Common Knowledge

Milena Scaccia

December 15, 2008

Abstract

In a game, an item of information is considered common knowledge if all players know it, and all players know that they know it, and all players know that they know that they know it, *ad infinitum*. We investigate how the information structure of a game can affect equilibrium and explore how common knowledge can be approximated by Monderer and Samet's notion of common belief in the case it cannot be attained.

1 Introduction

A proposition A is said to be *common knowledge* among a group of players if all players know A, and all players know that all players know A and all players know that all players know that all players know A, and so on *ad infinitum*. The definition itself raises questions as to whether common knowledge is actually attainable in real-world situations. Is it possible to apply this infinite recursion?

There are situations in which common knowledge seems to be readily attainable. This happens in the case where all players share the same state space under the assumption that all players behave *rationally*. We will see an example of this in section 3 where we present the Centipede game.

A public announcement to a group of players also makes an event A common knowledge among them [12]. In this case, the players ordinarily perceive the announcement of A, so A is implied simultaneously is each others' presence. Thus A becomes common knowledge among the group of players.

We observe this in the celebrated Muddy Foreheads game, a variant of an example originally presented by Littlewood (1953). In this game of n players, each player is to determine whether he has mud on his forehead whilst being able to observe the state of the $n-1$ other players, and not his own. If a sage announces to the n players that at least one of them has a muddy forehead, then this information becomes common knowledge. Everyone knows that at least one player has a muddy forehead, and everyone knows that everyone knows that at least one player has a muddy forehead, and so on. If, at each period, the sage questions the players as to whether they have mud on their foreheads, and under the assumption the players are rational and truthful, it can be proven by induction that if k players have muddy foreheads, then by the k^{th} period, all k will confess to having mud on their foreheads. In section 3, we revisit this game in greater detail and observe how the presence and absence of common knowledge can drastically affect equilibrium in this game. We investigate how crucial the assumption of common knowledge is and how certain theories collapse in the event common knowledge is absent [7].

The above view of common knowledge may be applicable only in limited contexts. In the case that knowledge is derived from sources that are not completely reliable (as in sending and receiving information through a medium which is not perfectly reliable), then common knowledge cannot be achieved [8].

Consider the case of processors in a distributed system that must jointly execute a complex computer protocol [3]. The processors are to coordinate themselves by means of communicating to each other what they know. It has been established that a necessary condition for the processors to be able to perfectly synchronize themselves is that of common knowledge. But given that messages sent and received between them have the possibility of failing, then common knowledge seems to be an excessively strong requirement that may rarely be achieved in practice.

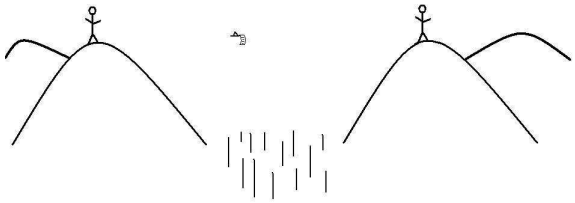

Figure 1: The Coordinated Attack Problem

Consider a version of Gray's original coordinated attack problem which can serve as an analogy to the communicating processors. The game is as follows: Suppose there are two allied armies situated on opposite hilltops wishing to attack their foe, residing in the valley (Figure 1). The state of the enemy *Prepared*, *Unprepared*, is only observable to the first commander. When he sees that the enemy is unprepared, he wants to alert the second commander. The first commander will not attack unless he is absolutely sure the second commander will do the same. If they attack at the same time while the enemy is unprepared, they will win the battle, otherwise they will be defeated. Their only means of communication is through a messenger pigeon, which with positive probability may not make it to the other side at some point during its rather dangerous trek. In the case that the message sent from the first commander to the second does arrive successfully, both commanders now know the message, but the first commander cannot be sure that the second knows it. Hence, the second will need to send a confirmation to the first. Assuming it arrives, we now have that commander 1 knows that commander 2 knows, but we do not have that commander 2 knows that commander 1 knows that he knows. This will require commander 1 to send another confirmation. Wouldn't this be enough for the two to know they will both attack? No, because again, we have that commander 1 is not absolutely certain that his last message was received, and so on and so forth. We must convince ourselves that under this protocol, in which communication is not 100% reliable, no matter how many trips the pigeon makes, there is no iron-clad guarantee that the opposing side will attack at the same time.

Given that the attainment of common knowledge is not possible in such contexts, we are left with the open problem of determining how common knowledge may be approximated. In section 4, we go on to apply Monderer and Samet's proposed way of achieving a reasonable level of coordination by approximating common knowledge with the weaker concept of *common belief*.

In the last section, we comment on the protocol used in the coordinated attack game as well as provide a brief discussion on the possible limitations of the proposed solution to approximating common knowledge.

In the next section, we formalize the definition of common knowledge and give a brief background on concepts needed for the rest of this paper.

