongly evidenced, that we need to have the capability to represent default assumptions, that would then be retractable[JWW86]. This implies that we need some kind of default logic and defeasible reasoning process. It appears that we may need psychological studies to address the issue of how rationality and the propositional attitudes used in comprehension work, so that the appropriate formalisms can be chosen. I will examine this further in section 9.

5 The Model Theory of Common Knowledge

Barwise tackles the problem of an infinite regress discussed in section 4.1, by first arguing that reality is not wellfounded anyway, and then modeling it using nonwellfounded sets. Second, he finesses the knowledge/beliefs problems by saying he is only looking at 'information', which he assumes to be out there in the world. Third, he addresses a question that I have not yet raised, which is whether the different definitions of CK in fact define the same phenomena. Barwise develops a framework in which he can compare the approaches above without presupposing that in fact they are equivalent, and thus is able to demonstrate what assumptions are necessary for them to be equivalent.

To explain Barwise's position let us first consider the fact that the fixed point and the shared-environment approach are blatantly circular. Many of the criticisms directed at the notion of CK have revolved around just this circularity (see section 4.1). Barwise lays the ground for his treatment of these facts by arguing that reality is not well founded. Then he uses Aczel's theory of non-well-founded sets[Acz88].

His examples of the non-wellfoundedness of reality rest on the demonstration that many real world situations are self referential. These include Grice's M-intentions, as given in Schiffer's definition of speaker meaning in section 2.4. have others recognize various things including the very intention. But all kinds of simpler situations are circular as well. Consider any self-referential statement, such as *This announcement will not be repeated.* Or imagine the physical situation of two parallel mirrors of the same size, facing each other, one A with an "X" painted on it and the other B with an "O". This is a simple finite physical situation but A reflects the "O" on B, but also reflects B reflecting the "X" on A etc.

In fact Barwise wants to shift the focus of attention away from individual mental states, to something that people have evidence that they share, namely their physical environments. I will return to this in section 6. The situation that drives much of Barwise's discussion is that of playing stud poker. At a poker table, everyone can see all the up cards of each player, and see each other seeing all the up cards, and see each other seeing each other see all the up cards, etc. Properties of cards such as being 'up' or 'down' are inadequate to represent the situation, because this treatment would not extend naturally to versions of the game in which, for instance, you had to show your 'up' cards only to the person on your right, and each player thus has different information. We use this situation to model the information each participant has.

Section 3 introduced the two relations S and H. S is a relation to a situation s, where ϕ obtains[8]. The relation H allows us to construct level 0 statements. The statements that one can make of the form "$p_i S\phi$" where ϕ is a level 0 statement are called level 1 statements. It is easily shown that

[8]Barwise rejects the standard modal logic interpretation of S as a relation between players and sets of possible worlds, due to the logical omniscience problem.

we cannot represent all the information about the game with just level 0 and level 1 statements. For instance we cannot distinguish the situation where everyone sees your cards because they are up, or by accident, such as by them being reflected in a mirror. In the former case where it is CK what cards everyone sees, there is a lot more information. In fact one can show that no finite level captures all the information that is represented in the finite, albeit circular situation represented by the public situation s.

In what follows, S is interpreted as 'having the information that'. This distinguishes between having information and knowledge. Knowledge is taken to be having information in such a way as to be able to use it, and is considered to be a problem of cognitive science rather than logic. With information one is justified in concluding that if you have the information that ϕ, and also the information that ϕ implies ψ, then you have the information that ψ. As Barwise puts it, information travels at the speed of logic, knowledge and belief at the speed of cognition and inference. By assuming an ideal reasoner for the model theory, he intends to characterize shared information rather than shared knowledge.

The relationship of non-well-founded sets to situation theory is found through the sets of facts that hold in a particular situation. These sets of facts are derived from the situation in question by the use of a forgetful functor, M, which applied to a situation, forgets the structure of the situation, and returns the set of facts that hold in that situation. These facts are the canonical model of a situation, thus facilitating the use of Aczels' theory of non-well-founded sets to model circular situations which are paradigmatic for the establishment of CK. Once we have established this correspondence, we can then compare the three definitions given for CK, iterate, fixed point and shared environment, to determine whether they are in fact the same.

5.1 Fixed point compared with iterated

We need a few more definitions than what we have used up to now. Remember that situations are modeled as sets of facts. The neutral facts are sometimes called infons.

> Definition 1: The models of situations and infons form the largest classes SIT, INFON such that[9]:
>
> - INFONS are of the form $< Hpc >$ or $< Sps >$, with p a player, c a card and s in SIT.
> - A set s is in SIT if s is a subset of INFON.

Facts are the result of applying a forgetful functor to a situation. The forgetful functor loses all the structure of the situation and returns a set of facts. A fact σ HOLDS in a situation s, $s \models \sigma$ is defined as:

- $s \models (pHc)$ iff $< Hpc > \in s$.
- $s \models (pSs_0)$ iff there is an s_1 such that $< Sps_1 > \in s$, and for each $\sigma \in s_0$, $s_1 \models \sigma$.

[9]The wellfounded situations and wellfounded facts form the smallest classes, Wf-Sit and Wf-Fact satisfying these conditions.

By the second clause, if a player in s, sees or has the information s_1, and if s_1 satisfies each $\sigma \in s_0$, then in s, that same player sees or otherwise has the information s_0. If we were speaking of knowledge rather than information, this would not be reasonable since it would imply ideal reasoning capability.

A situation s_0 is a SUBSITUATION of s_1, $s_0 \sqsubseteq s_1$, iff

- If $< Hpc > \in s_0$ then $< Hpc > \in s_1$
- If $< Sps > \in s_0$ then there is an s_i such that $s \sqsubseteq s_i$ and $< Sps_i > \in s_1$.

The first order facts in a subsituation must hold in the parent situation. The second order facts, constructed from the operator S, must hold in some situation that is a subsituation of the parent situatioin. Thus at some level of subsituations the fact holds.

The notion of HOLDS IN and SUBSITUATION are related by(p 209):

- If $s_0 \subseteq s_1$ then $s_0 \sqsubseteq s_1$.
- If $s_0 \sqsubseteq s_1$, and $s_1 \sqsubseteq s_2$, then $s_0 \sqsubseteq s_2$.
- $s_0 \sqsubseteq s_1$ iff for every situation s, if $s_1 \sqsubseteq s$, then $s_0 \sqsubseteq s$.
- If $s_0 \models \sigma$ and $s_0 \sqsubseteq s_1$, then $s_1 \models \sigma$.
- For all situations $s_0 \sqsubseteq s_1$, and $s_1 \models \sigma$ for each $\sigma \in s_0$, are equivalent.

Thus a subset relation on facts, provides a subsituation relation on situations. Additionally the subsituation relation is transitive, as one would expect, and monotonic.

Two situations s_0, s_1, are INFORMATIONALLY EQUIVALENT, $s_0 \equiv s_1$, if the same facts hold in them. By the above definitions, two situations are informationally equivalent iff they are subsituations of one another. It is possible for distinct situations to be informationally equivalent. Suppose s_0 is a proper subset of the set s_1 of facts, then compare the situation of one fact, {(Tom S s_1)}, with another with two facts, {(Tom S s_0),(Tom S s_1)}. Clearly these two situations are distinct yet informationally equivalent.

Something very similar to this situation is what is going on with respect to the difference between the iterate representation of CK and its fixed point counterpart. In order to make this comparison, we need to define the sequence of facts that are generated from a particular fact.

The transfinite sequence θ^α, for α an ordinal, of wellfounded facts associated with an arbitrary fact θ is defined by induction on ordinals as:

$(pHc)^0 = (pHc)$
$(pSs)^0 = (pS1^{st}ord(s))$
and for $\alpha > 0$

$(pHc)^\alpha = (pHc)$
$(pSs)^\alpha = (pSs^{<\alpha})$
where $s^{<\alpha} = \sigma^\beta \mid \sigma \in s, \beta < \alpha$

Similarly for any situation s, we define the transfinite sequence , $\{s^\alpha \mid \alpha \in Ordinals\}$ by letting $\{s^\alpha = \sigma^\alpha \mid \sigma \in s\}$.

For example, consider the fixed point situation we looked at before.

$$s_c = \{(p_1 S(s_u \bigcup s_c)), (p_2 S(s_u \bigcup s_c))\}$$

Let τ_1 be the first fact and τ_2 the second. Then by the inductive definition given:

$$\tau_1^0 =< S, p_1, s_u >$$
$$\tau_2^0 =< S, p_2, s_u >$$
$$\tau_1^1 =< S, p_1, \{s_u \bigcup \{< S, p_1, s_u >, < S, p_2, s_u >\}\} >$$
$$\tau_1^2 =< S, p_2, \{s_u \bigcup \{< S, p_1, s_u >, < S, p_2, s_u >\}\} >$$

We can continue forever, generating a hierarchy of wellfounded iterated facts. At each finite level, we have what is called a FINITE APPROXIMATION of the fixed point fact.

Also we must define entailment; a fact σ ENTAILS a fact τ, $\sigma \Rightarrow \tau$, if for every situation s, if $s \models \sigma$ then $s \models \tau$.

