Knowledge and action in groups

Rohit Parikh
City University of New York

Bridges 2 Workshop
Rutgers, September 18, 2015

Part I

Savagery for many agents

Inducing Beliefs: Shakespeare's Much ado about Nothing

At Messina, a messenger brings news that Don Pedro, a Spanish prince from Aragon, and his officers, Claudio and Benedick, have returned from a successful battle. Leonato, the governor of Messina, welcomes the messenger and announces that Don Pedro and his men will stay for a month.

Inducing Beliefs: Shakespeare's Much ado about Nothing

At Messina, a messenger brings news that Don Pedro, a Spanish prince from Aragon, and his officers, Claudio and Benedick, have returned from a successful battle. Leonato, the governor of Messina, welcomes the messenger and announces that Don Pedro and his men will stay for a month.

Beatrice, Leonato's niece, asks the messenger about Benedick, and makes sarcastic remarks about his ineptitude as a soldier. Leonato explains that "There is a kind of merry war betwixt Signior Benedick and her."

Inducing Beliefs: Shakespeare's Much ado about Nothing

At Messina, a messenger brings news that Don Pedro, a Spanish prince from Aragon, and his officers, Claudio and Benedick, have returned from a successful battle. Leonato, the governor of Messina, welcomes the messenger and announces that Don Pedro and his men will stay for a month.

Beatrice, Leonato's niece, asks the messenger about Benedick, and makes sarcastic remarks about his ineptitude as a soldier. Leonato explains that "There is a kind of merry war betwixt Signior Benedick and her."

Various events take place and Claudio wins the hand in marriage of Hero, Leonato's only daughter and the wedding is to take place in a week.

Don Pedro and his men, bored at the prospect of waiting a week for the wedding, hatch a plan to matchmake between Beatrice and Benedick who inwardly love each other but outwardly display contempt for each other.

Don Pedro and his men, bored at the prospect of waiting a week for the wedding, hatch a plan to matchmake between Beatrice and Benedick who inwardly love each other but outwardly display contempt for each other.

According to this strategem, the men led by Don Pedro proclaim Beatrice's love for Benedick while knowing he is eavesdropping on their conversation. Thus we have, using b for Benedick, d for Don Pedro and E for the event of eavesdropping,

$$K_b(E),\ K_d(E) \text{ and } \neg K_b(K_d(E))$$

All these conditions are essential and of course the plot would be spoiled if we had $K_b(K_d(E))$ instead of $\neg K_b(K_d(E))$. Benedick would be suspicious and would not credit the conversation.
The women led by Hero carry on a similar charade for Beatrice. Beatrice and Benedick, are now convinced that their own love is returned, and hence decide to requite the love of the other.

The play ends with all four lovers getting married.

Benedick's Decision problem

	love	nolove
propose	100	-20
nopropose	-10	0

Here *love* means "Beatrice loves me" and *nolove* the other possibility.

Benedick and the two florists

Suppose there are two florists in Messina. If there is a wedding they will have to furnish flowers and that many flowers are only available in Napoli. So to get flowers for a wedding they have to write to Napoli and put down a deposit.

They both know that Beatrice loves Benedick.

Suppose $K_f(K_b(L))$ and $\sim K_{f'}(K_b(L))$.

Florist 1 knows that Benedick knows that Beatrice loves him.
Florist 2 does not know.

Suppose $K_f(K_b(L))$ and $\sim K_{f'}(K_b(L))$.

Florist 1 knows that Benedick knows that Beatrice loves him.
Florist 2 does not know.

So florist 1 will put down a deposit with a supplier in Napoli and the second florist will not.

They have the same utilities and the same knowledge about the world but their knowledge about Benedick's knowledge is different.

Suppose $K_f(K_b(L))$ and $\sim K_{f'}(K_b(L))$.

Florist 1 knows that Benedick knows that Beatrice loves him. Florist 2 does not know.

So florist 1 will put down a deposit with a supplier in Napoli and the second florist will not.

They have the same utilities and the same knowledge about the world but their knowledge about Benedick's knowledge is different.

And if we know that Beatrice loves Benedick and Benedick knows this, then we can infer the knowledge of the two florists about Benedick's knowledge from the fact that one put down a deposit and the other did not.

Card game

For example suppose Ravi and Usha are playing cards. There is a mirror behind Usha so Ravi can see her cards. But Usha does not know this.

