Chapter 37
Three Views of Common Knowledge

Jon Barwise

Introduction

As the pioneering work of Dretske[1] has shown, knowing, believing, and having information are closely related and are profitably studied together. Thus while the title of this paper mentions common knowledge, I really have in mind the family of related notions including common knowledge, mutual belief and shared information. Even though I discuss common knowledge in this introduction, the discussion is really intended to apply to all three notions.

Common knowledge and its relatives have been written about from a wide variety of perspectives, including psychology, economics, game theory, computer science, the theory of convention, deterrence theory, the study of human-machine interaction, and the famous Conway paradox, just to mention a few. There are literally hundreds of papers that touch on the topic. However, while common knowledge is widely recognized to be an important phenomenon, there is no agreement as to just what it amounts to. Or rather, as we will see, what agreement there is presupposes a set of simplifying assumptions that are completely unrealistic. This paper offers a comparison of three competing views in a context which does not presuppose them to be equivalent, and explores their relationships in this context.[2]

Jon Barwise was deceased at the time of publication.

[1] F. Dretske, *Knowledge and the Flow of Information* (Cambridge, Mass.: Bradford Books/MIT Press, 1981).

[2] Obviously I have not read all, or even most, of the papers on common knowledge, so it could be that some or all of the points made in this paper are made elsewhere. If so, I would appreciate

J. Barwise (deceased)
Stanford University, Stanford, CA, USA

H. Arló-Costa et al. (eds.), *Readings in Formal Epistemology*, Springer Graduate Texts in Philosophy 1, DOI 10.1007/978-3-319-20451-2_37

I take it that these accounts are after characterizations of common knowledge in terms of ordinary knowledge, of mutual belief in terms of belief, and of having shared information in terms of having information. Such accounts should be compatible with, but presumably distinct from, an account that shows how it is that common knowledge comes about. They should also be compatible with some explanation of how common knowledge is used.

We are going to compare the following approaches to common knowledge: (1) the *iterate* approach, (2) the *fixed-point* approach, and (3) the *shared-environment* approach. In order to review these three accounts, let's consider a special case where there are just two agents, say p and q, with common knowledge of some fact σ. Let τ be this additional fact, of the common knowledge of σ. We are looking for a characterization of τ in terms of p, q, σ and ordinary (private) knowledge.

By far the most common view of common knowledge is that τ is to be understood in terms of *iterated* knowledge of σ: p knows σ, q knows σ, p knows q knows σ, q knows p knows, p knows q knows p knows σ, and so forth. On this account, for p and q to have common knowledge of σ is for all members of this infinite collection of other facts to obtain. This is the approach taken in David Lewis' influential book[3] on convention, for example. It is, without doubt, the orthodox account, at least in the field of logic. It is, for example, the one that is the basis of the mathematical modeling of common knowledge in the logic of distributed systems.[4]

The two other accounts we want to investigate replace this infinite hierarchy with some sort of circularity. One such account was explicitly proposed by Harman.[5] Harman's proposal is that the correct analysis of τ is as:

$$p \text{ and } q \text{ know } (\sigma \text{ and } \tau)$$

Notice that on this fixed-point account, τ is in some sense a proper constituent of itself. Harman seems to suggest that this is nothing but a succinct representation of the first infinite hierarchy.

This fixed point approach is also the view of common knowledge that is implicit in Aumann's pioneering paper modeling common knowledge in game theory, as

learning about it. But even if this is so, I am reasonably sure that the particular model I develop below is original, depending as it does on recent work in set theory by Peter Aczel.

[3]David Lewis, *Convention*, *A Philosophical Study* (Cambridge, Mass.: Harvard University Press, 1969).

[4]See, for example, the paper by Halpern and Moses, "Knowledge and common knowledge in distributed environments," Proc. 3rd ACM Symp. on Principles of Distributed Computing (1984), 50–61, and the paper by Fagin, Halpern and Vardi, "A model-theoretic analysis of knowledge: preliminary report," Proc. 25th IEEE Symposium on Foundations of C.S., 268–278.

