Knowledge, Games and the World

Rohit Parikh

City University of New York
Brooklyn College and CUNY Graduate Center

NASSLLI
July 11, 2016

Knowledge in Society

The peculiar character of the problem of a rational economic order is determined precisely by the fact that the knowledge of the circumstances of which we must make use never exists in concentrated or integrated form, but solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess.

Knowledge in Society

The peculiar character of the problem of a rational economic order is determined precisely by the fact that the knowledge of the circumstances of which we must make use never exists in concentrated or integrated form, but solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess.

The economic problem of society is thus not merely a problem of how to allocate "given" resources – if "given" is taken to mean given to a single mind which deliberately solves the problem set by these "data."

Knowledge in Society

The peculiar character of the problem of a rational economic order is determined precisely by the fact that the knowledge of the circumstances of which we must make use never exists in concentrated or integrated form, but solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess.

The economic problem of society is thus not merely a problem of how to allocate "given" resources – if "given" is taken to mean given to a single mind which deliberately solves the problem set by these "data." It is rather a problem of how to secure the best use of resources known to any of the members of society, for ends whose relative importance only these individuals know.

F. Hayek

Individualism and Economic Order

Topics

- Learning from situations
- Common Knowledge
- Theory of Mind
- Rational Behavior
- Campaigning

Learning from what is said

Three people A, B, C walk into a coffee shop. One of them orders cappuccino, one orders tea, and one orders icecream. The waiter goes away and after ten minutes *another* waiter arrives with three cups. “Who has the cappuccino?” “I do,” says A. “Who has the tea?” “I do,” says C.

Learning from what is said

Three people A, B, C walk into a coffee shop. One of them orders cappuccino, one orders tea, and one orders icecream. The waiter goes away and after ten minutes *another* waiter arrives with three cups. "Who has the cappuccino?" "I do," says A. "Who has the tea?" "I do," says C.

Will the waiter ask a third question?"

Consider the possible situations for waiter 2. They are

1) CTI **2) CIT**
3) TCI **4) TIC**
5) ICT **6) ITC**

Consider the possible situations for waiter 2. They are

1) CTI   2) CIT
3) TCI   4) TIC
5) ICT   6) ITC

When A says that he has the cappuccino, 3,4,5,6 are eliminated.

Consider the possible situations for waiter 2. They are

1) CTI 2) CIT
3) TCI 4) TIC
5) ICT 6) ITC

When A says that he has the cappuccino, 3,4,5,6 are eliminated. The waiter now has,

1) CTI, 2) CIT

Consider the possible situations for waiter 2. They are

1) CTI   2) CIT
3) TCI   4) TIC
5) ICT   6) ITC

When A says that he has the cappuccino, 3,4,5,6 are eliminated. The waiter now has,

1) CTI,   2) CIT

When C says that he has the tea, 1 is eliminated.

Now 2 alone is left and the waiter knows that B has the icecream.

A butler enters a hotel room to clean it and make the bed, but he encounters a woman guest, coming out of the bathtub and not even wearing a towel.
"Excuse me, sir," says the butler, and leaves the room.

A butler enters a hotel room to clean it and make the bed, but he encounters a woman guest, coming out of the bathtub and not even wearing a towel.
"Excuse me, sir," says the butler, and leaves the room.

Why did the butler say, "Excuse me, **sir**"?

A butler enters a hotel room to clean it and make the bed, but he encounters a woman guest, coming out of the bathtub and not even wearing a towel.
"Excuse me, sir," says the butler, and leaves the room.

Why did the butler say, "Excuse me, **sir**"?

In the woman's mind there were two possibilities.

A butler enters a hotel room to clean it and make the bed, but he encounters a woman guest, coming out of the bathtub and not even wearing a towel.
"Excuse me, sir," says the butler, and leaves the room.

Why did the butler say, "Excuse me, **sir**"?

In the woman's mind there were two possibilities.

S1 = "The butler saw her clearly"

A butler enters a hotel room to clean it and make the bed, but he encounters a woman guest, coming out of the bathtub and not even wearing a towel.
"Excuse me, sir," says the butler, and leaves the room.