2 Background

2.1 Preliminaries

We present the basic definitions of Mutual Knowledge and Common Knowledge.

Definition 2.1 *An event is **Mutual Knowledge** if players know this event. Note this is weaker than common knowledge.*

Let $\mathcal{P}$ be a finite set of players, and let the space (Ω, Σ, μ) be a probability space, where Ω is the state space, Σ is the σ-field of events, and μ is the probability measure on Σ. Let $H_i,\ i \in P$ be a measurable partition of Σ and $H_i(\omega)$ be the element of H_i containing ω, $\omega \in \Omega$. The set $H_i(\omega)$ contains the set of states that is indistinguishable to player i when ω occurs. Last but not least, let $\mathcal{F}_i$ be the σ-field generated by H_i. We will need this when proving a theorem on common knowledge.

Let $K_i(E)$ denote the event *player i knows event E*. This set will consist of all states ω in which player i knows E:

$$K_i(E) = \{\omega : H_i(\omega) \subseteq E\}.$$

We refer to K as the "knowledge operator".

Hence the event *everyone knows E* (Mutual Knowledge) is defined by

$$\mathcal{K}(E) = \cap_{i \in \mathcal{P}} K_i(E) = \{\omega : \cup_{i \in \mathcal{P}} H_i(\omega) \subseteq E\}.$$

We refer to $\mathcal{K}$ as the "everyone knows" operator and can construct a recursive chain as follows:

The event *everyone knows that everyone knows E* (Second order mutual knowledge) is defined by

$$\mathcal{K}^2(E) = \cap_{i \in \mathcal{P}} K_i(\mathcal{K}(E)) = \{\omega : \cup_{i \in \mathcal{P}} H_i(\omega) \subseteq \mathcal{K}(E)\}.$$

Everyone knows that everyone knows that everyone knows E (third order mutual knowledge) is defined by

$$\mathcal{K}^3(E) = \cap_{i \in \mathcal{P}} K_i(\mathcal{K}^2(E)) = \{\omega : \cup_{i \in \mathcal{P}} H_i(\omega) \subseteq \mathcal{K}^2(E)\}.$$

Note that this is a decreasing sequence of events, for $K^{n+1}(E) \subseteq K^n(E)$ and that $K^0(E) = E$.

The event *everyone knows that everyone knows that everyone knows (ad infinitum) E* can be defined as the intersection of all sets of the form $\mathcal{K}^n(E)$:

$$\mathcal{K}^\infty(E) = \cap_{n \geq 1} \mathcal{K}^n(E),$$

where $\mathcal{K}^n(E) = \cap_{i \in \mathcal{P}} K_i(\mathcal{K}^{n-1}(E))$. Because $\mathcal{K}^\infty$ is a countable intersection of classes, it must itself be in this class.

Proposition 2.2 *An event E is **common knowledge** at state* $\omega \iff \omega \in \mathcal{K}^\infty(E)$.

Some properties of the "knowledge operator", K.

Property 1 $K_i \in \mathcal{F}_i$.

Property 2 *If* $E \in \mathcal{F}_i$, *then* $K_i(E) = E$.

Property 3 $K_i(K_i(E)) = K_i(E)$.

Property 4 *If* $F \subseteq E$ *then* $K_i(F) \subseteq K_i(E)$.

Property 5 *If* (A_n) *is a decreasing sequence of events, then* $K_i(\bigcap_n A^n) = \bigcap_n K_i(A^n)$.

Proof (Proposition 2.2).

$(\Longrightarrow)$ Suppose that event E is common knowledge at ω.

There exists an $\omega \in F \subseteq E$ such that $F \in \mathcal{F}_i \ \forall i \in \mathcal{P}$. By properties 2 and 4, $F = K_i(F) \subseteq K_i(E)$. Hence $F \subseteq \cap_i K_i(E)$. Recall we have defined this intersection to be one iteration of the "everyone knows" operator, $\mathcal{K}$. So $F \subseteq \mathcal{K}(E)$. Since $\mathcal{K}^{n+1}(E) \subseteq \mathcal{K}^n(E)$, then by induction on $n,\ n \geq 1,\ F \subseteq \mathcal{K}^n(E)$. But this implies that $E' \subseteq \cap_{n\geq 1}\mathcal{K}^n = \mathcal{K}^\infty$. Therefore, $\omega \in F \subseteq \mathcal{K}^\infty$.

$(\longleftarrow)$ Conversely, suppose that $\omega \in \mathcal{K}^\infty$. It will suffice to show that $\mathcal{K}^\infty(E) \subseteq E$ and that $\mathcal{K}^\infty(E) \in \mathcal{F}_i \ \forall i \in \mathcal{P}$. Because $\mathcal{K}$ satisfies the decreasing sequence of events property, then for $n \geq 1$, $\mathcal{K}^\infty(E) \subseteq \mathcal{K}^n(E) \subseteq \mathcal{K}(E) \subseteq K_i(E) \subseteq E$.