[5]See Gilbert Harman's review of *Linguistic Behavior* by Jonathan Bennett, *Language* 53 (1977): 417–24.

was pointed out by Tommy Tan and Sergio Ribeiro da Costa Werlang.[6] Aumann suggests that this approach is equivalent to the iterate approach. Tan and Ribeiro da Costa Werlang develop a mathematical model of the iterate approach and show that it is equivalent to Aumann's fixed point model. Similarly, one sees from the work of Halpern and Moses, that while they start with the iterate approach, in their set-up, this is equivalent to a fixed point. One of the aims of this paper is to develop a mathematical model where both iterate and fixed point accounts fit naturally, but where they are *not* equivalent. Only in such a framework can we explicitly isolate the assumptions that are needed to show them equivalent. We will see that these assumptions are simply false (in the case of knowledge), so that the issue as to which of the two, if either, is the "right" analysis of the notion is a live one.

The final approach we wish to discuss, the shared-environment approach, was proposed by Clark and Marshall,[7] in response to the enormous processing problems associated with the iterate account. On their account, p and q have common knowledge of σ just in case there is a situation s such that:

- $s \models \sigma$,
- $s \models p_1$ knows s,
- $s \models p_2$ knows s.

Here $s \models \theta$ is a notation for: θ is a fact of s. The intuitive idea is that common knowledge amounts to perception or other awareness of some situation, part of which includes the fact in question, but another part of which includes the very awarenesses of the situation by both agents. Again we note the circular nature of the characterization.

What Are We Modeling: Knowing or Having Information?

It is these three characterizations of common knowledge, and their relatives for the other notions of mutual belief and shared information, that we wish to compare. Among common knowledge, mutual belief, and shared information, we focus primarily on the case of having information, secondarily on the case of knowledge. Part of the claim of the paper is that these two notions are often conflated, and that it is this conflation that lends some credibility to the assumptions under which the first two approaches to common knowledge are equivalent. So I need to make clear

[6] R. J. Aumann, "Agreeing to disagree," *Annals of Statistics*, 4 (1976), 1236–1239, and the working paper "On Aumann's Notion of Common Knowledge – An alternative approach," Tan and Ribeiro da Costa Werlang. University of Chicago Graduate School of Business, 1986.

[7] H. Clark and C. Marshall, "Definite reference and mutual knowledge," in *Elements of Discourse Understanding*, ed. A. Joshi, B. Webber, and I. Sag (Cambridge: Cambridge University Press, 1981), 10–63.

what I take to be the difference between an agent p knowing some fact σ, and the agent simply having the information σ.

Here I am in agreement with Dretske.[8] Knowing σ is stronger than having the information σ. An agent knows σ if he not only has the information σ, but moreover, the information is "had" in a way that is tied up with the agent's abilities to act. When might this not be the case? The most notorious example (and by no means the only one) is when I know one fact σ, and another fact σ' logically follows from σ, but I disbelieve the latter because I don't know that the one follows from the other. Obviously there is a clear sense in which I have the information σ', but I certainly don't know it in the ordinary sense of the word. Another arises with certain forms of perceptual information. If I see the tallest spy hide a letter under a rock, then there is a clear sense in which I have the information the tallest spy has hidden the letter. However, if I don't know that he is a spy, say, then I don't know that the tallest spy has hidden a letter. Information travels at the speed of logic, genuine knowledge only travels at the speed of cognition and inference.

Much of the work in logic which seems to be about knowledge is best understood in terms of having information. And for good reason. For example, in dealing with computers, there is a good reason for our interest in the latter notion. We often use computers as information processors, after all, for our own ends. We are often interested less in what the computer does with the information it has, than in just what information it has and what we can do with it. Or, in the design of a robot, we may be aiming at getting the robot to behave in a way that is appropriate given the information it has. One might say, we are trying to make it know the information it has.

So, as noted earlier, this paper focuses primarily on the case of having information. The model I am going to develop originated with an analysis of shared perceptual information,[9] but it also works quite well for primary epistemic perceptional information[10] and the relation of having information.

Let me say all this in another way, since it seems to be a confusing point. In the section that follows, I could interpret the model as a model of knowledge if I were to make the same idealization that is made in most of the literature on common knowledge. However, part of what I want to do here is make very explicit just what the role of this idealization is in the modeling of common knowledge. Thus, I am forced to work in a context where we do not make it. Once we are clear about its role, we can then decide if we want to make it.

[8] Op. Cit.