Why did the butler say, "Excuse me, **sir**"?

In the woman's mind there were two possibilities.

S1 = "The butler saw her clearly"
S2 = "The butler did not see her clearly"

The butler's remark eliminated S1 and saved her from embarrassment.

Kripke structures

Kripke structure M for knowledge for n knowers consists of a space W of states and for each knower i a relation $R_i \subseteq W \times W$.
There is a map π from $W \times A \longrightarrow \{0, 1\}$ which decides the truth value of atomic formulas at each state.

We now define the truth values of formulas as follows:

1. $M, w \models P$ iff $\pi(w, P) = 1$
2. $M, w \models \neg A$ iff $M, w \not\models A$
3. $M, w \models A \wedge B$ iff $M, w \models A$ and $M, w \models B$
4. $M, w \models K_i(A)$ iff $(\forall t)(wR_i t \rightarrow M, t \models A)$

$K_i(A)$ holds at w, (i knows A at w) iff A holds at all states t which are R_i accessible from w.

Some Consequences

If R_i is reflexive then we will get $K_i(A) \to A$ (veridicality) as a consequnce.
Moreover, regardless of the propeties of R_i, we have,

1. If A is logically valid, then A is known
2. If A and $A \to B$ are known, then so is B

Some Consequences

If R_i is reflexive then we will get $K_i(A) \rightarrow A$ (veridicality) as a consequnce.
Moreover, regardless of the propeties of R_i, we have,

1. If A is logically valid, then A is known
2. If A and $A \rightarrow B$ are known, then so is B

This is the well known problem of Logical Omniscience since we are attributing knowledge properties to agents which they do not actually have.

Some Consequences

If R_i is reflexive then we will get $K_i(A) \rightarrow A$ (veridicality) as a consequnce.
Moreover, regardless of the propeties of R_i, we have,

1. If A is logically valid, then A is known
2. If A and $A \rightarrow B$ are known, then so is B

This is the well known problem of Logical Omniscience since we are attributing knowledge properties to agents which they do not actually have.

Still, in small settings, such assumptions are reasonable.

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.

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.

We apply this insight to a problem similar to the well known Muddy Children problem.

Numerical Foreheads

Two players Ann and Bob are told that the following will happen. Some positive integer n will be chosen and *one* of n, $n+1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

Thus if Ann has 5 and Bob has 6, then Ann knows that her number is either 5 or 7 and Bob knows that his number is either 6 or 4.

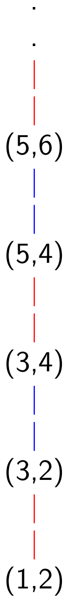
After this is done, they are asked repeatedly, beginning with Ann, if they know what their own number is.

Theorem 1: In those cases where Ann has the even number, the reponse at the nth stage will be, “my number is $n + 1$”, and in the other cases, the response at the $(n + 1)$st stage will be “my number is $n + 1$”. In either case, it will be the person who *sees* the smaller number, who will respond first.

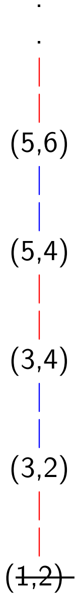
Theorem 1: In those cases where Ann has the even number, the reponse at the nth stage will be, "my number is $n+1$", and in the other cases, the response at the $(n+1)$st stage will be "my number is $n+1$". In either case, it will be the person who *sees* the smaller number, who will respond first.

Homework: Prove this theorem!

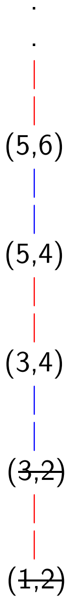
Start situation



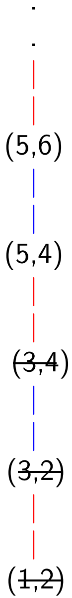
Bob has just said, I don't know my number



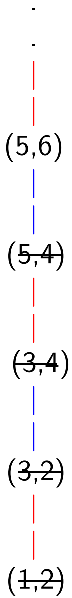
Ann said no also



Bob said a second "no"



Ann said a second “no”



Bob knows his number is 6

Common Knowledge

Defined independently by Lewis and Schiffer. (But apparently the notion already occurs in the doctoral dissertation of the philosopher Robert Nozick). Used first in Game theory by Aumann.