Card game

For example suppose Ravi and Usha are playing cards. There is a mirror behind Usha so Ravi can see her cards. But Usha does not know this.

And there is a mirror behind Ravi, and Usha can see Ravi's cards. But Ravi only knows about Usha's mirror. And Usha only knows about Ravi's mirror.

Card game

For example suppose Ravi and Usha are playing cards. There is a mirror behind Usha so Ravi can see her cards. But Usha does not know this.

And there is a mirror behind Ravi, and Usha can see Ravi's cards. But Ravi only knows about Usha's mirror. And Usha only knows about Ravi's mirror.

Suppose Ravi has the Queen of spades. Let Q represent this fact. Then we have $K_r(Q)$, of course, $K_u(Q)$, and naturally $K_u K_r(Q)$.

Card game

For example suppose Ravi and Usha are playing cards. There is a mirror behind Usha so Ravi can see her cards. But Usha does not know this.

And there is a mirror behind Ravi, and Usha can see Ravi's cards. But Ravi only knows about Usha's mirror. And Usha only knows about Ravi's mirror.

Suppose Ravi has the Queen of spades. Let Q represent this fact. Then we have $K_r(Q)$, of course, $K_u(Q)$, and naturally $K_u K_r(Q)$.

But we do not have $K_r K_u(Q)$.

Reivising Kripke structures when an announcement is made

Suppose we are given a Kripke structure $\mathcal{M}$. Then some formula φ is announced publicly.

Reivising Kripke structures when an announcement is made

Suppose we are given a Kripke structure $\mathcal{M}$. Then some formula φ is announced publicly.

The new Kripke structure is then obtained by deleting all states in $\mathcal{M}$ where φ did not hold.

Theory of Mind

A group of children are told the following story:
Maxi goes out shopping with his mother and when they come back, Maxi helps mother put away the groceries, which include chocolate. There are two cupboards, red and blue. Maxi puts the chocolate in the red cupboard and goes out to play.

Theory of Mind

A group of children are told the following story:
Maxi goes out shopping with his mother and when they come back, Maxi helps mother put away the groceries, which include chocolate. There are two cupboards, red and blue. Maxi puts the chocolate in the red cupboard and goes out to play.

While Maxi is gone, mother takes the chocolate out of the red cupboard, uses some of it to bake a cake, and then puts the rest in the blue cupboard.

Now Maxi comes back from play and wants the chocolate

Now Maxi comes back from play and wants the chocolate

Where will Maxi look for the chocolate?

Now Maxi comes back from play and wants the chocolate

Where will Maxi look for the chocolate?

Children at the age of five or more say, *In the red cupboard.*

Now Maxi comes back from play and wants the chocolate

Where will Maxi look for the chocolate?

Children at the age of five or more say, *In the red cupboard.*

But children up to the age of three or four say, *Maxi will look in the blue cupboard.*

Now Maxi comes back from play and wants the chocolate

Where will Maxi look for the chocolate?

Children at the age of five or more say, *In the red cupboard.*

But children up to the age of three or four say, *Maxi will look in the blue cupboard.*

What three year old children lack, according to psychologists Premack and Woodruff is a Theory of Mind

Now Maxi comes back from play and wants the chocolate

Where will Maxi look for the chocolate?

Children at the age of five or more say, *In the red cupboard.*

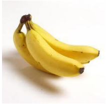
But children up to the age of three or four say, *Maxi will look in the blue cupboard.*

What three year old children lack, according to psychologists Premack and Woodruff is a Theory of Mind
But why do we infer what Maxi believes from what he would do?

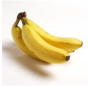
Animal Cognition

Do animals have a Theory of Mind?

BBBBBBBBBBBBBBBBBBBBBBBBBBBBBB



Food 1



Food 2

What chimps think about other chimps

In the last slide, the chimp at the bottom is subservient to the dominant chimp at the top and has to decide which group of bananas to go for.

What chimps think about other chimps

In the last slide, the chimp at the bottom is subservient to the dominant chimp at the top and has to decide which group of bananas to go for. In experiments, the sub-chimp tends to go for Food 1 which the dom-chimp cannot see. Is there use of epistemic logic by the sub-chimp? This is an issue of some controversy

Inducing false beliefs in the tigers of the Sundarbans

The Sundarbans are an area at the border of India and Bangladesh where lush forests grow and tiger attacks on humans have been common.