[9] See ch. 2 of Fred Dretske, *Seeing and Knowing* (Chicago: University of Chicago Press, 1969); or J. Barwise, "Scenes and other Situations", *Journal of Philosophical Logic* 78 (1981): 369–97; or ch. 8 of J. Barwise and J. Perry, *Situations and Attitudes* (Cambridge, Mass.: Bradford Books/MIT Press, 1983).

[10] See ch. 3 of *Seeing and Knowing* or ch. 9 of *Situations and Attitudes.*

Summary of Results

Our results suggest that the fixed point approach gives the right theoretical analysis of the pretheortic notion of common knowledge. On the other hand, the shared-environment approach is the right way to understand how common knowledge usually arises and is maintained over an extended interaction. It does not offer an adequate characterization of the pretheoretic notion, though, since a given piece of common knowledge may arise from many different kinds of shared environments. The fixed point gets at just what is in common to the various ways a given piece of common knowledge can arise.

What about the iterate approach? We will show that for the relation of having information, the fixed-point approach is equivalent to the iterate approach, provided we restrict ourselves to finite situations. Without this assumption, though, the iterate approach, with only countably many iterations, is far too weak. In general, we must iterate on indefinitely into the transfinite.

Not only is the iterate approach too weak. When we move from having information to knowing, then even two iterations are unjustified. In general, the iterate approach is incomparable and really seems to miss the mark. We will see just what assumptions *are* needed to guarantee that the iterate account is equivalent to the fixed-point account.

Modeling Shared Information

In developing our model, we will follow the general line used in *The Liar*[11] in three ways. First, we take our metatheory to be ZF/AFA, a theory of sets that admits of circularity. We do this because ZF/AFA offers the most elegant mathematical setting we know for modeling circularity. Space does not permit us to give an introduction to this elegant set theory. We refer the reader to Chap. 3 of this book, or to Aczel's lectures[12] for an introduction.

Second, we follow the approach taken in *The Liar* in paying special attention to "situations," or "partial possible worlds." As far as this paper goes, the reader can think of a situation as simply representing an arbitrary set of basic facts, where a fact is simply some objects standing in some relation. Actually, in this paper, situations play a dual role. On the one hand they represent parts of the world. On the other hand they represent information about parts of the world. Thus, for example, we will define what it means for one situation s_0 to support another situation s_1, in the sense that s_0 contains enough facts to support all the facts in s_1.

[11] J. Barwise and J. Etchemendy, *The Liar: An Essay on Truth and Circularity* (New York: Oxford University Press, 1987).

[12] P. Aczel, *Non-well-founded Sets* (CSLI Lecture Notes (Chicago: University of Chicago Press, 1987 (to appear)).

Finally, on the trivial side, we also follow *The Liar* in considering a domain of card players as our domain to be modeled. We use this domain because it is simple, and because the existence of common knowledge is absolutely transparent to anyone who has ever played stud poker. And while the example is simple, there is enough complexity to illustrate many of the general points that need making. However, there is nothing about the results that depend on this assumption. You could replace the relation of having a given card with any relation whatsoever, and the results would still obtain.

Example 37.1 Simply by way of illustration, we have a running example, a game of stud poker. To make it very simple, we will use two card stud poker,[13] with two players, Claire and Max. We will assume that the players have the following cards:

Player	Down card	Up card
Claire	$A\spadesuit$	$3\clubsuit$
Max	$3\spadesuit$	$3\diamond$

Except for the rules and the idiosyncrasies of the other players, all the information available to the players is represented in this table. Note that based on what he sees, Max knows that he has the winning hand, or at least a tie, but Claire thinks she has a good chance of having the winning hand. The question before us is how best to model the informational difference between up cards and down cards.

Notice how different this situation would be from draw poker, where all cards are down, even if each player had cheated and learned the value of the second card. Anyone who has played poker will realize the vast difference. The reason is that in the standard case, the values of all the up cards is common knowledge, but in the second it isn't. Our aim, then, is to use tools from logic to model the three approaches to the common knowledge and shared information present in such a situation.