With the help of property 5, we also have that $\mathcal{K}^\infty(E) \subseteq \mathcal{K}^{n+1}(E) \subseteq \bigcap_{n\geq 1} K_i(\mathcal{K}^n(E)) = K_i(\bigcap_{n\geq 1}\mathcal{K}^n(E)) \subseteq K_i(\mathcal{K}^\infty(E)) \subseteq \mathcal{K}^\infty(E)$. This shows that $\mathcal{K}^\infty(E) \in \mathcal{F}_i$.

∎

2.2 Computing Common Knowledge

To better grasp the notion of common knowledge, we use the Muddy Foreheads game to illustrate a simple example on how to determine what is common knowledge among a set of players.

We consider a simple case with only three players. Let us compute player 1's information partition. Denote the case where a player has a muddy forehead by 1 and that in which he does not by 0. Given he can see the other two players and not his own forehead, he would not be able to distinguish between the state in which he has a muddy forehead, and that in which he does not, regardless of the other players' foreheads. In the case that there is exactly one muddy forehead and it belongs to player 1, then he will know he has a muddy forehead. In the case that there are no muddy foreheads, the sage will announce so, and player 1 will know he does not have mud on his forehead. Hence, player 1's information partition H_1 will consist of the following sets, where each individual set contains a set of states that is indistinguishable to player 1.

$$H_1 = [\{000\}, \{100\}, \{001, 101\}, \{010, 110\}, \{011, 111\}].$$

We can construct analogous information partitions for players 2 and 3.
Fudenberg and Tirole show that if all three foreheads are clean and all players are informed of this in public by a sage, then event E= "There are no muddy faces"= 000 is common knowledge when it occurs. So $\mathcal{K}(E) = \{000\}$, $\mathcal{K}(\mathcal{K}(E)) = \mathcal{K}(E) = E = \{000\}$, and so on. Iterating the "everyone knows operator"

will always give us back state 000: $\mathcal{K}^{\infty} = 000$. Everyone will be able to distinguish that they have a clean forehead and will know that every other player knows that every other player,..., knows they have clean foreheads. The event 000 is common knowledge when it occurs.

Similarly, the event E= "At least one player has a muddy forehead" = $\{100, 001, 101, 010, 110, 011, 111\}$ can be shown to be common knowledge when it occurs. One iteration of the "everyone knows operator" gives: $\mathcal{K}(E) = \{100, 001, 101, 010, 110, 011, 111\} = E$. Assuming common knowledge of rationality, then everyone knows that everyone knows there is at least one muddy forehead, and so on, i.e. $\mathcal{K}^2(E) = E$, $\mathcal{K}^3(E) = E$,... and so on. This illustrates how $\mathcal{K}$ acts as a fixed point operator in the case of common knowledge.

Let us attempt to compute common knowledge in an environment in which it does not exist. Suppose there is no sage who makes any sort of public announcement. In this case, player 1's information partition H_1 will consist of the following sets (similarly for players 2 and 3),

$$H_1 = [\{000, 100\}, \{001, 101\}, \{010, 110\}, \{011, 111\}].$$

Let us begin applying the "everyone knows" operator on the event E ="there exists at least one muddy forehead". In the case that there is exactly one muddy forehead, not everyone will know the event E for the player whose forehead is muddy does not know that there is at least one muddy forehead. Thus we cannot include states 000, 100, 001, 010 in the space $\mathcal{K}(E)$. In the case that there are at least two muddy foreheads, then everyone will know E. Hence $\mathcal{K}(E)$ will consist solely of the following states:

$$\mathcal{K}(E) = \{111, 110, 101, 011\}.$$

Let use see what happens when we iterate again. Now we want to know in what states everyone knows that everyone knows E. Let us consider the case in which there are exactly two muddy foreheads, that is, the states 110, 101, and 011. In state 110, player 1 sees only one muddy forehead, hence knows event E. Since he does not know his own status, he cannot determine what player 2 sees. He thinks: if he has mud on his forehead, then player 2 will see one muddy forehead. If he does not, then player 2 would see zero muddy foreheads. He is unsure. That is, he is not able to distinguish between these two states ($\{110,010\}$), hence he does not know that player 2 knows E. We can argue similarly for the remaining states in which there are exactly two muddy foreheads. Hence the only state in which all players knows that all players know E is the state in which all three players have muddy foreheads:

$$\mathcal{K}(\mathcal{K}(E)) = \mathcal{K}(111, 110, 101, 011) = 111.$$

Iterating again, $\mathcal{K}(111)$ results in a null set, for no player can distinguish 111 from the state in which he does not have mud on his forehead and that in which he does. There does not exist an ω where E is common knowledge [3].