Theorem: Let θ be some fact(Barwise's Props 2 and 3 and Thm. 5, p 210-211).

1. For all α, $\theta \Rightarrow \theta^\alpha$.
2. For any situation s and ordinal α, $s^\alpha \sqsubseteq s$.
3. If θ is not wellfounded, then the hierarchy of approximations never terminates. In particular for each $\alpha < \beta$, θ^α does not entail θ^β
4. If each approximation fact θ^α holds in a situation s, then so does θ.

Statement (1) says that a simple circular situation entails all of its approximations, ie. each iteration of the inductive definition. In addition, by (2), each set of approximations is a subsituation of the original situation, so the original situation supports any facts that are supported by the set of approximations. Finally, for circular situations, the hierarchy of approximations never terminates, new facts are generated at each iteration, and if all these facts hold in a situation then so does the original circular fact.

This means that the finite approximations of a circular fact will be equivalent with respect to finite situations. However the iterates themselves form an infinite situation. If we drop the restriction to finite models, one must look at the whole transfinite sequence of approximations. No initial segment is enough, and thus the iterate approach is actually weaker than the fixed point approach[10].

5.2 Fixed point compared with shared environment

Lewis's definition of CK is called the shared environment definition, and Barwise's version of it is the shared situation approach.

[10]Mislove etal. show that if the iterates are defined differently then the fixed point is the limit of the iterates.

In order to model the shared situation approach, Barwise introduces a simple second order language for making existential statements about situations, called CONDITIONS. An example of a CONDITION is:

$$\exists e[e \models ((EllenH3\clubsuit) \bigwedge (EllenSe) \bigwedge (TomSe))]$$

This condition is a shared environment analysis of the fact that Tom and Ellen share the information that Ellen has the 3♣. The variables $e_1, e_2, \ldots$ range over situations, with constants for the cards and players. The atomic statements are those of the forms (pHc) and (pSe_j). The set of conditions are the smallest set containing the atomic statements and closed under conjunction, existential quantification over situations and the rule: if Φ is a situation, so is $e_j \models \Phi$[11]. Finally given any function f which assigns situations to variables, we define $s \models \Phi[f]$ as[12]

1. If Φ is an atomic statement, then $s \models \Phi[f]$, iff the appropriate fact is an element of s. For instance, if Φ is (pse_j) then $s \models \Phi[f]$ iff $< S, p, f(e_j) > \in s$.
2. If $\Phi = \Phi_1 \bigwedge \Phi_2$, then $s \models \Phi[f]$, iff $s \models \Phi_1[f]$ and $s \models \Phi_2[f]$.
3. If $\Phi = \exists e_j \Phi_0$, then $s \models \Phi[f]$ iff there is a situation s_j so that $s \models \Phi_0[f(e_j/s_j)]$.
4. If $\Phi = (e_j \models \Phi_0)$ then $s \models \Phi[f]$ iff the situation $s_j = f(e_j)$ satisfies $s_j \models \Phi_0[f]$.

Two conditions are strongly equivalent if they hold of the same situations, or sequences of situations for multiple condition variables. Two conditions are informationally equivalent if they entail the same facts. Clearly any two situations which are strongly equivalent are informationally equivalent, but the converse does not hold. If two conditions have a minimal model in common, they are informationally equivalent. Once one shows that every condition has a minimal model, then the converse also holds, and it is clear that minimal models characterize conditions up to informational equivalence.

For example, define:

$$\Phi(e_1) = [e_1 \models ((p_1 HQ\spadesuit) \bigwedge (p_1 Se_1) \bigwedge (p_2 Se_1))]$$

$$\Psi(e_1) = \exists e_2 [e_1 \models ((p_1 HQ\spadesuit) \bigwedge (p_1 Se_1) \bigwedge (p_2 Se_2)) \bigwedge e_2 \models ((p_1 HQ\spadesuit) \bigwedge (p_1 Se_1) \bigwedge (p_2 Se_2))$$

Any model of $\Phi(e_1)$ is a model of $\Psi(e_1)$, since e_1 and e_2 can be the same situation. However there are models of $\Psi(e_1)$ which are not models of $\Phi(e_1)$. For instance consider the case where p_1's $Q\spadesuit$ is a down card instead of an upcard. However p_1 has a mirror directly behind her and p_2 has a mirror directly behind him. The point is that they have exactly the same information. But p_1 sees the situation through a different mirror, and thus gets the information in a different way. Therefore there must be two different situations represented by the two reflections. On the other hand consider the fact that s is a minimal model of both situations:

[11] $\models$ does double duty here.

[12] This definitions are from Barwise's TARK paper. The version in the Situation in Logic chapter uses sequences of situations rather than a function f; I don't know if the increase in notation is worth it.

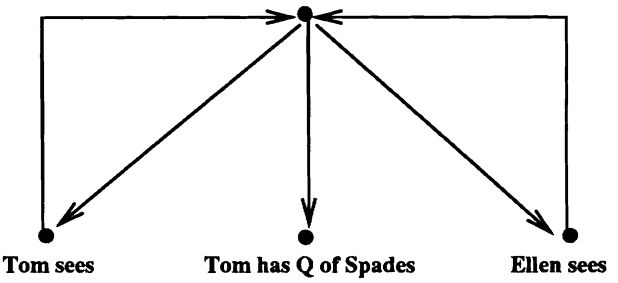

Figure 4: Mutual knowledge between Ellen and Tom that Tom has $Q\spadesuit$

$$s = \{< Hp_1 Q\spadesuit >, < Sp_1 s >, < Sp_2 s >\}$$

Barwise claims that what is going on is that the two conditions represent different ways in which p_1 and p_2 might share the information that p_1 has the Q♠. The first situation would represent a completely shared physical environment. The two situations share a minimal model, but they are clearly not equivalent. Barwise claims that neither one can be the right characterization of the shared situation. The fixed point characterization is 'cleaner' because only the facts of the situation are represented, without any possibility of representing that the shared information arose in different ways.

> THEOREM: Every condition Φ has a minimal model among the hereditarily finite situations.

This means that any condition can be approximated by a fixed point situation. This is because the fixed point situation will always provide a minimal model for the condition in question. Aczel's definition of hereditarily finite is used here. A hereditarily finite situation, is one in which the set of facts in it can be finitely pictured. (See p 7, [Acz88]). A finitely pictureable set can be pictured by an accessible pointed graph(apg), which consists of a set of nodes, with one distinguished as the point, and a set of directed edges, such that each node is accessible from the point via a directed path.

Mislove etal. clarify Barwises's use of Aczels theory[MMO90].

> Definition: A situation infon graph or sigraph, is an apg G with the following properties:
>
> 1. Each node of G has a type. This type is either situation or infon.
> 2. If a node n has type situation, then every child of n has type infon.
> 3. If a node has type infon, then either
>   - n is labelled with a card fact (p Has c) and n has no children, or
>   - n is labelled with (p Sees) for some p, and in addition n has exactly one child and this child is of type situation.

Then if we want to represent some shared knowledge such as that *Ellen and Tom mutually see that Tom has the* $Q\spadesuit$, we can do so with a sigraph as defined in Figure 4.

Here the top node, the point, is of type situation and the other nodes are of type infon. The node at the bottom is labelled with a level 0 fact, and the other nodes involve the **sees** relation. If we

check each arc in the graph, we can see that each of the arrows instantiates one of the clauses in Clark and Marshall's COMMON KNOWLEDGE INDUCTION SCHEMA. Thus a potential connection exists between Clark and Marshall's notion of the shared-environment schema representing a single mental entity and the theorem above along with the graph representation which is clearly finite. The theorem shows that every circular shared-situation has a minimal model among the hereditarily finite situations. The connection between the notions involved depends on one believing that finitely picturable sets can be thought of as a single mental entity, and are thus more psychologically plausible than sets that can't be finitely pictured.

The theorem above also has a corollary:

COROLLARY: Two conditions are informationally equivalent iff they have a minimal model in common[13].

On the graph-based interpretation, that would mean that we could draw an apg representing all the information that both situations share. Unfortunately this is not as useful as one might assume. Although Barwise suggests that the shared environment approach allows one to distinguish the ways in which the CK might have come about, in fact, if one wants to maintain a minimal model for the two different situations, all that one can distinguish is **that** it came about in different situations. If any statements about **how** these situations are in fact different are included, the minimal model is lost. However the situation which is characterized by the minimal model will be a subsituation of the two we are comparing, since that represents just what information is shared. Consider two situations in which we both know the same fact σ.

- **My Situation**
  - $s_1 \models \sigma$
  - $s_1 \models I\ know\ s_1$
  - $s_1 \models you\ know\ s_2$
- **Your Situation**
  - $s_2 \models \sigma$
  - $s_2 \models I\ know\ s_1$
  - $s_2 \models you\ know\ s_2$

These two situations have a minimal model in the fixed point representation which would be that our situations are the same. There is just one situation, call it our situation, s.