Inducing false beliefs in the tigers of the Sundarbans

The Sundarbans are an area at the border of India and Bangladesh where lush forests grow and tiger attacks on humans have been common.

Fishermen and bushmen then created masks made to look like faces to wear on the back of their heads because tigers always attack from behind. This worked for a short time.

Inducing false beliefs in the tigers of the Sundarbans

The Sundarbans are an area at the border of India and Bangladesh where lush forests grow and tiger attacks on humans have been common.

Fishermen and bushmen then created masks made to look like faces to wear on the back of their heads because tigers always attack from behind. This worked for a short time.

In 1987 no one wearing a mask was killed by a tiger, but 29 people without masks were killed.

Inducing false beliefs in the tigers of the Sundarbans

The Sundarbans are an area at the border of India and Bangladesh where lush forests grow and tiger attacks on humans have been common.

Fishermen and bushmen then created masks made to look like faces to wear on the back of their heads because tigers always attack from behind. This worked for a short time.

In 1987 no one wearing a mask was killed by a tiger, but 29 people without masks were killed.

Unfortunately the tigers eventually realized it was a hoax, and the attacks resumed.

The tiger in the bathroom

Suppose I know that T there is a tiger in your bathroom. I also know that you need to go.

If $\sim K_y(T)$ then you will proceed to the bathroom.

If $K_y(T)$ then you will go the neighbor's apartment and ask if you can use his bathroom. Or perhaps you will call your mother for advice.

So I can infer what you know from what you do.

Maxi again

As we saw we infer what the children believed about what Maxi believed from what they thought he would do to find the chocolate.

Savage showed us how to infer an agent's utilities and the agent's subjective probabilities from the agent's choices (actions). There are some problems with Savage's theory raised by Allais, Ellsberg, Kahneman and Tversky. But the theory is still respected at least as a first approximation.

Savage's theory was in the context of decision theory. Can we come up with such an account in the multi-agent case? In a situation where there are several agents, can we infer what j believes is believed by i from j's actions?

The florist example and the Maxi examples showed that sometimes we can.

A formal framework

We have n players and some propositions P about the world whose truth value they may or may not know. T is all truth assignments on P.

We define an epistemic game with n players to be a map F from T (truth assignments) and S (stategy profiles) to P (payoff profiles) So $(F(t, s))_i$ is the payoff to player i when the truth values are according to t and the strategy profile is s.

We let $s_i^- = s$" to mean the strategy profile of all players other than i. We will drop the subscript i when clear from the context.

Let s, s' be strategies for i. we let $s <_t s'$ to mean $(\forall s")(F(t,(s,s")) < F(t,(s',s"))$ (We will usually assume that payoffs for i are never the same so that we need not worry about $<$ and $\leq$.) In other words s' is better than s no matter what the other players do.
We write $s <_\varphi s'$ t mean that for all $t \models \varphi$, $s <_t s'$

Theorem

If $s <_\varphi s'$ and $\psi \models \varphi$ then $s <_\psi s'$)
If $s <_\varphi s'$ and $s <_\psi s'$) then $s <_{\varphi \vee \psi} s'$

Corollary

The set $\{\varphi | s <_\varphi s'\}$ is a filter in the boolean algebra.

Note that if a rational player knows φ and $s <_\varphi s'$ then the agent will not play s.
Moreover if j knows that i knows φ and $s <_\varphi s'$ then j knows that the agent i will not play s, and j only needs to respond to strategies other than s. Indeed what j knows about what other players know allows j to reduce the strategy profiles that he needs to respond to.

Goal To define the notion of rationalizability relative to a given Kripke structre and an epistemic game.

Conjecture: Two non-bisimilar Kripke structures yield different sets of rationalizabile strategies in *at least one* epistemic game.

Conjecture: Every strategy rationalizable relative to a Kripke structure is rationalizable in the usual sense. The reverse of course is not true.

Part II

This part reports joint work with Cagil Tasdemir and Andreas Witzel.
The results have appeared in *International Game Theory Review*

Knowledge leads to action

So far a great deal of work has been done on reasoning about knowledge, studying how knowledge is transferred and updated, and even some paradoxical theories of knowledge.