2.3 Common Beliefs

We introduce the concept of common belief which will be used in section 4 of the paper. Common belief can be defined similarly to common knowledge only that the phrase *players knows E at* ω is replaced by

players p-believe E at ω. Thus, an event is common p-belief if everyone believes it with probability at least p, and everyone believes with probability at least p that everyone believes it with probability at least p, and so on *ad infinitum*.

Definition 2.3 *A player i* ***p-believes*** *E at* ω *if he believes that E occurs with probability at least p at* ω*. More formally,* $\mu_i(E|H_i(\omega)) \geq p$*, where* μ_i *is player's i probability distribution over* Ω*.*

We replace our definition from section 2.1 of "player i knows E", by "player i p-believes E" as follows:

$$B_i^p(E) = \{\omega : \mu(E|H_i(\omega) \geq p\}.$$

We can define n^{th}-order mutual p-belief analogously to n^{th}-order mutual knowledge as we have done in section 2.1.

Hence the event *everyone p-believes E* (Mutual Knowledge) is defined by

$$\mathcal{B}^p(E) = \cap_{i\in\mathcal{P}} B_i^p(E)$$

Everyone knows that everyone knows that everyone knows...(k times) E (k^{th}-order mutual knowledge) is defined by

$$\mathcal{B}_k^p(E) = \cap_{i\in\mathcal{P}} B_i^p(\mathcal{B}_{k-1}^p(E)))$$

Hence the event *everyone knows that everyone knows that everyone knows (ad infinitum) E* can be defined as the intersection of all sets of the form $\mathcal{B}_n^p(E)$:

$$\mathcal{B}_\infty^p(E) = \cap_{n\geq 1} \mathcal{B}_n^p(E),$$

where $\mathcal{B}_n^p(E) = \cap_{i\in\mathcal{P}} B_i(\mathcal{K}^{n-1}(E))$.

Proposition 2.4 *An event E is* ***common p-belief*** *at state* ω *if* $\omega \in \mathcal{B}_\infty^p(E)$.

The proof is similar to that of proposition 2.2.

Common p-belief is a weakening of common knowledge. It has been shown by Monderer and Samet that if *perfect* coordination is achievable in a game where there is common knowledge of the structure of the game, then *approximate* coordination is achievable when there is common p-belief, for some p sufficiently close to 1 [7,8].

From this result, Kajii and Morris show that if $\Sigma_{i\in\mathcal{P}} p_i < 1$, then $P(\mathcal{B}_\infty^p(E))$ is close to 1 whenever $P(E)$ is close to 1. This is implied in the following proposition which they call the *Critical Path Result*:

Proposition 2.5 *If* $\Sigma_i p_i < 1$*, then the probability that event E is common p-belief is at least*

$$1 - (1 - P(E))\frac{1 - min(p_i)}{1 - \Sigma_i p_i}.$$

We will not prove, but we will use Monderer and Morris' results in section 4 of the paper to show that by approximating with common belief, there exists an equilibrium where commanders are able to coordinate an attack under the condition that the probability the enemy is unprepared is close to 1.

In the next section, we examine how the presence and absence of common knowledge can affect the Nash equilibrium in the Muddy Foreheads and the Centipede games.

3 The Effect of Common Knowledge on Equilibrium

We see how the presence and lack of common knowledge affects equilibrium in the Muddy Foreheads game. We also briefly comment on the relationship between common knowledge and backward induction using the Centipede game.

3.1 The Muddy Foreheads Game

We elaborate on the previously mentioned muddy foreheads game of three players by introducing payoffs associated to players' actions as follows. We make use of the same payoffs Fudenberg and Tirole use in [4]. Each player receives:

1. a payoff of δ^t if he confesses in period t while having a muddy forehead. $0 < \delta < 1$, so that the longer it takes a player to confess, the smaller his payoff will be. This will give players an incentive to confess immediately once they know their forehead is muddy.
2. a payoff of 1 if he does not confess while having a clean forehead.
3. a large negative payoff of -100 if he confesses while having a clean forehead.
4. a small negative payoff of -1 if he does not confess while having a muddy forehead.

The payoffs can be summarized in the following table.

	Muddy	Clean
Confess	δ^t	-100
¬Confess	-1	1

Table 1: Payoff Matrix - Muddy Forehead Game

Thus, it is in each player's own interest to not confess to a muddy forehead unless they are absolutely certain that this is the case.

The game begins with a sage publicly announcing that at least one player has a muddy forehead, if and only if this is the case. This makes a fact that may have been previously knows to all the players common knowledge. Following the announcement, players will be able to determine their own type by the actions of the other players [13].