- **Our Situation**
  - $s \models \sigma$
  - $s \models I\ know\ s$
  - $s \models you\ know\ s$

[13] Unfortunately, two informationally equivalent conditions do not have to have exactly the same minimal models. (See [Bar88] for an example).

For instance, suppose you and I both received an email message stating the fact σ that *Barwise is moving to Indiana.* We each can verify by the address header that we had both received it and in fact we both know that we both make a practice of reading address headers to see who gets which messages. We also must both know that we read our mail daily. Thus after a one day time lag, we are justified in believing that it is CK that Barwise is moving to Indiana. However the situations in which we both came to know this are different. Presumably you were sitting in your office reading your mail, situation s_1 and I was sitting in my office reading mine, situation s_2. Let us represent this by single facts in each situation: $\theta \in s_1$, and $\psi \in s_2$. Thus we have two situations:

- **My Situation**
  - $s_1 \models \sigma$
  - $s_1 \models I\,know\,s_1$
  - $s_1 \models you\,know\,s_2$
  - $s_1 \models \theta$
- **Your Situation**
  - $s_2 \models \sigma$
  - $s_2 \models I\,know\,s_1$
  - $s_2 \models you\,know\,s_2$
  - $s_2 \models \psi$

These two situations no longer have a minimal model in common. They will however have a subsituation which has a minimal model, namely the one that excludes the distinguishing facts θ and ψ.

5.3 Summary

Barwise concludes that the fixed point definition captures exactly the pretheoretic notion of common knowledge, whereas the shared-environment account allows one to represent the different ways in which knowledge might come about. As I have shown, all it really lets you represent is that it came about differently. He rejects the suggestion that the iterate approach might characterize **how** it is that the CK really gets used. The only way the iterated and fixpoint approaches are the same is under the assumptions of ideal reasoners, and that it is CK that all the players are ideal reasoners. But without these assumptions, it is clear that an agent may not make the necessary inferences from the fixed-point representations that are needed to generate the infinite sequence characteristic of the iterate approach. Thus if we assume that the fixed-point approach is the right characterization, the facts of the iterate representation are not really facts at all. In contrast, Tan and da Costa Werlang [TW86] claim to have proved that the fixed point approach is equivalent to the iterate approach. Obviously they made these assumptions without noting it.

When comparing the fixed point and shared environment approach Barwise notes that the fixed point representation will always provide a minimal model for a shared-situation representation. However, under the shared-situation approach, two different situations may entail the same facts

and yet have different minimal models. This is because there are many different situations that can give rise to a particular piece of CK.

Barwise uses the relation 'sees' because he assumes that each player can see what there is for him to see. This seems uncontroversial in the case of cards, which have a conventional meaning, and are clearly either 'up' or 'down'. However this physical situation is not like a linguistic one. The mental states involved in 'understanding what there is to understand' or 'believing what there is to believe' are clearly not so straightforward.

Barwise conjectures that the notion of CK is not really that useful, that it is a necessary but not sufficient condition for action. On his view, what suffices for common knowledge to be useful is that it arises in some fairly straighforward shared situation, which provides a basis for perceivable situated action. This action then produces further shared situations. That is, what makes a shared situation work is not just that it gives rise to common knowledge but that it provides a stage for maintaining CK through the maintenance of a shared situation.

6 The Structure of Common Knowledge

One question that we have not yet addressed is how CK might be structured. What might be the interplay between a structure and processing models for comprehension? Given the the difference between belief and knowledge we might imagine that: (1) structure would tie in with what CK is used for, (2) Inferences would have different status depending on what type of assumptions have been made and whether or not they are defeasible, (3) The type of evidence supporting CK may well influence whether a certain set of facts gets invoked in producing or restricting implicatures.

Barwise treats CK simply as a set of propositions[Bar88]. Joshi suggested that it is necessary to distinguish between generic and particular cases of CK[Jos82]. Prince distinguishes known entities from salient ones. Parikh suggests that subjective probabilities need to be attached to CK facts[Par90]. Parikh also suggests the need to distinguish between facts that are salient and those that are just known. He notes the difference between the known facts when Ann notices that Bob's knee is touching hers, which of course he knows as well, and the situation in which everything is just the same as before but Ann has said to Bob *Your knee is touching mine.* The explicit mention of a fact changes its CK status. Schiffer also distinguished between a belief and an activated belief in order to explain such speech acts as reminders. Clark and Marshall propose a specific structure for CK, that it should be organized to show its bases. I will examine their proposal in more detail later in this section. First I want to review the various bases for CK that have been proposed.

6.1 Bases for Common Knowledge

Barwise assumes a physical situation as the main way that CK comes about, but showed that the shared-environment approach allows one to have informationally equivalent situations, which nevertheless reflect that there were different situations from which the CK arose. In the main, he restricts himself to physical situations and conventionally signalled facts, such as a card being the Q♠, and being 'up'. On his view,

> ...just what one knows is highly dependent on one's external circumstances – where they are, what they see, what is going on around them more generally, and what things

> are relatively stable in their environment. By focusing on the objective circumstances of individuals and their role in the determination of what people know and believe, we can exploit shared circumstances in the explanation of shared understanding([Bar88], p 197).

Although other researchers argue for other sources of CK, when Barwise talks about 'what is going on around someone more generally' and 'what things are stable in someone's environment' it isn't clear whether one might interpret these sources as being generic, commonsense knowledge or aspects of one's community membership. In general, Barwise seems to assume that physical situations just 'offer up' the facts which hold in them. In fact this position is somewhat controversial[FP81]. Whether they do will in general depend on which representation of a fact is assumed to hold.

Lewis and Schiffer both base their definition of CK on the idea that if one knows something one knows **how** it is that one knows it, and this provides the basis for determining whether someone else might know it as well. Schiffer discusses it in terms of properties such as being a normal person and having one's eyes open and facing the candle. Lewis's definition rests on some state of affairs $\mathcal{A}$ providing a BASIS for common knowledge. The types of bases he discusses include only(p. 57):

- Agreement
- Salience (a weaker basis than agreement)
- Precedents, especially past conformity to a convention

Lewis does point out the role of salience in CK, as did Schiffer with his notion of **activated** belief. Nevertheless these bases seem to be tightly linked to his notion of convention in terms of coordination games; they do not even include the physical basis which has been proposed to be the strongest basis by others. For instance, Lewis says little about the linguistic bases for CK. His first proposed language, $\mathcal{L}$, a two-sided signalling system, ignores the effects of context. In extending $\mathcal{L}$, he notes that $\mathcal{L}$ may contain indexical sentences whose truth conditions depend on the utterer, on his intended audience, or on the time and place of the utterance. Furthermore $\mathcal{L}$ may contain anaphoric sentences whose truth conditions depend on the previous conversation or intended subsequent conversation, or $\mathcal{L}$ may depend on the surroundings of their utterance. Lewis seems to believe that determination of the actual context for an utterance is trivial. He states that the occasion of utterance is identified with a pair of a possible world and a spatiotemporal location. Given this, 'all the further information we need about the context will be forthcoming. We will have uniquely identified the utterer, his intended audience, the previous conversation, the surroundings and so on.'

Schelling suggested that finding the key to a coordination game depend on imagination, analogy, precedent, accidental arrangement, symmetry, aesthetic or geometric configuration, ethical reasoning, and who the parties are and what they know about each other. Of these, precedent, ethical reasoning and who the parties are and what they know about each other could clearly be bases for CK.

Clark and Marshall propose a taxonomy of bases for CK by their strengths. The next section examines the details of this proposal.

6.2 Classification of Bases for Common Knowledge by Strengths

As we noted before Clark and Marshall propose that there are three factors that allow one to infer CK:

Evidence + Assumptions + Induction Schema = CK.

Because the induction schema is fixed, there is a tradeoff between the strength of the evidence and the number of assumptions that must be made. In addition, due to the multiple sources for CK, it must be classified in human memory in order to show its sources in a person's experience. This will predict the speed with which humans access CK[14].

They propose an organization based on the TYPE of evidence and assumptions. CK might additionally be organized by **whether** it is based on evidence or on assumptions, that is whether the belief is defeasible or not. Clark and Marshall don't address this. They propose that in order to support felicitous reference, memory must be organized into two components. The first is encyclopedic knowledge organized by community membership, and perhaps hierarchically as well by what everyone knows, what Philadelphia residents know, what Penn students know etc.. The second is a personal diary of event structures. One must be able to access this diary based on which individuals were present at a particular event in order to use the copresence heuristics and thus determine CK. In addition the diary must have another level of organization, based perhaps on recency or significance. This idea is not developed here in any more detail, but has been studied under the rubric of attentional state[JW81, GJW86].

The main distinctions are between lasting and temporary kinds of CK, between several kinds of temporary CK, and between generic and particular knowledge. Clark and Marshall propose four bases deriving from the different types of evidence for CK and their associated auxiliary assumptions(See Figure 5, from [CM81] p. 43). Contrary to Barwise, they show that although the physical situation may be the primary basis, there are other bases as well, for instance community membership such as discussed by Prince[Pri78].