A is common knowledge in a group G if everyone knows it, everyone knows that everyone knows it, everyone knows that everyone knows that everyone knows it,.....

Aumann showed that common knowledge implies same opinion.

Common Knowledge

Defined independently by Lewis and Schiffer. (But apparently the notion already occurs in the doctoral dissertation of the philosopher Robert Nozick). Used first in Game theory by Aumann.

A is common knowledge in a group G if everyone knows it, everyone knows that everyone knows it, everyone knows that everyone knows that everyone knows it,.....

Aumann showed that common knowledge implies same opinion.

Geanakoplos and Polemarchakis showed that communication between two agents leads to common knowledge and same opinion.

Common Knowledge

Defined independently by Lewis and Schiffer. (But apparently the notion already occurs in the doctoral dissertation of the philosopher Robert Nozick). Used first in Game theory by Aumann.

A is common knowledge in a group G if everyone knows it, everyone knows that everyone knows it, everyone knows that everyone knows that everyone knows it,.....

Aumann showed that common knowledge implies same opinion.

Geanakoplos and Polemarchakis showed that communication between two agents leads to common knowledge and same opinion.

Parikh and Krasucki showed that among n agents communicating in pairs, common opinion about some quantity can come about without most agents communicating with others.

Aumann's argument

Column

Row

$v_{1,1}$	$v_{1,2}$	$v_{1,3}$	$v_{1,4}$
$v_{2,1}$	$v_{2,2}$	$v_{2,3}$	$v_{2,4}$
$v_{3,1}$	$v_{3,2}$	$v_{3,3}$	$v_{3,4}$
$v_{4,1}$	$v_{4,2}$	$v_{4,3}$	$v_{4,4}$

Now Row's value v is
$v_1 = (1/4)[v_{1,1,} + v_{1,2} + v_{1,3} + v_{1,4}]$
And Column's value w is
$w_1 = (1/4)[(v_{1,1,} + v_{2,1} + v_{3,1} + v_{4,1}]$

Now Row's value v is

$v_1 = (1/4)[v_{1,1,} + v_{1,2} + v_{1,3} + v_{1,4}]$

And Column's value w is

$w_1 = (1/4)[(v_{1,1,} + v_{2,1} + v_{3,1} + v_{4,1}]$

But, these values are common knowledge. So,

Now Row's value v is

$v_1 = (1/4)[v_{1,1,} + v_{1,2} + v_{1,3} + v_{1,4}]$

And Column's value w is

$w_1 = (1/4)[(v_{1,1,} + v_{2,1} + v_{3,1} + v_{4,1}]$

But, these values are common knowledge. So, v_1 must equal v_2 where

$v_2 = (1/4)[v_{2,1,} + v_{2,2} + v_{2,3} + v_{2,4}]$

and similarly for v_3 and v_4. The v_i are all the same!

Now Row's value v is

$v_1 = (1/4)[v_{1,1,} + v_{1,2} + v_{1,3} + v_{1,4}]$

And Column's value w is

$w_1 = (1/4)[(v_{1,1,} + v_{2,1} + v_{3,1} + v_{4,1}]$

But, these values are common knowledge. So, v_1 must equal v_2 where

$v_2 = (1/4)[v_{2,1,} + v_{2,2} + v_{2,3} + v_{2,4}]$

and similarly for v_3 and v_4. The v_i are all the same!

Thus v_1 must equal

$(1/16)[\Sigma v_{i,j} : i \leq 4, j \leq 4]$

and similarly for w_1.