1. Period 1: If there is exactly one muddy forehead, the player with the mud observes no muddy foreheads, and thus infers he has the mud. He confesses immediately due to the discount factor δ.
2. Period 2: If in period 1, no one confessed, then it is common knowledge that all players know that there are at least two muddy foreheads. Otherwise, we would have expected someone to confess in the first period.
   If there are exactly two muddy foreheads then the two players with mud observe only one clean

forehead. Since they know that there is more that one muddy forehead from period 1, then they infer that they must be the second player with mud. Hence they confess in the second round.

3. Period 3: If no player confessed in the second round, this means that each player saw two muddy foreheads. Each player will make the deduction that all of them have mud on their foreheads, thus they should all confess in the third period.

More generally, in an *n*-player game in which there are *k* muddy foreheads, all players with mud will confess by the k^{th} period. Thus, *equilibrium* arises after *k* periods of iterated rationality.

3.2 Removing the Common Knowledge Assumption in the Muddy Foreheads Game

Let us tweak the game to observe what would happen in the case in which there is no public announcement at the beginning of the game, i.e. removing the assumption of common knowledge (while still keeping the assumption of rationality).

We consider the 3-player game. Each player can see the foreheads of their opponents, but in this case, whether they see no muddy foreheads, one muddy, or two muddy, it will not be enough information for players to infer their state at any period in the game. In this case, player 1's information partition would be $H_1 = [\{000, 100\}, \{001, 101\}, \{010, 110\}, \{011, 111\}]$. Take for example, the case in which there is exactly one muddy forehead. The players with clean foreheads each observe one muddy forehead and thus know there is at least one muddy forehead while the player with the mud sees no muddy foreheads, and thus does not know that there is at least one muddy forehead. For all he knows, there can be no muddy foreheads. If we iterate the everyone knows operator, we saw in section 2.2 that it will collapse at some point, because we will not find an ω in which the event is common knowledge.

Because he believes muddy and clean is equally likely, he will not confess. If he confesses while having a clean forehead, he suffers a large negative payoff. So it is in his own interest to stay quiet, and receive, at worst, a negative payoff of -1. Hence the players learn nothing from their opponents' play and will continue to believe, at each round, that muddy and clean are equally likely. Thus no player will have the incentive to deviate from their choice of not confessing as it would render a large negative expected payoff.

So the Nash equilibrium in this case is for all players to never confess, even when all their foreheads are muddy. As in the coordinated attack game, when there is a lack of common knowledge between players, they choose to play it safe.

Drawing a comparison, in an *n*-player game in which *all* players have muddy foreheads,

1. if common knowledge is satisfied, then all players receive a payoff of $\delta^n \geq 0$ by the *n*-th round.
2. if common knowledge is not satisfied, the Nash is to not confess, and thus all players receive payoff -1.

We conclude that the players were better off when common knowledge was satisfied.

3.3 Backward Induction and Common Knowledge

We comment on how common knowledge of rationality among players has strong implications for backward induction.

Definition 3.1 ***Backward induction*** *is an iterative process for solving finite sequential games. The last player, who must choose between leaves of the game tree, makes a choice that maximizes his payoff. Then, the second-to-last player makes a choice maximizing his payoff. The process continues until we reach the beginning of the game at which point all players' actions have been determined. Effectively, one determines the Nash equilibrium of each subgame of the original game.*

This can be manifested in the Centipede Game (Rosenthal 1981) in which two players must take turns choosing either to move down in which they take a slightly larger share of the increasing pot, or to move right in which case they pass the pot to the other player.

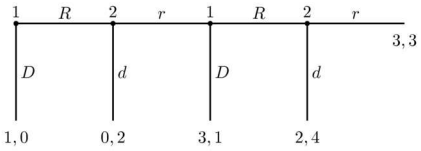

Figure 2: The Centipede Game

If we were to reach the last round of the game, and assuming that player 2 is rational, he would choose to move down and receive a payoff of 4 rather than to move right and receive a payoff of 3. However, given that player 2 will choose to move down, then player 1 should choose to move down in the second to last round, receiving 3 instead of 2. But knowing this, player 2 ought to move down in the third to last round, taking a slightly higher payoff than he would have received by allowing the first player to defect in the second to last round. This reasoning proceeds until we have reached the first node of the tree, concluding that the best action for the first player is to move down on the first round. Thus the Nash Equilibrium is to get (1,0).

In his paper "Backward Induction and Common Knowledge of Rationality", Aumann's proves that *if common knowledge of rationality obtains in a game of perfect information, then the backward induction outcome is reached* [2]. In order to have reasoned backwards to the first node, we needed that the players were rational. Aumann states that simple rationality on the part of each player is not enough - the players must ascribe rationality to each other. (Otherwise player 1 may believe that the next player may choose to go right instead of down, giving him the opportunity of receiving a greater payoff on his next turn). So if it is common knowledge that all players are rational, then they know the next player will choose to move down in order to maximize his payoff, meaning player 1's best response is to choose to go down on the first round.