We reiterate that we use this simple card domain simply by way of making things concrete. We could equally well treat the more general case, if space permitted. We use S for the relation of seeing (or more generally of having information), H for the relation of having a card, and appropriate tuples to represent facts involving these relations. Thus, the fact that Max has the $3\spadesuit$ will be represented by the triple $\langle H, \text{Max}, 3\spadesuit\rangle$. The fact that Claire sees this will be represented by $\langle S, \text{Claire}, \{\langle H,$

[13]For the reader unfamiliar with two card stud poker, here is all you need to know to follow the example. First each player is dealt one card which only he is allowed to see, and there is a round of betting. Then each player is dealt one card face up on the table and there is another round of betting. Hands are ranked and players bet if they think their hand is best. But they can also drop out of the round at any point. After both rounds of betting are over, the hands are displayed, so that all players can see who won. As far as the ranking, all that matters is that a hand with a matching pair is better than a hand with no pairs. But among hands with no pairs, a hand with an ace is better than a hand with no ace.

Max, 3♠⟩}⟩. The question is how to adequately represent the common knowledge, or public information, of the up cards, like the fact ⟨*H*, Max, 3♢⟩ that Max has the 3♢. Thus for our formal development we have primitives: players $p_1, \ldots, p_n$, cards A♠, K♠, ..., 2♣, and relations *H* for the relation of having some card and *S* for the relation of seeing or otherwise having the information contained in some situation.

Comparing the Iterate and Fixed Point Accounts

Definition 37.1

1. The (models of) *situations* and *facts*[14] form the largest classes *SIT*, *FACT* such that:
   - $\sigma \in FACT$ iff σ is a triple, either of the form $\langle H, p, c\rangle$, where p is a player and c is a card, or of the form $\langle S, p, s\rangle$, where p is a player and $s \in SIT$.
   - A set s is in *SIT* iff $s \subseteq FACT$.
2. The *wellfounded situations* and *wellfounded facts* form the smallest classes *Wf-SIT* and *Wf-FACT* satisfying the above conditions.

Routine monotonicity considerations suffice to show that there are indeed largest and smallest such collections. If our working metatheory were ordinary ZF set theory, then these two definitions would collapse into a single one. However, working in ZF/AFA, there are many nonwellfounded situations and facts. A fact $\sigma = \langle R, a, b\rangle$ being in some situation s represents the fact of the relation R holding of the pair a, b in s, and is said to be a *fact of s*.

Example 37.1, Cont'd The basic situation s_0 about which player has which cards is represented by the following situation: $s_0 =$

$$\{\langle H, \text{Claire}, A\spadesuit\rangle\ ,\ \langle H, \text{Max}, 3\diamondsuit\rangle\ ,\ \langle H, \text{Claire}, 3\clubsuit\rangle\ ,\ \langle H, \text{Max}, 3\spadesuit\rangle\}$$

Abbreviations We sometimes write $(p_i\ Hc)$ for the fact $\langle H, p_i, c\rangle$, and similarly $(p_i\ Ss)$ for the fact $\langle S, p_i, s\rangle$. We write $(p_i\ S\sigma)$ for $(p_i\ Ss)$ where $s = \{\sigma\}$. All of our facts are atomic facts. However, our situations are like conjunctive facts. Hence we sometimes write $\sigma \wedge \tau$ for the situation $s = \{\sigma, \tau\}$, and so we can write $(p_i\ S\ (\sigma \wedge \tau))$ for $p_i\ Ss$, where $s = \{\sigma, \tau\}$. Similarly when there are more conjuncts.

Example 37.1, Cont'd With these tools and abbreviations, we can discuss the first two approaches to the public information about the up cards in our example. Toward this end, let $s_u =$

$$\{\langle H, \text{Claire}, 3\clubsuit\rangle\ ,\ \langle H, \text{Max}, 3\diamondsuit\rangle\}$$

which represents situation concerning the up cards.

[14] In order to keep this paper within bounds, I am restricting attention only to positive, nondisjunctive facts.

Iterates On this account, the fact that s_u is public information would be represented by an infinite number of distinct wellfounded facts: (Claire Ss_u), (Max Ss_u), Claire S(Claire Ss_u), (Max S(Claire Ss_u)), etc., in other words, by a wellfounded though infinite situation.

Fixed-Point On this account, the fact that s_u is publicly perceived by our players can be represented by the following public situation s_p:

$$s_p = \left\{\text{Claire } S\left(s_u \cup s_p\right), (\text{Max } S\left(s_u \cup s_p\right)\right\}$$

By contrast with the iterate approach, this situation contains just two facts. However, it is circular and so not wellfounded. The Solution Lemma of ZF/AFA guarantees that the sets used to represent the situation s_p exists.