Now Row's value v is
$v_1 = (1/4)[v_{1,1,} + v_{1,2} + v_{1,3} + v_{1,4}]$
And Column's value w is
$w_1 = (1/4)[(v_{1,1,} + v_{2,1} + v_{3,1} + v_{4,1}]$

But, these values are common knowledge. So,
v_1 must equal v_2 where
$v_2 = (1/4)[v_{2,1,} + v_{2,2} + v_{2,3} + v_{2,4}]$
and similarly for v_3 and v_4. The v_i are all the same!

Thus v_1 must equal
$(1/16)[\Sigma v_{i,j} : i \leq 4, j \leq 4]$
and similarly for w_1.

Thus $v = w$.

Using Aumann's reasoning,
Milgrom and Stokey proved a famous
No Trade theorem!

If A is selling a stock to B, and B is buying it, then obviously A thinks the stock will go down and B thinks it will go up. But this fact is common knowledge! By a proof based on Aumann, it cannot be common knowledge that they have different views of the stock and the sale cannot take place.

“I’d never join any club that would have me for a member”

Groucho Marx

But what if the value is not common knowledge?

Will communication help?

GP argument

Column

Row

2	3	5	4
7	8	9	10
3	2	5	4
5	4	3	2

At this point Row announces that her expected value is 3.5, and column eliminates row 2

	Column			
Row	(2)	3	5	4
	~~7~~	~~8~~	~~9~~	~~10~~
	3	2	5	4
	5	4	3	2

Now column announces that his value is 3.33, and row eliminates columns 2,3

Column

Row				
	2	3	5	4
	7	8	9	10
	3	2	5	4
	5	4	3	2

Now Row announces his value as 3 = (2+4)/2 and Column eliminates row 3, 4, announcing his value as 2.

Column

Row

2	3	5	4
7	8	9	10
3	2	5	4
5	4	3	2

At this point Row eliminates column 4, also announces his value at 2, and they have consensus.

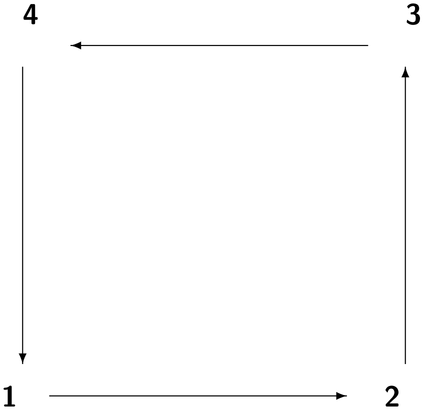
	Column			
Row	2	3	5	4
	7	8	9	10
	3	2	5	4
	5	4	3	2

A brief overview of the [PK] result:
Suppose we have n agents connected in a strongly connected graph. They all share initial probability distribution, but have now received, each of them, a finite amount of private information. Thus their estimate of the probability of some event or the expected value of some random variable v may now be different.

Let g be a function which, at stage n picks out a sender $s(n)$ and a recipient $r(n)$. $s(n)$ sends his latest value of v to $r(n)$ who then revises *her* valuation of v.

If the graph G is strongly connected, and for each pair of connected agents i, j, i repeatedly sends his value of v to j, then eventually all estimates of the value of v become equal.

Parikh-Krasucki result



Sketch of proof of the P-K result

Each agent i has a personal partition $\mathcal{P}_i$. Let $\mathcal{P}$ be the common refinement of the $\mathcal{P}_i$. The $\mathcal{P}_i$ are finite and so is $\mathcal{P}$.

Sketch of proof of the P-K result

Each agent i has a personal partition $\mathcal{P}_i$. Let $\mathcal{P}$ be the common refinement of the $\mathcal{P}_i$. The $\mathcal{P}_i$ are finite and so is $\mathcal{P}$. Each time an agent receives information, (s)he deletes at least one piece of $\mathcal{P}$ from her current set of possibilitities.

Sketch of proof of the P-K result

Each agent i has a personal partition $\mathcal{P}_i$. Let $\mathcal{P}$ be the common refinement of the $\mathcal{P}_i$. The $\mathcal{P}_i$ are finite and so is $\mathcal{P}$. Each time an agent receives information, (s)he deletes at least one piece of $\mathcal{P}$ from her current set of possibilitities.