In the next section, we return to the coordinated attack game in which it seems impossible to attain common knowledge through unreliable communication.

4 The Coordinated Attack Game

4.1 Truncating the Knowledge Hierarchy

When we introduced this game, we left off by stating that the pigeon is to make an infinite number of trips between the two commanders in order to achieve common knowledge. Let us convince you of this. Even though it seems sensible to believe that for m large, m messages would sufficiently approximate common knowledge of the enemy being unprepared, (i.e. that $\mathcal{K}^m \approx \mathcal{K}^\infty$), we will see through Rubinstein's theory, that it is not possible to approximate common knowledge with m^{th}-level mutual knowledge. Before attempting to prove this, we give more structure to the game by presenting the payoff matrices in the cases where the enemy is prepared and that in which he is not. It is further assumed that all players have common prior probabilities: P(*message failing to be delivered*) = $\varepsilon > 0$, and P(*enemy prepared*) = δ where $\varepsilon < \delta$.

		Commander 2	
		Attack	¬Attack
Commander 1	Attack	(1,1)	(-M,0)
	¬Attack	(0,-M)	(0,0)

Table 2: Payoff Matrix when Enemy is Unprepared

		Commander 2	
		Attack	¬Attack
Commander 1	Attack	(-M,-M)	(-M,0)
	¬Attack	(0,-M)	(0,0)

Table 3: Payoff Matrix when Enemy is Prepared

The play-it-safe strategy for both armies is to not attack. They may run the risk of not obtaining positive payoff this way, but they can avoid attacking alone and receiving large negative payoff.

Proposition 4.1 *Truncating common knowledge hierarchy at any finite level can lead agents to behave as though they had no mutual knowledge at all.*

Rubinstein has proved the above using the "Electronic Mail Game" which is also a game involving coordination, somewhat similar to the coordinated attack problem. We shall present the proof by induction applied to the coordinated attack problem.

Proof. Suppose that the enemy is initially unprepared. Let N_1 be the number of messages that commander 1 sends and N_2 be the number of messages commander 2 sends.

BASE CASE:

If N_2=0, then commander 2 did not send a reply for he did not receive any messages from commander 1. This may be due to the following two possibilities:

1. $N_1 = 0$ AND $N_2 = 0$: The enemy is prepared, so commander 1 did not send a message. This occurs with probability δ.
2. $N_1 = 1$ AND $N_2 = 0$: The enemy is unprepared, but commander 1's message failed to deliver. This occurs with probability $(1-\delta) * \varepsilon$.

In either case, commander 2 believes that with probability $\frac{\delta}{\delta+(1-\delta)*\varepsilon} > 1/2$ that the enemy is prepared. We see that expected utility of commander 2 in the case he chooses not to attack is greater than that of if he chooses to attack:

$$E(U_2(Attack)|N_2 = 0) \leq -M(1/2) + 1(1/2) < 0.$$
$$E(U_2(\neg Attack)|N_2 = 0) = 0.$$

Commander 2 is better off playing it safe and not attacking, no matter what commander 1 choose to do.

Let's see what happens to commander 1's expected payoff: If the enemy is prepared, then no message is sent and clearly his best response is to not attack. If the enemy is unprepared, and he receives no confirmation to his first message, then he believes that commander 2 did not receive his message with a probability of:

$$\frac{(1-\delta)*\varepsilon}{(1-\delta)*\varepsilon+(1-\delta)*epsilon*(1-\varepsilon)} = \frac{\varepsilon}{\varepsilon+(1-\varepsilon)\varepsilon} > 1/2.$$

Hence, commander 1 believes that with probability greater than $1/2$, his message never arrived and thus with probability greater than $1/2$, commander 2 will not attack. Once again, the expected payoffs for commander 1 are the same as those above. His best response is to not attack.

I.H.:
For all $N_i < n$, each commander's best response is to not attack so that the unique Nash of the game is $(\neg Attack, \neg Attack)$.

STEP CASE:
Assume $N_1 = n$. That is, commander 1 sent n messages and no more, for he did not receive a response after his last message. $(K_1K_2)^{n-1}(enemy\neg prepared)$ is true, but it is not true that $(K_1K_2)^n(enemy\neg prepared)$. He is uncertain whether the last message was received successfully. Thus, he does not know which of the two following events actually occurred.

A : his n^{th} message was lost, i.e. $N_2 = n-1$.

B : his n^{th} message was not lost, but commander 2's n^{th} response was lost.