COMMUNITY COMEMBERSHIP supports facts such as *everyone who lives in Philadelphia knows who the mayor is*, and is normally preserved over long periods of time. PHYSICAL COPRESENCE, combined with attention, is strong evidence of CK such as that given by the up cards in a poker game. IMMEDIATE COPRESENCE relies only on my assumptions about your rationality, attention and our simultaneous presence in the same environment. But reference time may vary with respect to the time of physical copresence. An expression like *the dog we saw yesterday* which typifies prior physical copresence relies on the added assumption of recallability, and my saying *I wonder how that dog got in here* relies on potential physical copresence, with the added assumption of locatability, ie. you being able to visually locate the referent. Obviously physical copresence of all types is temporary.

Many things that are referred to have only been mentioned in conversation, such as a dog, in *I saw a dog yesterday.* Indeed many things can **only** be referred to in conversation, future actions, imaginary creatures such as mermaids etc.. But LINGUISTIC COPRESENCE is a bases for common knowledge, since I can then make a definite reference to *the dog*, or *the mermaid.* However linguistic copresence is normally weaker evidence for CK than physical copresence. Whereas seeing is believing, hearing about something requires more, the understandability assumption. In addition Clark

[14]This is reminiscent of scripts, frames, schemas etc.

BASES	ASSUMPTIONS
COMMUNITY MEMBERSHIP	Community comembership, Universality of knowledge
PHYSICAL COPRESENCE Immediate	Simultaneity, Attention, Rationality
Potential	Simultaneity, Attention, Rationality, Locatability
Prior	Simultaneity, Attention, Rationality, Recallability
LINGUISTIC COPRESENCE Potential	Simultaneity, Attention, Rationality, Locatability, Understandability
Prior	Simultaneity, Attention, Rationality, Recallability, Understandability
INDIRECT COPRESENCE Physical	Simultaneity, Attention, Rationality, Recallability OR Locatability, Associativity
Linguistic	Simultaneity, Attention, Rationality, Recallability OR Locatability, Understandability, Associativity

Figure 5: Bases for Common Knowledge and Supporting Assumptions

and Marshall distinguish two kinds of linguistic copresence, each with one additional assumption[15]. Prior copresence, requires recallability, as in *I bought a candle, but it was broken* and potential linguistic copresence requires locatability, as in the *Because it was broken, I returned a candle I had just bought to the store.*

Both types of copresence are difficult to compare with community membership because the assumptions are so different. And often CK is established by a combination of the types given. For example if Ann says *I bought a candle yesterday, but the wick had broken off*, she must assume that to refer to the wick that her utterance of a candle establishes the indirect linguistic copresence of her, Bob and the wick. But this in turn must be based on her assumption that they belong to a community of people, for whom it is universally known that candles have wicks. Indirect physical copresence also relies on community membership, e.g. a physically present book may have an indirectly present author, *The author also wrote Sure of You.* This knowledge may be generic or particular, as in *I saw Ann yesterday. The baby is doing fine*, referring to some particular CK about Ann having a baby. The assumptions that are required to induce CK from indirect copresence is loosely termed ASSOCIATIVITY. This is what Prince called INFERRABLE[Pri81].

[15]While Clark and Marshall claim that linguistic copresence can never be immediate, a counterexample might be *This announcement will not be repeated.*

Clark and Marshall use the taxonomy developed above to determine when particular kinds of referring expressions can be used. The only further issue that interests us here is their claims about repairs. Repairs, ie. speaker corrections of their own referring expressions, give evidence as to the relative strength of the different bases for CK. A repair, they hypothesize, must strengthen the grounds; in order to succeed in repairing, it must remove some of the auxiliary assumptions. There are two ways to do this. The first is by leaving the type of grounds the same, but making the description more narrowly specified. For instance -

Ann: *A doctor I met last night introduced me to a lawyer, and she gave me some advice*
Bob: *Who did?*
Ann: *The lawyer*

Another way is by providing a different type of evidence, replacing one kind of copresence with another. By seeing just which kinds of replacements make sense, Clark and Marshall determine which kind of evidence is stronger.

We would expect immediate physical copresence to be the strongest and it is. For instance:

Ann: *That's my book over there.*
Bob: *Which one?* (Ann moves over and picks it up)
Ann: *This one.*

Direct linguistic evidence is demonstrably stronger than indirect as we would expect.

Ann: *I bought a candle today and the seal was broken.*
Bob: *What seal?*
Ann: *The seal on the wrapper around the candle.*

By making comparisons like these, Clark and Marshall conclude that direct physical copresence is stronger than indirect, but that any kind of physical copresence is stronger than linguistic copresence, that direct linguistic copresence is stronger than indirect, and that community membership cannot be ordered for strength with the other types of evidence because the assumptions are too different.

The idea of different beliefs having different strengths has been explored by Galliers within the area of assumption based truth maintenance systems (ATMS) and their role in a theory of belief revision[Gal89, Gal90]. She has modified the ATMS to include endorsements with each belief. These endorsements reflect the relative strengths with which beliefs are held and her system of endorsements is compatible with Clark and Marshall's proposal, except that she distinguishes strengths based on specific versus general community membership facts, and claims that all such facts are defaults, and as such are held with weaker strengths than firsthand information from linguistic or physical copresence. However this work is fairly preliminary and in the main this area has not received as much attention as it should. Where research have addressed beliefs of different strengths, there is not much consensus. Clark and Marshall have made some concrete suggestions, but there is clearly much work to be done.

7 Relevance

Two primary objections have led a number of researchers to object to the idea of common knowledge: (1) The psychological implausibility of CK due to the potential infinity of conditions that must be checked, and (2) whether indeed knowledge is the right propositional attitude. Sperber and Wilson give a third objection: The need for CK is predicated on an unrealistic model of language use, which they term the CODE MODEL. These three factors lead Sperber and Wilson to propose a model of language comprehension based on a PRINCIPLE OF RELEVANCE. Relevance theory replaces the need for CK with a principle from which to generate speaker's assumptions. This section examines their proposals.

7.1 The Code Model of Communication

According to Sperber and Wilson (hereafter SW), from Aristotle to modern semiotics, all theories of communication were based on the code model. A code is a system which pairs internal messages with external signals, thus enabling two information processing devices to communicate. The model they characterize is exactly the formal model of coding as proposed by Shannon. On their view, Schiffer and Lewis are code theorists, as are most researchers in pragmatics[16]. They note that although language can be seen as a code which pairs phonetic and semantic representations of sentences, more recent views show that the literal meaning of an utterance greatly underspecifies its intended meaning. The gap is filled not by more coding but by inference([Rev87], SW).

They point out that inferential processes and decoding processes are obviously quite different. The only way, then, that pragmaticians who hold to the code model but describe comprehension using inferential terms can remain consistent, is by holding that not only do the speaker and hearer use the same language, but they must also use the same set of premises. This enables a symmetric encoding/decoding process to be performed at the sending and receiving ends. This set of premises is the context, and for code theorists, the context used by the hearer must always be identical with the one the speaker thinks the hearer will use. This is their CK. But how are speakers and hearers to determine what assumptions they share from the ones they don't? Within the framework of the code model, CK is a necessity. In rejecting the code model of communication, SW reject the notion of common knowledge.

For SW, Grice's definition of meaning provides the point of departure for a new model of communication, the INFERENTIAL MODEL. Most pragmatic accounts, according to Sperber and Wilson, assume that the context for the comprehension of an utterance is fixed in advance, and undergoes no more than minor adjustments during the comprehension process[17]. They distinguish their account from those that use CK, on the basis that the context for an utterance is determined during comprehension and not before, based on the number of inferences that a hearer can derive.

[16] Lewis's idea of language as a two-sided signalling system is basically a code, and neither Schiffer or Lewis discuss inferential meaning since they were interested in the development of conventional aspects of meaning.

[17] They fail to review work on presupposition and accommodation[Lew79].

7.2 Arguments against Common Knowledge

Sperber and Wilson start from the perspective that language comprehension is easily achieved. Thus CK is implausible for three reasons: (1) The CK paradox as discussed by Clark and Marshall[18]. The problems with identifying CK and evaluating an infinity of conditions do not give rise to the expected problems of comprehension; (2) CK is not a sufficient condition for belonging to the context. In reality the context is much more restricted than just what is mutually known; (3) They try to demonstrate by examples that CK is not a necessary condition for comprehension either. I will now recap their arguments.

To show that CK is implausible, they note that Clark and Marshall assume that all evidence for CK is ultimately physical. Linguistic copresence is physical copresence at a linguistic or acoustical event and community membership must be established through linguistic or physical copresence. They claim that one might have to do a lot of inferencing in order to connect one's knowledge with a particular physical event. They also point out that assuming CK under physical copresence does not justify a particular description of the physical object. Clearly having seen someone bury a piece of paper under a rock might not support the viewer knowing that *the tallest spy buried a message under the rock.*

They next attack the sufficiency of CK for comprehension. CK is just too large in many cases to be searched as quickly as it apparently is. The restrictions they mention are derived from the current conversational context, or from what is currently being attended to. For example if Omar says *I am a good Moslem*, then Ann starts her inferencing process from her knowledge of Islam. These examples suggest that they believe in some sort of associative memory or focusing process. Clark and Marshall also suggested that CK be ordered by recency, or by significance(See as well [Gro77, JW81, Sid79]).