After a finite number of steps each agent has removed as many pieces as she is ever going to and the situation stabilizes.

Sketch of proof of the P-K result

Each agent i has a personal partition $\mathcal{P}_i$. Let $\mathcal{P}$ be the common refinement of the $\mathcal{P}_i$. The $\mathcal{P}_i$ are finite and so is $\mathcal{P}$. Each time an agent receives information, (s)he deletes at least one piece of $\mathcal{P}$ from her current set of possibilitities.

After a finite number of steps each agent has removed as many pieces as she is ever going to and the situation stabilizes.

The non-trivial part of the proof consists of showing that at that stage, all agents have the same value of the parameter in question.

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.*

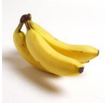
But children up to the age of three or four say, *Maxi will look in the blue cupboard.*

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

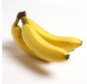
Animal Cognition

Do animals have a Theory of Mind?

BBBBBBBBBBBBBBBBBBBBBBBBBBBBB



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.

Tiger's Decision problem

	Face	No face
Attack	-10	10
No attack	0	0

Rational Behavior

Suppose that a man is hungry (desire), knows that there is a restaurant within two blocks (knowledge) and is able to walk the two blocks (ability). Then we can predict that he will go to that restaurant. Knowledge plus desire plus ability lead to action. Suppose that there are two restaurants nearby and we do not know his taste. Then we can predict that he will go to one, but not know which one. But we know that he will go to one of them. The strategy of not going to either is dominated for him by going to restaurant A as well as by going to restaurant B.

Note that *for him* going to restaurant A might dominate going to B. But we do not know that. His personal notion of domination is different from his notion as seen by us.

Suppose Jill knows that he prefers A and we do not. We both find out that he went to B.

Note that *for him* going to restaurant A might dominate going to B. But we do not know that. His personal notion of domination is different from his notion as seen by us.

Suppose Jill knows that he prefers A and we do not. We both find out that he went to B.

Jill will say, “I wonder why he went to B.” But we will not.

NB: *Intuitively a strategy* s *is dominated by* s' *if the outcome from* s' *is always better than the outcome from* s. *And here we interpret "always" as "in all situations compatible with our knowledge." Thus the notion of* dominated *has an epistemic (or doxastic) component.*

Suppose now that *we* know that the man is hungry, can walk two blocks and also know that he does not know about the restaurant. Then we will *tell* him about the restaurant. If we are nice people then the strategy of not telling him anything is dominated by telling him about the restaurants. If we do not tell him and he finds out later that we knew then he will reproach us. So telling is better.

If on the other hand we know he already knows about the restaurant then it would be rude to tell him anyway. He might be offended.

Thus our actions take place in a world of desires, knowledge (or beliefs) and abilities. And quite often not only our own beliefs and desires are involved but also what we know about the desires and beliefs of others.

Leonard Savage [15] worked out a theory in which by observing an agent's willingness to accept or reject certain bets we can discover both his beliefs (subjective probability) and his desire (utility). This theory has been questioned and has some difficulties pointed out by Ellsberg, Allais, and Kahneman and Tversky. But the theory is much respected and still taught routinely.

Thus our actions take place in a world of desires, knowledge (or beliefs) and abilities. And quite often not only our own beliefs and desires are involved but also what we know about the desires and beliefs of others.

Leonard Savage [15] worked out a theory in which by observing an agent's willingness to accept or reject certain bets we can discover both his beliefs (subjective probability) and his desire (utility). This theory has been questioned and has some difficulties pointed out by Ellsberg, Allais, and Kahneman and Tversky. But the theory is much respected and still taught routinely.

But Savage did not have a theory of what we do when other agents are involved and we know something about their desires and beliefs.

Two tools have come into prominence since Savage. One is *Game Theory* which deals with many agents taking actions, taking into account the other agents' desires. Typically the situation is described either in terms of a payoff matrix for a normal form game or in terms of a tree of actions and choices in the case of an extensive form game.