Knowledge leads to action

So far a great deal of work has been done on reasoning about knowledge, studying how knowledge is transferred and updated, and even some paradoxical theories of knowledge.

But we do not acquire knowledge, or transmit it to others just to pass the time.

Knowledge leads to action

So far a great deal of work has been done on reasoning about knowledge, studying how knowledge is transferred and updated, and even some paradoxical theories of knowledge.

But we do not acquire knowledge, or transmit it to others just to pass the time.

Knowledge means influence and power.

It is a commonplace that what we do depends on what we know. And given that most of us have at least the rudiments of a *theory of mind* (cf. Premack and Woodruff) we also know that what others do will depend on what *they* know.

Wife and husband - knowledge and temperament

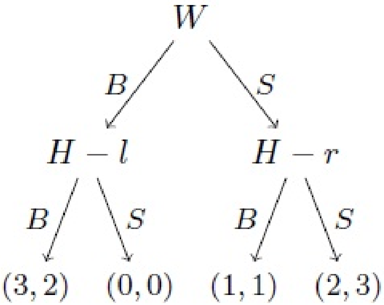


Fig. 1

In the last figure we assume that the wife moves first and the husband after.

In the last figure we assume that the wife moves first and the husband after.

We consider various scenarios involving the husband's knowledge and temperament. We assume that the wife knows the husband's payoffs and temperament and he does not know hers.

Case 1) Husband does not know wife's move (and she knows this).
a) He is aggressive. Then being aggressive, he will choose S (Stravinsky) for his move since the highest possible payoff is 3. Anticipating his move, she will also choose S, and they will end up with payoffs of (2,3).

b) If the husband is conservative, then not knowing what his wife chose, he will choose B since the minimum payoff of 1 is better than the minimum payoff of 0. Anticipating this, the wife will also choose B and they will end up with (3,2).

2) Finally if the husband will know what node he will be at, then the wife will choose B, the husband will also choose B and they will end up at (3,2).

The knowledge manipulator - KM

We consider now the question of how KM can create these various knowledge scenarios of the last example.

At each node, the manipulator has a set of n-tuples of signals such that if s is such an n-tuple then s_i is sent to player i. The set is common knowledge but the particular n-tuple chosen on any particular occasion is not.

KM is capable of creating all these three situations by means of signals, as well as the one we did not mention where the husband does not know but the wife does not know that he will not.
For case 1a), $s(H - l) = (l, a)$ and $s(H - r) = (r, a)$. The wife knows (if she did not already) which node they are at, but the husband will not.

For case 2, $s(H - l) = (l, l)$ and $s(H - r) = (r, r)$. Both will know which node they are at.
Finally if KM wants the wife to be in doubt whether the husband knows, he could make $s(H - l) = \{(l, l), (l, a)\}$ and $s(H - r) = \{(r, r), (r, a)\}$. Then if the wife chose left and receives an l, she will not know if the husband got an l or the neutral a. If KM does send (l, l) then the husband will know, but will also know that his wife did not know whether he *would* know.

Theorem

Any knowledge situation represented by a finite Kripke structure M can be created in a single signaling step.

Proof

The knowledge manipulator (KM) picks a world w in M and sends player i the signal (M, X_i^w) where $X_i^w = \{v | wR_i v\}$ and R_i is the accessibility relation of player i. This tells us the local history of i. The global history H_w is $(M, w, X_1^w, ..., X_n^w)$. An atomic formula p holds at H_w iff it holds at w. We now show by induction on the complexity of the formula F that $M, w \models F$ iff $H_w \models F$ where we use the [PR] history-based semantics to define $H_w \models F$

We have already noticed that atomic formulas behave correctly (by stipulation) and truth functions are clear. Consider $F = K_i(B)$.

- $M, w \models K_i(B)$ iff
- $(\forall v)(wR_i v \to M, v \models B)$ iff
- (IH) $(\forall v)(wR_i v \to H_v \models B)$ iff
- $(\forall H_v)(\lceil_i(H_v) = \lceil_i(H_w) \to H_v \models B)$
- iff $H_w \models K_i(B)$

We recall that $\lceil_i$ is the projection for i of a global history, and two histories H_w, H_v here have the same projection for i iff $X_i^w = X_i^v$ iff $wR_i v$ $\Box$