It will be useful for later purposes to have a notation for some of the situations that play a role in our example. First, let the situations s_1, s_2 represent the visual situations, as seen by each of Claire and Max, respectively, including both the up cards and what each sees about what the others see. Consider also the larger situation s_w that represents the whole. Let $s_w = s_0$ (from above) union the set of the following facts:

$$\langle S, \text{Claire}, (\text{Claire } HA\spadesuit)\rangle, \langle S, \text{Max}, (\text{Max } H3\spadesuit)\rangle,$$
$$\langle S, \text{Claire}, s_1\rangle, \langle S, \text{Max}, s_2\rangle$$

where the first two facts represent what each player sees about his own down cards, and, e.g., s_1 is everything relevant seen by Claire, with facts s_u (= the "up" cards, as above) plus the fact (S, Max, s_2). Notice that s_1 is a constituent of s_2, and vice versa, so that s_w is a circular, nonwellfounded situation.

The next task is to define what it means for a fact a to hold in a situation s, which we write $s \models \sigma$, so that we can show that the situation s_w does satisfy the fixed point fact situation s_p defined above, as well as the above iterates.

Definition 37.2 The relation $\models$ is the largest subclass of $SIT \times FACT$ satisfying the following conditions:

- s $\models (pHc)$ iff $\langle H, p, c\rangle \in s$
- $s \models (pSs_0)$ iff there is an s_1 such that $\langle S, p, s_1\rangle \in s$, and for each $\sigma \in s_0$, $s_1 \models \sigma$.

The motivation for the second clause should be fairly obvious. If, in s, a player p sees (or otherwise has the information) s_1, and if s_1 satisfies each $\sigma \in s_0$, then in s that same player p sees (or otherwise has the information) s_0. This would not be a reasonable assumption about the usual notion of knowledge, since knowledge is not closed under logical entailment.

There is a difference with the possible worlds approach that sometimes seems puzzling to someone familiar with the traditional modal approach to knowledge. In p.w. semantics, partial situations are represented by the set of all possible worlds

compatible with them. As a result, whereas we can use an *existential* quantifier in clause (2) of our definition over situations about which p has information, the p.w. approach is forced to use a universal quantifier over possible worlds.

The reader can verify that all of the facts of s_p and the hierarchy of iterates of our running example indeed hold in the situation s_w. We also note that it follows from the definition that for all facts σ, if $\sigma \in s$ then $s \models \sigma$. However, the converse does not hold.

We extend our notation a bit and write $s_1 \models s_2$ provided $s_1 \models \sigma$ for each $\sigma \in s_2$.

As a companion of this notion of holding in, there is a notion of hereditary subsituation.[15] Intuitively, s_1 is a hereditary subsituation of s_2, written $s_1 \sqsubseteq s_2$, if all the information present in s_1 is present in s_2.

Definition 37.3 The hereditary subsituation relation $\sqsubseteq$ is the largest relation on $SIT \times SIT$ satisfying: $s_1 \sqsubseteq s_2$ iff:

- If $\langle H, p, c\rangle \in s_1$, then $\langle H, p, c\rangle \in s_2$;
- If $\langle S, p, s_0\rangle \in s_1$, then there is an s such that $s_0 \sqsubseteq s$ and $\langle S, p, s\rangle \in s_2$.

Proposition 37.1

1. *If* $s_1 \models \sigma$ *and* $s_1 \sqsubseteq s_2$, *then* $s_2 \models \sigma$.
2. *For all situations* s_0 *and* s_1, *the following are equivalent:*

   (a) $s_1 \sqsubseteq s_2$
   (b) $s_2 \models \sigma$ *for each* $\sigma \in s_1$.

Proof Limitations of space in this volume prevent us from doing more than hint at the proofs of the results in this paper. In this case we note that (1) is a simple consequence of the maximality of the $\models$ relation. Likewise, the implication from (2a) to (2b) is a consequence of the maximality of the $\models$ relation. The converse is a simple consequence of the maximality of the $\sqsubseteq$ relation. □

We say that situations s_0, s_1 are *informationally equivalent*, $s_0 \equiv s_1$, if the same facts hold in them. This is clearly an equivalence relation on situations. By the above lemma, $s_0 \equiv s_1$ if and only if each is a hereditary subsituation of the other. Distinct situations are often informationally equivalent. For example, suppose s_0' is a proper subset of the set s_1' of facts. Consider the situation $s_1 = \{\langle S, \text{Max}, s_1\rangle\}$, where Max has the information s_1, with the situation s_0 where there are two facts, that Max has the information s_0' and that he has the information s_1'. Using the fact just mentioned it is clear that $s_0 \equiv s_1$.