To show that CK isn't necessary, they offer this example: Ann believes that Bob doesn't know which movie is playing at the Roxy tonight, but asks him if he has seen it anyway just to annoy him. Bob knows what Ann is up to and also just happens to know which movie is playing, so he replies that yes, he has seen it, Ann infers that she was wrong in her belief, and the fact that Ann and Bob mutually know which movie is playing tonight gets added to the CK.

Some of their arguments against CK revolve around the certainty with which certain beliefs are held. For instance, when I hear you speak, I "don't need to be sure that a remark is say English, but only to have sufficient ground for assuming that it is". According to them, an argument that subjects take feasible steps necessary for achieving certainty, although they know that those steps will never be enough, is implausible because it ignores processing costs[19].

7.3 Communication is really one-sided

In rejecting the notion of CK, SW also reject the basic premise that language is a coordinated activity, as was assumed by Clark and Marshall, and Schiffer. Lewis assumes that language has both a one-sided and a collaborative aspect(Ch 4, p 177-181). SW argue for a one sided view of communication. They agree that communication requires some degree of coordination between communicator and audience, but

[18]Clark and Marshall felt they solved the paradox with the adoption of Lewis's definition of CK as an induction schema.

[19]It isn't clear at this point if they are talking about speaker's or hearer's processing costs.

> ..ask yourself what are the grounds for assuming that responsibility for coordination is equally shared between communicator and audience, and that both must worry symmetrically what the other is thinking. Asymmetrical coordination is often easier to achieve, and communication is an asymmetrical process anyhow. Consider what would happen in ballroom dancing if the responsibility for choosing steps was left equally to both partners (and how little help the CK framework would be for solving the resulting coordination problems in real time). Coordination problems are avoided or considerably reduced, in dancing by leaving the responsibility to one partner who leads, while the other has merely to follow([SW86] p 43)[20].

The responsibility is left to the communicator to make correct assumptions about what the audience will have accessible and be likely to use in the comprehension process. The responsibility for avoiding misunderstanding also lies with the speaker, so that all the hearer has to do is go ahead and use whatever information comes most easily to hand, and presumably never ask themselves what **this** speaker might have meant by x.

7.4 The principle of relevance

SW propose replacing the role of CK with the PRINCIPLE OF RELEVANCE[SW82]:

> *The speaker tries to express the proposition which is the most relevant one possible to the hearer*

Relevance is defined as the ratio of input to output, where input is the amount of processing one must do, and output is the number of contextual implications (CIs) that one can derive from an utterance. The number of CI's must be finite, to be calculated in a short amount of time, and to facilitate the comparison of different interpretations with respect to their degrees of relevance. However most utterances could derive an infinity of trivial implications, e.g. $p \rightarrow p \bigwedge p$, and it is clear that humans do not waste their time deriving such trivial implications. So SW attempt to limit their deductive inference process to one which only derives nontrivial contextual implicatures (hereafter NTCI's), by postulating that there are no and-introduction or or-introduction rules. Thus we may view relevance R, as a function of the context, the amount of processing needed and the number of NTCI's.

Consider Sperber and Wilson's example of a Flag-seller F and a Mad Passerby P.

(2) F: Would you like to buy a flag for the Royal National LifeBoat Institution?

P: Thanks, I always spend my holidays with my sister in Birmingham.

In order to understand this fully, Sperber and Wilson claims that the hearer must supply at least the following premises:

1. Birmingham is inland.

[20] I have to question that even dancing is a one-sided activity. Certainly I have no hope of following your lead without any idea of what steps you might be trying to execute. The fact that there are a finite set of known steps and likely breakpoints in their combination is critical to my recognition of your intentions.

2. The Royal National LifeBoat Institution is a charity.
3. Buying a flag is a way of subscribing to a charity.
4. Someone who spends his holidays inland has no need of the services of the Royal National LifeBoat Institution.
5. Someone who has no need of the services of a charity cannot be expected to subscribe to that charity.

But someone who cannot derive the NTCI in (3) would be unable to determine the relevance of the passerby's response.

(3) The passerby cannot be expected to subscribe to the Royal National LifeBoat Institution.

They propose that as the hearer attempts to comprehend an utterance, they can expand the context in 3 dimensions: (1) the conversational record, (2) encyclopedic knowledge, and (3) the current physical environment. While these expansions are clearly reminiscent of Clark and Marshall's linguistic, community, and physical copresence heuristics, SW claim that the hearer does not have to worry about CK while doing these expansions. What guides these expansions is the proposed principle of relevance.

When a hearer is searching for the relevance of the utterance, each expansion of the context will provide more NTCI's at the cost of increased processing. They conjecture that the amount of processing remains roughly constant over a certain stretch of discourse so that this search must be limited to a small domain. The speaker must have grounds for thinking that the hearer has an easily accessible context in which they can get enough NTCI's. They acknowledge that the common ground would be such a context but that there must be others, since, for example, I can answer your question of *What time is it?*, with full relevance and without worrying about what in fact that answer might imply for you.

Although it is not clear exactly how it is linked, relevance seems to be dependent on awareness, or perhaps with the difference between explicit and implicit belief. Consider SW's example of a situation in which Peter and Mary are at home watching TV. The fact that the TV is on provides grounds that the electric company is not on strike. But according to SW whether Mary believes that there is no strike at the electric company depends on whether or not Mary asked herself if there was a strike on or not[SW86].

7.5 When Common Knowledge is Needed

SW claim that in general the inferences involved in determining the NTCI's need not be intended ones. For example if (4) is common knowledge and (5) is uttered,

(4) Ann is a nuclear physicist.

(5) Bob is in love with Ann.

Then they presume that (6) should become CK, whether or not the speaker intended this inference to be drawn.

(6) Bob is in love with a nuclear physicist.

However, according to SW, there are two cases where a speaker must assume a specific piece of context, (a) definite reference and (b) intended inferences. First consider the case of intended inferences. For example:

(7) A: Will you have some brandy?

B: You know I am a good Moslem.

B must assume that A knows that good Moslems don't drink alcohol in order to expect her to be able to see the relevance of his utterance.

The second case is definite reference, as Clark and Marshall argued. SW repeat Clark and Marshall's claim that anaphors cause one to search the linguistic context, deictics cause one to look in the physical context and proper nouns cause one to search encylopedic memory. They discuss cases of 'expanding the context' to get the referent of a definite referring expression. For example, (8) requires the assumption of the premise in (9):

(8) I have read John's novel. The character of Eliza is so moving.

(9) There exists a character called Eliza in John's novel.

They claim that the hearer is justified in making this assumption by the principle of relevance.

8 Arguments against the relevance of Relevance

SW propose that relevance is intended to replace CK as a means of making inferences about the goals and beliefs of others. CK is rejected on the basis of being an unattainable idealization, requiring an infinity of inferences. This rejection, however seems to be based on two false premises: the assumption that knowledge has to be certain and the assumption that CK requires an infinite chain of explicitly held assumptions. Under the fixed point or shared environment representations, it seems reasonable to assume that CK can be achieved in a finite amount of time by considering the representative set objects as single units.

Their example of CK not being necessary is actually one of a breakdown in communication. Ann's intention is to produce a referring expression that Bob will not be able to understand, but she does not achieve her intended effect of annoying Bob. The fact that CK gets updated as a result isn't the point. It is not clear what theories would predict about a breakdown in communication.

SW's *Bob is in love with a nuclear physicist* example makes the ideal reasoner assumption, that I will deduce everything possible from what was said and add that to the common ground. This is implausible on two grounds: (1) Humans are not ideal reasoners (2) Unintended inferences should not be part of the common ground. In addition SW contradict themselves by claiming that what is part of CK depends on one's awareness.

They reject Clark and Marshall's classification of bases for CK, because it is all ultimately physical, as though this is significant criticism. The statement *All the information you have about others*

ultimately comes to you through your senses doesn't seem to be debatable, unless perhaps you are a Siamese twin, but this is a paraphrase of SW's criticism. In addition they use a subset of this classification in their own discussion of CK used to achieve felicitous definite reference.

8.1 The deductive component

SW at least address the nature of the inference process for language, but the system they propose is very poorly motivated. It certainly is true that one does not want to derive such statements as $p \bigwedge p$ from p, however the idea that one can do without introduction rules at all is questionable. Gazdar and Good point out that in order to arrive at nontrivial inferences, that it is sometimes necessary to use a trivial inference rule[GG82]. For example consider the nontrivial inference:

$$\begin{array}{l} (P \bigvee Q) \rightarrow R \\ P \\ \hline R \end{array}$$

This can be shown to be valid by:

$$\begin{array}{l} (P \bigvee Q) \rightarrow R \\ P \\ P \bigvee Q \quad \text{[disjunction introduction]} \\ \hline R \quad \text{[modus ponens]} \end{array}$$

Sperber and Wilson reply that a rule would be stored in memory such that at the first step from $(P \bigvee Q) \rightarrow R$, a hearer would derive $(P \rightarrow (Q \rightarrow R)$ immediately, and then be able to derive the same conclusion without ever using disjunction introduction. Others have pointed out that such a rule might be required for any possible number of premises, and that it isn't clear whether SW mean that the first step is a schema or has to be instantiated in memory for every possible propositional combination[Rev87].