Two tools have come into prominence since Savage. One is *Game Theory* which deals with many agents taking actions, taking into account the other agents' desires. Typically the situation is described either in terms of a payoff matrix for a normal form game or in terms of a tree of actions and choices in the case of an extensive form game.

Another tool is *Epistemic Logic* which allows us to represent the states of knowledge or belief of many agents. We can represent "Ravi knows that it is raining" or "Yang does not know that Ravi knows that it is raining" using Kripke structures.

However, these two tools have not been brought together in the way we need. Of course there is the field of Epistemic Game Theory but it investigates certain models originally developed by Harsanyi and investigated further by Brandenburger and others. See [8] for example. The more down to earth area which we are proposing does not exist.

Here is a story about the TV detective Adrian Monk. A woman has fallen off a fifteenth floor balcony and is lying on the pavement, dead of course. A policeman arrives and a little later Adrian Monk arrives.

“We don’t know yet if it was murder or suicide” says the policeman.

Here is a story about the TV detective Adrian Monk. A woman has fallen off a fifteenth floor balcony and is lying on the pavement, dead of course. A policeman arrives and a little later Adrian Monk arrives.

“We don’t know yet if it was murder or suicide” says the policeman.

“It was murder,” says Monk.

Here is a story about the TV detective Adrian Monk. A woman has fallen off a fifteenth floor balcony and is lying on the pavement, dead of course. A policeman arrives and a little later Adrian Monk arrives.

“We don’t know yet if it was murder or suicide” says the policeman.

“It was murder,” says Monk.

“But you just arrived! How can you know?”

Here is a story about the TV detective Adrian Monk. A woman has fallen off a fifteenth floor balcony and is lying on the pavement, dead of course. A policeman arrives and a little later Adrian Monk arrives.

"We don't know yet if it was murder or suicide" says the policeman.

"It was murder," says Monk.

"But you just arrived! How can you know?"

"She was in the middle of painting her nails," says Monk.

Now we all agree that a woman does not paint half her nails and then commit suicide without painting the other half. Perhaps she was in despair because she ran out of paint but that is not very likely. Murder is the far more plausible explanation. Since Monk is unable to answer the question, “Why might she commit suicide at *that* moment?” he concludes that it was murder.

When people or animals or children do something, we look for an explanation in terms of their beliefs and desires. Often the desires are known to us. A hungry animal wants to eat. A tired child wants to go to bed. And then from their actions we conclude what their beliefs are. Sometimes we go in the other direction and manipulate their beliefs so as to get them to act in a certain way. "If you are a good girl, you will get ice cream after the dinner!"

This work is about inferring beliefs from actions, and inferring preferences from observing actions. We offer some examples from literature, from real life and offer a formal framework. But we will be guided by the following intuition. If, given an agent's desires and beliefs, action s is definitely worse than action s' then the agent will not do action s. If the agent *does* do action s then we can conclude that we were wrong about the beliefs or desires.

This work is about inferring beliefs from actions, and inferring preferences from observing actions. We offer some examples from literature, from real life and offer a formal framework. But we will be guided by the following intuition. If, given an agent's desires and beliefs, action s is definitely worse than action s' then the agent will not do action s. If the agent *does* do action s then we can conclude that we were wrong about the beliefs or desires.

Of course the agent might be irrational. All of us have met irrational people. But the scenarios we consider are so simple that irrationality is unlikely to be an explanation.

With language using creatures, we do sometimes just ask, to find out their beliefs. But this method is not available with small children or animals. And it may not work even with adults who might have a reason to deceive. So a formal theory of inferring beliefs from actions is likely to have value.

Cardinal and Ordinal Utilties

Suppose Bob prefers chololate icecream to vanilla and vanilla to strawberry.
His utilities for the three icecreams might be 10, 9, 1
Or they might be 10, 2, 1.

Cardinal and Ordinal Utilties

Suppose Bob prefers chololate icecream to vanilla and vanilla to strawberry.
His utilities for the three icecreams might be 10, 9, 1
Or they might be 10, 2, 1.
In both cases the question "Would you like chocolate or vanilla?" would be answered the same way.