To compare the iterate and the fixed point approaches, we will show how an arbitrary fact θ (or situation s) gives rise to a transfinite sequence of wellfounded facts θ^α (or wellfounded situations s^α), for arbitrary ordinal α, finite or infinite. We use Tr for the conjunction of the empty situation, a fact that holds in every situation.

[15] In more recent joint work with Aczel, a generalization of this relation takes center stage.

Definition 37.4 The transfinite sequence $\langle \theta^\alpha \mid \alpha \in \mathit{Ordinals} \rangle$ of wellfounded facts associated with an arbitrary fact θ is defined by induction on ordinals as follows: for any θ, $\theta^0 = Tr$, and for $\alpha > 0$ we have:

$$(pHc)^\alpha = (pHc)$$
$$(pSs)^\alpha = (pS\ \ s^{<\alpha})$$

where

$$s^{<\alpha} = \left\{ \sigma^\beta \,\middle|\, \sigma \in s, \beta < \alpha \right\}$$

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

The reader should verify that if we apply this definition to the fixed point fact in our example, we generate the iterates for all the finite ordinals, but then we go on beyond them into the transfinite.

We say that a fact σ *entails* a fact τ, written $\sigma \Rightarrow \tau$, if for every situation s, if $s \models \sigma$ then $s \models \tau$.

Theorem 37.2 Let θ be some fact.

1. *For all* α, $\theta \Rightarrow \theta^\alpha$.
2. *If each approximation fad* θ^α *holds in a situation* s, *then so does* θ.
3. *Assume that* κ *is a regular cardinal, and that* s *is a situation of size less than* κ. *If each approximation* θ^α, *for* $\alpha < k$, *holds in* s, *then so does* θ.

Proof The first is proved by means of a routine induction on α. The second is a consequence of the maximality of $\models$ and is not too difficult to prove. The third is a strengthening of the second involving routine cardinality considerations. □

Corollary 37.3 *Let* θ *be any fact, and let* s_w *be the set of all finite approximations of* θ. *Then, for any finite situation* s, $s \models \theta$ *iff* $s \models s_w$.

Refinement (3) of (2) of Theorem 37.2, and so the above corollary, were not present in the original working paper referred to above. They were discovered later in joint work with Peter Aczel. This result shows that the finite approximations of a circular fact will be equivalent to it, with respect to *finite* situations. This is a bit unsatisfactory, since the iterates themselves form an infinite situation. Still, it is the best we can hope for. However, in general, when we drop this restriction to finite models, one must look at the whole transfinite sequence of approximations. No initial segment is enough, as simple examples show. In this sense, the usual iterate approach is actually weaker than the simpler fixed-point approach.

When we move from having shared information to knowing, additional considerations must be brought to bear, as we will see below.

Comparing the Fixed Point and Shared Environment Approaches

To compare the shared environment approach with the fixed point approach, we introduce a simple second-order language which allows us to make existential claims about situations of just the kind made in the shared environment approach. We call the statements of this language ∃-statements. Before giving the definition, let's give an example. The following ∃-statement

$$\exists e\,[e = ((\text{Claire } H3\clubsuit) \wedge (\text{Claire } S\ \ e) \wedge (\text{Max } S\ \ e))]$$

is one shared environment analysis of the fact that Claire and Max share the information that Claire has the 3♣. Notice that what we have here is a simple, finite, wellfounded statement, but one that could only hold of nonwellfounded situations. Similarly, there is a fairly simple ∃-statement explicitly describing the situation s_w in our running example.

To define our language, we introduce variables $e_1, e_2, \ldots$ ranging over situations, in addition to constants for the cards and players. In fact, we do not bother to distinguish between a card or player and the constant used to denote it in statements. For atomic statements we have those of the form $(p_i Hc)$ (where P_i is a player and c is a card) and $(p_i Se_j)$. The set of ∃-statements forms the smallest set containing these atomic statements and closed under conjunction (∧), existential quantification over situations $(\exists e_j)$ and the rule: if Φ is a statement so is $(e_j \models \Phi)$. We are thus using ⊨ both for a relation symbol of our little language, as well as a symbol in our metalanguage. No more confusion should result from this than from the similar use of constants for cards and people. Finally, given any function f which assigns situations to variables, we define what it means for a statement Φ to hold in a situation s relative to f, written $s \models \Phi[f]$, in the expected way.