Clearly the debate about what kind of inference process is at work has not been settled. While Sperber and Wilson should be congratulated for drawing attention to the role of inference in conversational understanding, the current formulation is too underspecified to be of much use. In the next section, I will discuss another objection to SW's formulation. I will return to the nature of inference in conversation in section 9.

8.2 Communication isn't one sided

According to Sperber and Wilson's formulation of relevance *the speaker tries to express the proposition which is the most relevant one possible to the hearer*[SW82]. This section shows that the relevance of an utterance must sometimes depend on which of the hearer's views are taken into account.

One problem with the formulation of relevance is that an utterance which tells the hearer something they are already certain of is ir-Relevant. This is because presumably the hearer will not be able to derive any NTCI's from a known utterance. This applies as well to statements of tautologies, such as *A rose is a rose.* For example if A tells B, *Ellen was caught cheating at poker*, and B already knows this, then relevance would suggest that B may search her context for some other Ellen in search of some NTCI's. These utterances can be viewed as mistakes by the speaker, but repetitions and reminders presumably should have no NTCI's as well. SW claim that despite the fact that the hearer should be unable to determine the relevance of an utterance that informs them of a fact that they already know, that in fact hearers can do this by using the principle of relevance as a guide. The hearer knows the speaker was trying to be relevant. But this requires the principle of relevance to be part of CK and to be used by the hearer to figure out what the speaker might have meant, which seems to contradict their one-sided view of communication(See Clark in [Rev87]).

Similarly suppose that A refers to the woman B was just with as *your girlfriend.* Actually the woman was B's sister, but B does have a girlfriend. B will certainly understand A only by knowing or inferring what A does and does not know. In both of these cases, understanding depends on the hearer using her common ground with the speaker. Speakers can only be as relevant as their CK allows, whereas relevance theory predicts that addressees will make highly inappropriate inferences in situations in which speakers have made mistakes in referring(Millikan, p725 [Rev87], Gerrig, p. 718 [Rev87]).

Consider SW's example of Ann, who offers Omar a glass of brandy and receives this reply: *You know I am a good Moslem.* SW say that if Ann knows that brandy is alcoholic and that good Moslems do not drink alcohol, she can infer that Omar will not have a glass of brandy. She can also infer that Omar intended her to draw that specific inference, without which his utterance will not be relevant. But what if Ann believes that Omar believes that she knows nothing about Moslems or what if Ann also believes that Omar believes that she thinks good Moslems love alcohol. Then she would take him as accepting the glass of brandy. Or imagine that Ann also believes that Omar believes that she believes that Omar holds that good Moslems don't drink alcohol, so she will take him as refusing her again. Ann must therefore replicate Omar's beliefs about her beliefs([Cla82], p 127).

In SW's definition of relevance, it is difficult to determine what distinction they make between implicit and explicit belief. They seem to imply that only explicit belief can count as something being relevant. However it is clear that listeners do not have to access a relevant assumption consciously before a speaker can say something that makes it possible to comprehend an utterance. What is necessary is that it be mutually known to both parties that the listener is capable of using this tacitly shared information at the right moment. For instance:

(10) A: Are you going to the party tonight?

B: I hear Jack's coming.

Obviously understanding B's reply requires that A know something about B's attitudes about Jack. But it also requires that B knows that A knows this about her. It certainly needn't be the case that A had seriously thought about what B's attitude toward Jack is, but the reason B says what she does and doesn't make a more direct response to the question is that she intends A to base his inference not just on any knowledge of beliefs he has but on their CK. In Grice's terms, there is an 'authorized' inference, but SW's formulation leads to the generation of both authorized and unauthorized inferences(Gibbs, [Rev87], p. 718).

Finally, consider the fact that in SW's passerby examples, the set of premises they supply aren't the only set that a hearer might supply in order to make sense of the speakers response([Rev87], Wilks). Consider an alternative set:

1. The Royal National LifeBoat Institution is a charity that provides cheap holidays for poor elderly people.
2. The speaker is a shabby elderly looking person.
3. Someone who already has holiday provisions will not need of the services of the Royal National LifeBoat Institution.
4. Someone who has no need of the services of a charity cannot be expected to subscribe to that charity.

As Wilks points out, it doesn't matter that (1) is a false belief, as was SW's (5), for belief attribution in communication cannot require that we attribute to others only beliefs we happen to hold. Wilk's premise (2) is a belief of the hearer about the speaker, rather than a belief of the speaker. But it might not be appropriate for the hearer to attribute this belief to the speaker. Thus one must consider whose mental space the inferencing is supposed to take place in.

9 The Role of Inference for Modeling Common Knowledge in Extended Dialogue

The main problem with a usable notion of CK is that what conversants infer under normal circumstances should become part of the common ground, but the mechanism of conversational based inference is not known. However conversation based inference does have a number of properties that distinguish it from other inferential systems. The question that I wish to address in this section is how CK is established and maintained over the course of an extended dialogue. This case of CK, the conversational record, as a subset of the common ground, would seem to be one of the simplest cases of evidenced assumptions. It would seem that conversational partners keep track of the information status of propositions that are assumed together and that the conversation provides evidence for. The trick is that conversational inferences must become a part of CK if they were intended by the speaker. But how do we determine what the conversational inferences are?

It is certainly beyond the scope of this paper to review the deductive systems that might be appropriate for conversational inference. Possibilities might be given by various modifications of standard logical rules[Kon85, JK79]. A number of authors have proposed that language provides it's own clues for the control of inferencing[JW81, Gro77, Sid79, Pri78]. In the following sections I will discuss what makes conversational inference special. Then I will focus on a particular kind of conversational inferences called presuppositions, and look at examples of some presuppositions in naturally occurring dialogues.

9.1 Conversational Inference

One striking aspect of conversational inference is that it is so fast. Less than 5% of speech is delivered in overlap, yet the pauses between speakers are rarely more than 250 milliseconds[Lev83][21].

[21] Most of this discussion is taken from Levinson [Lev85].

Not only is conversational inference fast, it must also be correct most of the time, since there is little time for multiple inferential attempts given the demands of the ongoing conversation. Some researchers who study language assume that general principles of inference used in problem solving or vision are the same ones used in conversational inference, but language based inferences have a special property in having been designed to have a single solution, somewhat like a puzzle.

First, let us consider some of the inferences that have been called Gricean implicatures. One kind, generated from Grice's QUANTITY MAXIM: *Make your contribution as informative as is required*, generate what have been called SCALAR IMPLICATURES. For instance

(11) I ate some of the cookies.

implicates I didn't eat all of them, ie I made the strongest statement I could. In contrast, the maxims also license implicatures that enrich the bare facts of the utterance.

(12) I turned the key and the engine started.

implicates that the engine started because I turned the key, although I didn't explicitly say so. Thus the maxims are sometimes in conflict. The first example assumes that the speaker made the strongest statement possible, while the second assumes that the hearer should read more into the utterance than the speaker said. One property of implicatures is that they are defeasible. Thus in 13

(13) A: How many ewes do you have?

B: Certainly 120.

the implicature of no more than 120 that would be expected from the MAXIM OF QUANTITY, can be cancelled by the additional premise 14:

(14) A is an agricultural inspector and B must have a minimum number of sheep to get a subsidy.

So the inferencing process must be defeasible. In addition, conversational inference often consists of adding new premises in order to reach a conclusion, as in systems of abductive inference such as in Hobbs[Hob86]. For instance:

(15) A: Where are my chocolates?

B: The dog is certainly sleeping very soundly.

Hobb's system depends on making plausible assumptions, but the interpretation of 15 depends on adding such premises as *Dogs eat chocolates* and *Dogs sleep soundly after eating chocolates.* It seems unlikely that these premises could come from background knowledge. As Levinson clearly points out, it is hard to imagine what kind of inferential process could yield a unique and determinate solution, rapidly, while supplying both a conclusion dependent on missing premises and missing premises dependent on the conclusion one is attempting to calculate.

Schelling's coordination games consist of situations in which participants are able to coordinate with minimal clues. Perhaps something similar goes on in language that would enable us to explain how communication of determinate inferences might be achievable from minimal linguistic cues. Levinson suggests that utterance interpretation is in effect a coordination problem in which the hearer may take it that the problem for the speaker is to choose, as a determinate clue to his communicative intentions, an utterance which will lead to successful coordination on the recovery of those intentions. However in coordination games the goals are mutually known and the means have to be coordinated on. In conversation, the means are given by the utterance and the goal must be coordinated on[Lev85].

Some authors have provided evidence that the syntactic structure of language controls inferencing by indicating which discourse entities the inferences should be about. Certain syntactic positions in utterances have more prominent roles than others[Pri85]. Since a speaker can choose to say something any number of ways, information as to which entity is being focused upon can be brought to bear to constrain inferencing.