The probability that the last message from commander 1 to commander 2 was lost can be expressed as,

$$\mu(A|N_1 = n) = \frac{\mu(N_1 = n|A)\mu(A)}{\mu(N_1 = n|A)\mu(A) + \mu(N_1 = n|B)\mu(B)}$$
$$= \frac{1*\mu(A)}{1*\mu(A) + 1*\mu(B)}$$

$$= \frac{\varepsilon}{\varepsilon + (1-\varepsilon)\varepsilon}$$
$$> 1/2.$$

Hence, $\mu(A|N_1 = n) > 1 - \mu(A|N_1 = n) = \mu(B|N_1 = n)$, i.e. event A is more likely than event B. So in the case that $N_1 = n$ and $N_2 = n-1$, by the induction hypothesis, commander 1 assesses that commander 2 will choose not to attack. The expected utility of commander 1 choosing to attack while not receiving an n^{th} confirmation from commander 2 is thus:

$$E(U_1(Attack)|N_1 = n) \leq -M(1/2) + 1(1/2) < 0.$$

And the expected utility of commander 1 choosing to not attack while not receiving an n^{th} confirmation from commander 2 is:

$$E(U_1(\neg Attack)|N_1 = n) = 0.$$

We compute commander 2's expected utilities analogously.

Hence $(\neg Attack, \neg Attack)$ is the unique Nash of the coordinated attack game for all n finite.

∎

Thus no number of messages will suffice in order to achieve common knowledge of the desire to attack under the current protocol. In the next subsection, we restore this unbounded hierarchy by replacing common knowledge with common belief. We explore how under particular circumstances, commanders may have the incentive to attack.

4.2 Approximating Common Knowledge with Common Beliefs

We have seen that under their communication protocol, the commanders will never achieve common knowledge, hence never achieve a coordinated attack. But Morris and Rubinstein state that the fact that they could not achieve common knowledge does not exclude the possibility that with positive probability they will both attack.

The general idea here is that under certain circumstances, the commanders may have incentive to attack if they assign a common high probability to the enemy being unprepared (δ small), to the communication being reliable ($\varepsilon < \delta$), and to the other side attacking. In other words, we may weaken the notion of common knowledge to that of common belief to achieve approximate coordination.

In the example of the coordinated attack problem seen thus far, the commanders did not commit to strategies before the communication stage. Secondly, the commanders were allowed to have different objectives, i.e. a commander would prefer the other side to attack alone than for his side to attack alone - he receives greater payoff this way. Perhaps, if they were to commit to an action protocol before communication took place, coordinated attack may take place with high probability. But what if we make matters simple and tweak the payoff matrices so that the two commanders have the same objective? In order to remove the conflict of interest, suppose the new payoff matrices were as follows:

		Commander 2	
		Attack	¬Attack
Commander 1	Attack	(1,1)	(-M,-M)
	¬Attack	(-M,-M)	(0,0)

Table 4: New Payoff Matrix when Enemy is Unprepared

		Commander 2	
		Attack	¬Attack
Commander 1	Attack	(-M,-M)	(-M,-M)
	¬Attack	(-M,-M)	(0,0)

Table 5: New Payoff Matrix when Enemy is Prepared

We note that not attacking is no longer a "play-it-safe" strategy. For if one side does *not* attack while the other side does, they can do just as bad! Morris and Shin describe the "both attack"-equilibrium as risk-dominant in the sense that there exists a probability $p < 1/2$[1] such that if one commander assigns this probability to the other side attacking, his best response is to attack. They take this p to be $\frac{M}{2M+1}$. Given M large, this comes sufficiently close to $1/2$.

So the "both-attack" equilibrium occurs so long as it is common $\frac{M}{2M+1}$-belief that the enemy is unprepared. And given that the state in which the enemy is unprepared occurs with probability close to 1, then with the use of proposition 2.6 from section 2.3, we are able to show that common belief of the enemy being unprepared also occurs with probability close to 1. Consider proposition 2.6 in the case of two players, where $p_1 = p_2 = \frac{M}{2M+1}$. Then $\Sigma_i p_i = 2p$. So the condition $\Sigma_i p_i < 1$ reduces to $p < 1/2$. Now, given the probability the enemy is unprepared is $(1-\delta)$, we have that it is common $\frac{M}{2M+1}$-belief that the enemy is unprepared with probability

$$\leq 1-(1-(1-\delta))(\frac{1-p}{1-2p})$$

$$= 1-\delta(\frac{1-\frac{M}{2M+1}}{1-\frac{2M}{2M+1}})$$

$$= 1-\delta(M+1)$$

$$\approx 1.$$

So if, with sufficiently high probability, it is common $\frac{M}{2M+1}$-belief that the enemy is unprepared, then with sufficiently high probability, the equilibrium behavior implied by common belief approximates behavior implied by common knowledge.

[1]Recall from the proof of proposition 4.1, that if at any point commander 1 does not receive a confirmation from commander 2, then he assigns probability less than 1/2 to the second commander receiving his message. Thus it cannot be that it is common p-belief that the enemy is unprepared where $p \geq 1/2 \implies p < 1/2$.