Cardinal and Ordinal Utilties

Suppose Bob prefers chololate icecream to vanilla and vanilla to strawberry.
His utilities for the three icecreams might be 10, 9, 1
Or they might be 10, 2, 1.
In both cases the question “Would you like chocolate or vanilla?” would be answered the same way.
But in the first case, Bob would prefer vanilla to a 50-50 chance of chocolate and strawberry.
In the second case he would prefer a 50-50 chance of chocolate and strawberry to vanilla
But in social situations, cardinal utilities are a bit hard to find out.

Ann and Bob

Ann lives in San Francisco and Bob is visiting her from out of town.

Ann and Bob

Ann lives in San Francisco and Bob is visiting her from out of town.

After giving him dinner, Ann asks, “Would you like some icecream?”

Ann and Bob

Ann lives in San Francisco and Bob is visiting her from out of town.

After giving him dinner, Ann asks, "Would you like some icecream?"
"Do you have chocolate?" says Bob.
"I am sorry I do not but I do have vanilla and strawberry" says Ann.

Ann and Bob

Ann lives in San Francisco and Bob is visiting her from out of town.

After giving him dinner, Ann asks, “Would you like some icecream?”
“Do you have chocolate?” says Bob.
“I am sorry I do not but I do have vanilla and strawberry” says Ann.
“Vanilla, then” says Bob.
Now Ann knows that for Bob the utilities have the order
$u_c > u_v > u_s$
but she does not know the cardinal utilities.

Further queries

“Oh, I am sorry, I do have chocolate. Would you like vanilla or a 50-50 chance of chocolate and strawberry?”
Bob is puzzled but says, “Vanilla.”

Further queries

"Oh, I am sorry, I do have chocolate. Would you like vanilla or a 50-50 chance of chocolate and strawberry?"
Bob is puzzled but says, "Vanilla."
"And how about vanilla versus a 70% chance of chocolate and 30% chance of strawberry?"
"*What is the matter with Ann?* " says Bob to himself and then says aloud,
"Actually the doctor told me to avoid icecream. How about just coffee?"

Further queries

"Oh, I am sorry, I do have chocolate. Would you like vanilla or a 50-50 chance of chocolate and strawberry?"

Bob is puzzled but says, "Vanilla."

"And how about vanilla versus a 70% chance of chocolate and 30% chance of strawberry?"

"*What is the matter with Ann?* " says Bob to himself and then says aloud,

"Actually the doctor told me to avoid icecream. How about just coffee?"

Ann now knows that $u_v > .5(u_c + u_s)$ and also knows that she will not find out more without being rude.

And this is why we will confine ourselves to ordinal utilities to understand human behavior.

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.

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.

Benedick's Decision problem

	love	nolove
propose	100	-20
nopropose	-10	0

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

Note that our utilities are ordinal and not cardinal. We often know other people's preferences but not how they would respond to bets.

As things were, neither of the strategies propose and nopropose dominated the other.
But once he comes to learn love, proposing dominates not proposing.

Note that proposing always dominated not proposing, in reality.But in ‘Benedick’s world’, neither dominated the other.

The play ends with all four lovers getting married.

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.

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.

The tiger in the bathroom

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

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.

Bathroom Decision problem

	tiger	no tiger
Own bathroom	-1000	10
Nbr bathroom	-10	-10

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 about i's beliefs from j's actions?

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

Behavior to Preferences

Jack to Bill; *I am sorry to hear about the fire at your warehouse last night*.

Bill: *Shhh! It is tonight!*

To understand the structure of the joke consider two similar examples.

Jack to Bill: *I understand your daughter's wedding was magnificent!*

Bill: *Actually the wedding is next week.*

Jack to Bill: *I understand your wife's surprise birthday party was magnificent!*

Bill: *Shhh! Actually the party is next week.*

Why is the first example a joke but not the other two?

And why does *Shhh!* occur in the first and third examples but not in the second?