Levinson suggests that the interactivity of language provides one means for people to coordinate on their contributions. Any failure to elicit the desired response, can be followed by a more explicit linguistic cue, until coordination is achieved. Reference repair is one example of this as was discussed in section 6. I want to suggest that the use of utterances that can be characterized as being redundant in standard information theoretic terms, is another way in which participants coordinate. In the next sections, I will look at some examples.

9.2 Presupposition

Seuren defines presuppositions as elements that are supposed to ensure that the information necessary for the interpretation of the utterance is stored in the conversational record before the utterance is interpreted. As such, these should be among the simplest kind of inferences that a discourse participant makes. Seuren notes that sometimes the presupposition will already be represented in the discourse representation because it has been uttered as a separate utterance, but most of the time it will be supplied posthoc[Seu88].

A particular type of presuppositions are what has been called existential presuppositions[Pri78]. According to Clark and Marshall any definite reference presupposes that the referent is in CK, or you can readily infer the existence of such an entity, thus definite referring expressions are commonly believed to presuppose the existence of their referents. As Prince notes, Kempson goes to great lengths to show that the phrase *the neighbours* in (16), has no presupposition.

(16) No the neighbours didn't break it - we don't have any neighbours.

But Prince claims that it does have the presupposition that is expected, namely that *the neighbours* do exist. But this is a particular person's presupposition. Namely whoever produced the previous utterance, such as A in the interchange below.

(17) A: Did your neighbours break the window?

B: No the neighbours didn't break it - we don't have any neighbours.

As Prince points out, there is a need to distinguish between the beliefs of the two participants. When B makes her utterance, she has added that A assumed the existence of neighbours, and this is what makes her reference felicitous. From a logical point of view, such a discourse should be contradictory. But it is not self-contradictory. If the use of the phrase *your neighbours* adds the fact that *B has neighbours* to the context, this fact must be able to be retracted, and the discourse context updated([Pri78], p 421).

Besides this anaphoric use of presupposition, Prince notes that presuppositions are often accompanied by inferencing instructions, that instruct the hearer to 'defer' attribution of a speaker belief. A simple case with respect to nouns is the use of an adjective such as *alleged.* The use of *If* in an *If, then* construction also instructs the hearer to defer attribution of the proposition in the if-clause.

For instance:

(18) If Jack's children are bald, then he has children.

In summary, it seems to be quite important, even with respect to presuppositions in a speaker's utterance, to distinguish two things: (1) Whose presupposition is being used at a particular point in the dialogue, (2) What the role of discourse markers such as *Either, or* and *If, then* are as guidelines of the hearer's inference processes. I will look at the use of this type of 'inference instruction' in the next section.

9.3 Redundancy and Interactivity

What I would like to do now is to push my hypothesis that one function of redundancy in language is to help discourse participants increase the strength of the evidence for CK. Another important aspect of interactive communication is the possibility for continually monitoring understanding and increasing the strength of a description, or checking an assumption with another participant.

In the discourse situation where the purpose of the conversation is the interactive transfer of expertise, the participants must ensure that the beliefs of the expert become CK, in just the right way to support action by the non-expert. If presuppositions are one of the least debatable types of inference, and if *if* is used to mark a proposition whose attribution should be delayed, we would not expect the proposition *the 2/3 is yours* to occur in utterance H3 below.

```
(from hg211.rno)
H1. alright how does the income break down?
D1. about 2/3 to 1/3
H2. the 2/3 i may i assume is yours?
D2. right
H3. if the 2/3 is yours, you can get 2/3 of the taxes and the interest
D3. will there be any problem in the future, let's say if her income should
    increase and the percentage changes.
```

While Prince never claims that this is the only reason for using *if* in discourse, the question of why a known proposition would be marked in this way is still there. The truth of the proposition certainly cannot be in question for the participants. Nor can the speaker H, be instructing the hearer D, to defer attribution of the proposition as a speaker belief, because D1 and H2 should have already established that both participants believe that *the 2/3 is D's.* The claim is that the

redundancy ensures that the action that is recommended becomes part of a common known set of actions for achieving the goal, in just the right way. The proposed action is contingent upon the world being a certain way; the contingent nature must be CK.

The other facility that conversation provides for achieving coordination is its interactivity. Consider H's first utterance, in which he marks *all of these are 6 month certificates*, goes on to say *i presume they are*, and then **waits**. E's reply *yes* ensures the CK status of the proposition, as well as the fact that E is attending to that particular point, and a general action description is forthcoming.

```
hg211.rno
H1. well it's difficult to tell because we're so far away from any of um --
    but i would suggest this
    -- if all of these are 6 month certificates and i presume they are
E1. yes
H2. then i would like to see you start spreading some of that money around
E2. uh hu
```

A much more extended example of the role of interactivity in combination with redundancy is given below. I don't intend to analyze this in detail, and I have no suggestions about how to handle this formally. But I would like to point out what I think is really striking. Consider the sequence of interchanges from H2 to M8 where M echoes *without penalty*, signalling her understanding by **repeating** at various points what H has said. They interrupt one another. M checks her understanding of M5, which was implied by H4, since there must be some switching around in order to get 2000 each. This reformulation is anticipated before its completion and confirmed at H5 and H6. Finally at M10, M summarizes what H has just told her throughout the extent of the previous exchange, and which she apparently had already been confirming as they went along. In H10, H confirms her summary, and affirms another fact that had only been implied previously. This kind of interactive additive phenomenon has been documented with respect to the production of referring expressions in Clark and Wilkes-Gibbes[CW86]. Typically a conversation provides many opportunities for monitoring understanding, clarification and hypothesis testing.

```
(hg222.rno)
M1. anyway what happens in the future if i do get a job where i could uh contribute
    to my own ira, the local bank person indicated that that would mean we would
    pay a penalty on anything that had ever been put in the spousal account
H1.  nope not so -- if you'll put in 22-50 right now --
M2. um hm
H2. then you go out, you get a job, and you earn let's say another 2000
M3. um hm
H3. you may put in another 17-50 in your account
M4. into my account
H4. right so that you have 2000 each
M5. so it could be switched out of
H5. it could be
M6. my own
H6. that is correct it could be moved around so that each of you  have 2000
M7. i see..
H7. without penalty
M8. without penalty -- and the fact that i have a an account of my own from a
```

```
    couple of years ago when i was working doesn't affect this at all
H8. oh no you can have as many ira's  as you wish you can .. as a matter of fact
    you can have 4 or 5 in one year or 10
M9. ok
H9. if you want to put 200 bucks into each of 10 institutions you can do that
.
.
M10. yeah -- ok so we can go ahead and do it to a spousal one and then change it
     later when
H10. right and remember you can split that spousal one any way you wish
M11. ok fine
H11. if indeed you intend to work it might be simpler if you split it on a
     2000 - 2-50 basis
M12. ok good idea
H12. all right?
```

The proposal here is that participants exploit the fact that there are going to be opportunities for further interaction and the potential for redundancy to ensure common knowledge. This is what allows the inferences to be made "so fast" in conversation. Repetitions and reformulations of one another's utterances, as a way of testing and refining knowledge about a task, are common. Many theories of how the common ground is established and maintained, suggest that one purpose of doing so is to ensure that you never tell the same person the same thing twice. Indeed, without looking at conversations like these, it is hard to believe that a participant would tell another what they just told you. However this may one way that experts and advisees transfer expertise.

10 Conclusion

This paper has discussed the motivation behind common knowledge. It is argued that it is necessary for joint action in general and for language use as a particular kind of joint action. However, I hope to have demonstrated that this term has been broadly interpreted. Most generally, common knowledge is used to describe the knowledge that is evidenced in reflexive reasoning, reasoning about other's knowledge and about their knowledge of your knowledge, etc. The term has also been used to refer to facts or objects which are mutually salient, since this will normally support reflexive mental attitudes. One of the main problems for a theory of common knowledge is whether knowledge is the appropriate mental attitude. It seems as though probabilistic beliefs might approximate the cognitive phenomenon more closely. Another problem that has been widely noted in the literature has to do with the circular nature of the characterization. My belief is that treating CK as deriving from an induction schema as Clark and Marshall did solves the problem. Barwise also showed that Aczel's theory of nonwellfounded sets can give a solid mathematical characterization of the circular situations which support common knowledge.

The main problem with a usable notion of CK is that inference must play a critical role in what exactly is CK. Conversational inference has a number of properties that distinguish it from other inferential systems, such as being apparently abductive and probabilistic, but a precise characterization of it is an unsolved problem. I suggest that in cases where ensuring CK really matters, participants in dialogue exploit opportunities for redundancy to do so.

11 Acknowledgements

I would like to thank the members of my WPE II reviewing committee: Susan Davidson, Aravind Joshi, and Scott Weinstein, and others who read drafts or participated in discussions (both synchronous and asynchronous) while I was writing this: Steve Whittaker, Phil Stenton, Mark Gawron, Barbara DiEugenio, Lew Creary, Owen Rambow, and Herb Clark. This paper is much better than it would have been without their comments.

References

[Acz88] Peter Aczel. *Non-Well-Founded Sets.* CSLI, 1988.