5 Discussion and Conclusion

We have seen how the information structure of a game can affect the equilibrium of the game. In the Muddy Foreheads game, if it was not common knowledge that at least one player had mud on his forehead, then the equilibrium was for no player to confess, as opposed to players confessing truthfully by a certain period when common knowledge was satisfied. We have also seen in the Centipede game that if the backward induction outcome is not reached then the players did not possess common knowledge of rationality.

In these games, common knowledge of the environment seemed to be readily attainable among players for it followed from the fact that the players perceived an event simultaneously, shared the same state space, and from the assumption that all players were logically competent. In other more complex situations, such as the coordinated attack problem in which players communicate information to each other through an unreliable medium, we have seen that the attainment of common knowledge was not possible. In this case, we have explored a possible way of approximating the notion of common knowledge with that of common p-belief in order to achieve a coordinated attack with high probability.

The communication protocol used in the coordinated attack problem may have appeared foolish and unrealistic. It would seem feasible for the commanders to coordinate an attack with high probability if they would agree to abide by an action protocol prior to the communication phase. For example, their action protocol could be to attack if they receive at least one message. In this case, we do not require the notion of common knowledge or common p-belief in order to achieve a reasonable level of coordination. But this suggests other open questions in areas away from that of approximating common knowledge. That is, in the area of the design of optimal action protocols in systems that are subject to communication failures.

Despite seeming to have no practical interest, the coordinated attack problem may serve some purpose in determining how to deal with situations having similar unfortunate information structure. In his paper, "Comparing the Robustness of Trading Systems to Higher Order Uncertainty", Shin demonstrates that if the fact that trade is feasible is not common knowledge among traders in a decentralized market, then efficient trade is not attained [10]. The problem of trading in a decentralized market in which traders obtain noisy observations of the true state of payoffs seems to have similar implications as the coordinated attack problem.

On another note, we are led to ask the question: in the case that common knowledge cannot be attained, to what extent can the notion of common knowledge be relaxed in order to achieve coordination? It was possible to relax common knowledge in the coordinated attack game, a game that was binary and finite. There were only two players who chose from two possible actions: $Attack$ or $\neg Attack$, and there were two possible outcomes: *Win* or *Lose*. Given that the authors of papers surveyed thus far have only really considered binary and finite games, this may suggest that approximation using common p-belief may be feasible only in limited contexts. In [11], Shin seems to suggest that common p-belief will not suffice in approximating common knowledge in multi-player games in which players choose actions from a continuum, and in which there are many possible outcomes of coordination.

The question remains: does there exist a form of approximate common knowledge that would suffice in order to achieve coordination in such complex contexts? This is yet to be determined.

6 References

1 Aumann, R., 1976, “Agreeing to Disagree”, *Annals of Statistics 4*, 1236-9.

2 Aumann, R., 1995. “Backward Induction and Common Knowledge of Rationality”, *Games and Economic Behavior*, 8, 6-19.

3 Fagin, R., Halpern, J., Moses, Y., and Vardi, M., 1995, *Reasoning about knowledge*, Cambridge, MA: MIT Press.

4 Fudenberg, Drew, Tirole, Jean, 1991, *Game Theory, Chapter 14: Common Knowledge and Games*, Cambridge, Mass. MIT Press.

5 Halpern, Joseph Y., Moses, Yoram, 1990. “Knowledge and Common Knowledge in a Distributed Environment,” *Journal of the ACM* 37, 549-587.

6 Kajii, A., Morris, S., 1995. “The Robustness of Equilibria to Incomplete Information”, University of Pennsylvania, CARESS Working Paper #95-18, forthcoming in Econometrica.

7 Monderer, Dov and Samet, Dov. 1989, “Approximating Common Knowledge with Common Beliefs”, *Games and Economic Behavior 1*, 170-190.

8 Morris, Stephen, Song Shin, Hyun, 1997, “Approximate Common Knowledge and Co-ordination: Recent Lessons from Game Theory”, *Journal of Logic*, Language and Information 6, 171-190.

9 Rubinstein, Ariel, 1992. “The Electronic Mail Game: Strategic Behaviour Under ’Almost Common Knowledge”, *American Economic Review 79* (1989), 385-391.

10 Shin, H.S., 1996. “Comparing the Robustness of Trading Systems to Higher-Order Uncertainty”, *Review of Economic Studies*, 63, 39-60.

11 Shin, H.S.,Williamson, T., 1996. “How Much Common Belief is Necessary for a Convention?”, *Games and Economic Behaviour*, 13, 252-268.

12 Sillari, Giacomo, Vanderschraaf, Peter, 2001. “Common Knowledge”, *Stanford Encyclopedia of Philosophy*, $< http://plato.stanford.edu/entries/common-knowledge/index.html\#return-2.5 >$.

13 Weber, Roberto, A., 2001. “Behaviour and Learning in the ’Dirty Faces’ Game”, *Experimental Economics*, 4:229-242.