Definition 37.5

1. If Φ is an atomic statement, then $s \models \Phi[f]$ iff the appropriate fact is an element of s. In particular, if Φ is $(p_i Se_j)$, then $s \models \Phi[f]$ iff $\langle S, p_i, f(e_j)\rangle \in s$.
2. If Φ is $\Phi_1 \wedge \Phi_2$ then $s \models \Phi[f]$ iff $s \models \Phi_1[f]$ and $s \models \Phi_2[f]$
3. If Φ is $\exists e_j \Phi_0$ 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 Φ is $(e_j \models \Phi_0)$ then $s \models \Phi[f]$ iff the situation $s_j = f(e_j)$ satisfies $s_j \models \Phi_0[f]$.

A *closed* ∃-statement is one with no free variables, as usual. If Φ is closed, we write $s \models \Phi$ if some (equivalently, every) assignment f satisfies $s \models \Phi[f]$.

Notice that the ∃-statements are all finite and wellfounded. (The results that follow would hold equally well if we allowed infinite conjunctions and infinite strings of quantifiers, except for the word "finite" in Theorem 37.5 below.) Nevertheless, some of them can only hold of nonwellfounded situations, as the above example shows.

We want to show that any ∃-statement can be approximated in a certain sense by a fixed point situation. In particular, if we take as our ∃-statement one that expresses a shared environment approach to shared information, the resulting situation will be the one that characterizes the fixed point approach. Then, using the transfinite wellfounded iterates approximating the fixed point approach, we obtain a transfinite sequence of wellfounded facts approximating any ∃-statement.

Let us say that a situation s_Φ almost characterizes the ∃-statement Φ if $s_\Phi \models \Phi$ and for every situation $s \models \Phi$, we have $s \models s_\Phi$. For example, if we take our above example of an ∃-statement, then the following situation can easily be seen to almost characterize it:

$$s = \{\langle H, \mathit{Claire}, 3\clubsuit\rangle, \langle S, \mathit{Claire}, s\rangle, \langle S, \mathrm{Max}, s\rangle\}$$

Clearly our statement is true in this model. It is also easy to see that s is a hereditary subsituation of any situation which is a model of our statement, so by Proposition 37.1, s almost characterizes the statement. This definition is justified by the following result, which is an easy consequence of Proposition 37.1.

Proposition 37.4 *Suppose that the situation s almost characterizes the* ∃*-statement* Φ. *Then for any fact* σ, *the following are equivalent:*

1. σ *is entailed by* Φ, *i.e.*, σ *holds in all models of* Φ
2. $s \models \sigma$

The following is the main result of this paper. It shows the extent to which the shared environment approach can be approximated by the fixed point and iterate approaches.

Theorem 37.5 *Every* ∃*-statement* Φ *is almost characterized by some finite situation* s_Φ.

Proof First one establishes a normal form lemma for ∃-statements, where all the quantifiers are pulled out front. One then uses the Solution Lemma of AFA to define the desired situation. The proof that it almost characterizes the statement uses Proposition 37.1. □

However, there is a distinct sense in which ∃-statements are more discriminating than the situations that almost characterize them. For example, compare our above example of an ∃-statement with the following:

$$\exists e_1, e_2\,[e_1 \models ((\text{Claire } H3\clubsuit) \wedge (\text{Claire } S\ \ e_2)) \wedge e_2 \models ((\text{Claire } H3\clubsuit)\, J \wedge (\text{Max } S\ \ e_1))]$$

Clearly any model of our first statement is a model of our second. However, it is easy to see that there are models of our second that are not models of our first. (Think of a case where the card is not an up card, but is down, but where there are suitably placed mirrors.) On the other hand, these two statements are almost characterized by exactly the same situations. Or, in view of Proposition 37.4, the two statements entail the same facts, both wellfounded and circular.

Intuitively, what is going on here is that both of these statements represent ways in which Max and Claire might share the information that Claire has the 3♣. The first would be the one predicted by a literal reading of the Clark and Marshall account, but the second is clear in the spirit of that account. However, this means that since they are not equivalent, neither one can be the right characterization of the shared information. Rather, what they represent are two distinct ways, among many, that Max and Claire might have come to have the shared information. We leave it to the reader to work out analogous inequivalent ∃-statements that also give rise to the shared information in our running example.