[Bar88] Jon Barwise. *The situation in Logic-IV: On the Model Theory of Common Knowledge.* Technical Report No. 122, CSLI, 1988. Chapter 8 of The situation in Logic, also as Three Views of Common Knowledge in TARK2.

[BH79] Kenneth Bach and R. M. Harnish. *Linguistic Communication and Speech Acts.* MIT Press, 1979.

[CC82] Herbert H. Clark and T. B. Carlson. Speech acts and hearer's beliefs. In Neil V. Smith, editor, *Mutual Knowledge*, pages 1–37, Academic Press, New York, New York, 1982.

[Cla82] Herbert H. Clark. The relevance of common ground: comments on sperber and wilson's paper. In Neil V. Smith, editor, *Mutual Knowledge*, pages 124–127, Academic Press, New York, New York, 1982.

[CM81] Herbert H. Clark and Catherine R. Marshall. Definite reference and mutual knowledge. In Aravind K. Joshi, Bonnie Lynn Webber, and Ivan Sag, editors, *Elements of Discourse Understanding*, pages 10–63, Cambridge University Press, Cambridge, 1981.

[Coh84] Phillip R. Cohen. The pragmatics of referring and the modality of communication. *Computational Linguistics*, 10:97–146, 1984.

[CW86] Herbert H. Clark and Deanna Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22:1–39, 1986.

[FP81] J. A. Fodor and Z. W. Pylyshyn. How direct is visual perception?: some reflections on gibson's "ecological approach"*. *Cognition*, 9:139–196, 1981.

[Gal89] Julia R. Galliers. *A Theoretical Framework for Computer Models of Cooperative Dialogue, Acknowledging Multi Agent Conflict.* Technical Report 172, University of Cambridge, Computer Laboratory, New Museums Site, Pembroke St. Cambridge England CB2 3QG, 1989.

[Gal90] Julia R. Galliers. *Belief Revision and a Theory of Communication.* Technical Report 193, University of Cambridge, Computer Laboratory, New Museums Site, Pembroke St. Cambridge England CB2 3QG, 1990.

[GG82] Gerald Gazdar and David Good. On a notion of relevance*; comments on sperber and wilson's paper. In Neil V. Smith, editor, *Mutual Knowledge*, pages 88–100, Academic Press, New York, New York, 1982.

[GJW86] Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. Towards a computational theory of discourse interpretation. 1986. Unpublished Manuscript.

[Gri57] H. P. Grice. Meaning. *Philosophical Review*, LXVI, No. 3:377–388, 1957.

[Gri82] Paul Grice. Meaning revisited. In Neil V. Smith, editor, *Mutual Knowledge*, pages 223–245, Academic Press, New York, New York, 1982.

[Gro77] Barbara J. Grosz. *The Representation and Use of Focus in Dialogue Understanding.* Technical Report 151, SRI International, 333 Ravenswood Ave, Menlo Park, Ca. 94025, 1977.

[Har77] Gilbert Harman. Review of linguistic behaviour by j bennet. *Language*, 53:417–424, 1977.

[HM84] Joseph Halpern and Yoram Moses. Knowledge and common knowledge in a distributed environment. In *Proceedings of the Third Annual ACM Symposium on Principles of Distributed Computing*, pages 50–61, 1984.

[HM85] Joseph Halpern and Yoram Moses. A guide to the modal logics of knowledge and belief. In *Proc. International Joint Conference on Artificial Intelligence, Los Angeles*, pages 480–490, 1985.

[Hob86] Jerry R. Hobbs. *Discourse and Inference.* Technical Report, SRI International, 333 Ravenswood Ave., Menlo Park, Ca 94025, 1986.

[JK79] Aravind K. Joshi and Steve Kuhn. Centered logic: the role of entity centered sentence representation in natural language inferencing. In *Proc. International Joint Conference on Artificial Intelligence*, page pp., 1979.

[Jos82] Aravind K. Joshi. Mutual beliefs in question-answer systems. In Neil V. Smith, editor, *Mutual Knowledge*, pages 181–199, Academic Press, New York, New York, 1982.

[JW81] Aravind K. Joshi and Scott Weinstein. Control of inference: role of some aspects of discourse structure - centering. In *Proc. International Joint Conference on Artificial Intelligence*, pages pp. 385–387, 1981.

[JWW86] Aravind K. Joshi, Bonnie Lynn Webber, and Ralph M. Weischedel. *Some Aspects of Default Reasoning in Interactive Discourse.* Technical Report MS-CIS-86-27, University of Pennsylvania, Department of Computer and Information Science, 1986.

[KB67] R. M. Krauss and P.D. Bricker. Effects of transmission delay and access delay on the efficiency of verbal communication. *The journal of the Acoustical Society of America*, 41(2), 1967.

[Kon85] Kurt Konolige. Belief and incompleteness. In Jerry R. Hobbs and Robert C. Moore, editors, *Formal Theories of the Commonsense World*, pages 359–403, Ablex Publishing, 1985.

[Lev83] Stephen C. Levinson. *Pragmatics.* Cambridge University Press, 1983.

[Lev84] Hector Levesque. A logic of implicit and explicit belief. In *Proc. National Conference on Artificial Intelligence, Austin*, pages 198–202, 1984.

[Lev85] Stephen C. Levinson. What's special about conversational inference. In *1987 Linguistics Institute Packet*, 1985. In packet for 1987 Linguistics Institute.

[Lew69] David Lewis. *Convention.* Harvard University Press, 1969.

[Lew79] David Lewis. Scorekeeping in a language game. *Journal of Philosophical Logic*, 8:339–359, 1979.

[MMO90] Michael W. Mislove, Lawrence S. Moss, and Frank J. Oles. A different information ordering for situation theory. 1990. Manuscript.

[Moo85] Robert C. Moore. A formal theory of knowledge and action. In Jerry R. Hobbs and Robert C. Moore, editors, *Formal Theories of the Commonsense World*, pages 319–358, Ablex Publishing, 1985.

[MSHI78] John McCarthy, M. Sato, T. Hayashi, and S. Igarashi. *On the Model Theory of KNowledge.* Technical Report AIM-312, CS78-657, Stanford University, 1978.

[NJ83] Gopalan Nadathur and Aravind K. Joshi. Mutual beliefs in conversational systems: their role in referring expressions. In *Proc. International Joint Conference on Artificial Intelligence, Austin*, 1983.

[OC89] Sharon L. Oviatt and Philip R. Cohen. The effects of interaction on spoken discourse. In *Proc. 27th Annual Meeting of the Association of Computational Linguistics*, pages 126–134, 1989.

[Par90] Rohit Parikh. *Recent Issues in Reasoning about Knowledge.* Technical Report , Brooklyn College of CUNY, 1990.

[PC81] C. Raymond Perrault and Philip R. Cohen. It's for your own good: a note on inaccurate reference. In Aravind Joshi, Bonnie Webber, and Ivan Sag, editors, *Elements of Discourse Understanding*, pages 217–230, Cambridge University Press, Cambridge, 1981.

[Pow84] Richard Power. Mutual intention. *Journal for the Theory of Social Behaviour*, 14, 1984.

[Pri78] Ellen F. Prince. On the function of existential presupposition in discourse. In *Papers from 14th Regional Meeting*, CLS, Chicago, IL, 1978.

[Pri81] Ellen F. Prince. Toward a taxonomy of given-new information. In *Radical Pragmatics*, Academic Press, 1981.

[Pri85] Ellen F. Prince. Fancy syntax and shared knowledge. *Journal of Pragmatics*, pp. 65–81, 1985.

[Rev87] Peer Review. Precis of relevance and open peer commentary. *Behavioural and Brain Science*, 10:pp. 697–754, 1987.

[Sch60] Thomas C. Schelling. *The Strategy of Conflict.* Harvard University Press, 1960.

[Sch72] Stephen R. Schiffer. *Meaning.* Clarendon Press, 1972.

[Seu88] Pieter A.M. Seuren. The self-styling of relevance theory. *Journal of Semantics*, 5:123–143, 1988.

[Sid79] Candace L. Sidner. *Toward a computational theory of definite anaphora comprehension in English.* Technical Report AI-TR-537, MIT, 1979.

[Sta78] Robert C. Stalnaker. Assertion. In Peter Cole, editor, *Syntax and Semantics, Volume 9*, pages 315–332, Academic Press, 1978.

[SW82] Dan Sperber and Deidre Wilson. Mutual knowledge and relevance in theories of comprehension. In Neil V. Smith, editor, *Mutual Knowledge*, pages 61–87, Academic Press, New York, New York, 1982.

[SW86] Dan Sperber and Deidre Wilson. *Relevance, Communication and Cognition.* Basil Blackwell, Oxford, 1986.

[TW86] Tommy Chin-Chiu Tan and S. R. Werlang. Summary of on aumann's notion of common knowledge -an alternative approach. In *Proc. of Theoretical Aspects of Reasoning about Knowledge, 1986*, pages 253–258, 1986.

[Web78] Bonnie Lynn Webber. *A Formal Approach to Discourse Anaphora.* PhD thesis, Harvard University, 1978.