We conclude this section by observing that the results can be extended to the case where we allow disjunctions to occur in ∃-statements, if one also allows disjunctive facts.

Conclusions

In thinking about shared information and common knowledge, it is important to keep three questions separate: (i) What is the correct analysis of common knowledge? (ii) Where does it come from? (iii) How is it used?

It would be neat if these three questions got their answers from the three different approaches in the literature. The results discussed above prompt us to propose that the fixed-point approach is the right analysis of the notion, and that it typically arises through some sort of shared environment.

However, by definition, the epistemically neutral case we have been studying is divorced from questions of use. To think about how shared information gets used, we turn to the epistemic case. Let us suppose that the fixed-point approach, or something like it, characterizes common knowledge, and the shared-environment approach characterizes the way in which common knowledge commonly arises. Does it follow that the iterate approach approximates common knowledge, or perhaps how it is used?

It seems that it can't. A clear difference between having information and knowing arises in the respective relationships between the fixed-point facts and its approximations. In the nonepistemic case, it is a matter of logical entailment. However, in the latter case, the fixed-point fact will simply not entail the analogous approximations. To see why, let's consider an example.

Example 37.2 Consider the following situation s, where we use K for the relation of knowing of a situation:

$$\langle H, \text{Max}, 3\diamondsuit\rangle\ ,\ \langle K, \text{Claire}, s\rangle\ ,\ \langle K, \text{Dana}, s\rangle\ ,\ \langle K, \text{Max}, s\rangle$$

It seems clear that the fact

$$\theta = (\text{Max}\ H3\diamondsuit) \wedge (\text{Claire}\ K\theta) \wedge (\text{Dana}\ K\theta) \wedge (\text{Max}\ K\theta)$$

holds in this situation. However, is it a fact in this situation that, say, Max knows that Dana knows that Claire knows that he, Max, has the $3\diamond$? And even more iterations?

It seems clear that it will not in general be true. After all, some sort of inference is required to get each iteration, and the players might not make the inference. They are, after all, only 3 years old. And even if Claire makes her inference, Dana may have legitimate doubts about whether Claire has made her inference. But once one player has the least doubt about some other player's making the relevant inference, the iterated knowledge facts breaks down. That is, once the making of an inference is implausible, or even just in doubt, the next fact in the hierarchy is not really a *fact* at all.

It is usually said that the iterate account assumes that all the agents are perfectly rational, that is, that they are perfect reasoners. This example also shows that it in fact assumes more: it assumes that it is *common knowledge* among the agents that they are all perfectly rational. It is only by making this radical idealization, plus restricting attention to finite situations, that the iterate account is equivalent to the fixed-point account. And the idealization requires the very notion that one is trying to understand in the first place.

We began this section by asking three questions. We have proposed answers to the last two of them, and suggested that the third question, about how common knowledge is used, is not answered by the iterate approach. But then how *do* people make use of common knowledge in ordinary situations?

My own guess is that common knowledge per se, the notion captured by the fixed-point analysis, is not actually all that useful. It is a necessary but not a sufficient condition for action. What suffices in order for common knowledge to be useful is that it arise in some fairly straightforward shared situation. The reason this is useful is that such shared situations provide a basis for perceivable situated action, action that then produces further shared situations. That is, what makes a shared environment work is not just that it gives rise to common knowledge, but also that it provides a stage for maintaining common knowledge through the maintainence of a shared environment. This seems to me to be part of the moral of the exciting work of Parikh, applying ideas of game theory to the study of communication.[16]

It seems to me that the consequences of this view of common knowledge are startling, if applied to real world examples, things like deterrence (mutual assured destruction, say). Indeed, it suggests a strategy of openness that is the antithesis of the one actually employed. But that goes well beyond the scope of this conference.

Finally, let me note that the results here do not lend themselves to an immediate comparison with other mathematical models of common knowledge, especially the approaches in game theory. It would be interesting to see a similar analysis there, one that pinpoints the finiteness or compactness assumption that must be lurking behind the Tan and Ribeiro da Cost Werlang result.

[16]Prashant Parikh, "Language and strategic inference," Ph.D. Dissertation, Stanford University, 1987.