The three examples together

Jack to Bill; *I am sorry to hear about the fire at your warehouse last night*.

Bill: *Shhh! It is tonight!*

_____

Jack to Bill: *I understand your daughter's wedding was magnificent!*

Bill: *Actually the wedding is next week.*

_____

Jack to Bill: *I understand your wife's surprise birthday party was magnificent!*

Bill: *Shhh! Actually the party is next week.*

In the first case and the third case, the event is not supposed to be common knowledge. In the second one it is fine if it is common knowledge.

In the first case and the third case, the event is not supposed to be common knowledge. In the second one it is fine if it is common knowledge.

But why is the first one the only one which is a joke?

In the first case and the third case, the event is not supposed to be common knowledge. In the second one it is fine if it is common knowledge.

But why is the first one the only one which is a joke?

Because in this case we are making an inference from action (or rather inaction) to preferences.

Compare

Jack to Bill: *Your hat is on fire.*

Compare

Jack to Bill: *Your hat is on fire.*

Bill: *I know.*

Compare

Jack to Bill: *Your hat is on fire.*

Bill: *I know.*

We do not know what to make of Bill's response!

Rationalizability

We assume that the games in question are generic so that no player has the same payoff in two different situations.

When an agent has two strategies s, s' and s' is better than s (given his information) then the rational agent will not play s.

Another agent, realizing this will conclude that he need not worry about the first agent playing s.

He may then realize that his own stategy t is worse than t' so *he* will not play t.

(t' was worse than t only if the first agent played s)

This phenomenon leads to the iterated elimination of dominated strategies and what remains are the rationalizable strategies.

The basic work was done by Bernheim and Pearce in 1984. We generalize their notion but modify it. Consider the game

	Left	Right
Top	$(0,1)$	$(10,1)$
Middle	$(4,1)$	$(4,1)$
Bottom	$(10,1)$	$(0,1)$

The strategy *Middle* is not rationalizable for row as it is dominated (in expected value) by

$$[.5 \times (\mathit{Top})] + [.5 \times (\mathit{Bottom})]$$

But we could consider *Middle* as rational if row is risk-averse!

Two Games

P

	left	right
top	$(2, 1)$	$(1, 2)$
bottom	$(3, 2)$	$(2, 1$

$\neg P$

	left	right
top	$(2, 1)$	$(1, 2)$
bottom	$(1, 2)$	$(2, 1)$

Suppose that P is true but no one knows this. Then all of *top, bottom, left, right* are rationalizable.
Suppose now that P is true and both row and column know this. But column does not know that row knows.
Now top is dominated by bottom but column does not know that row will not play top, and for him both left and right are still rational.
Suppose column knows that row knows P. Then column knows that ros will not play top, and he will then not play right. They will converge to (bottom, left).

Each time there is a growth in knowledge the set of rationalizable strategies decreases.

The less you know, the more rational you are!!

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 $s <_t^i s'$ to mean $(\forall s")(F(t,(s,s")_i < F(t,(s',s")_i)$

(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 will also drop i when clear from the context.

If ϕ is a formula, we write $s <_\phi s'$ t mean that

for all $t \models \phi$, $s <_t^i s'$

So if i knows ϕ and i is rational, i will not play s

Theorem

If $s <_\phi s'$ *and* $\psi \models \phi$ *then* $s <_\psi s'$

If $s <_\phi s'$ *and* $s <_\psi s'$ *then* $s <_{\phi \vee \psi} s'$

Corollary

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

Note that if a rational player knows ϕ and $s <_\phi s'$ then the agent will not play s.
Moreover if j knows that i knows ϕ and $s <_\phi s'$ 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 structure (knowledge situation) and an epistemic game.

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

An *infon* α for agent i is a pair (t, S) where t is a truth assignment and S is an (n-1)-tuple of the other players' straregies.

In a generic game an infon α generates a unique strategy $b(\alpha)$ for agent i which yields the highest value given the infon.

A *state of knowledge* for agent i is a set X of infons.

A strategy s is *rational* for agent i relative to a state of knowledge X if $s \in \{b(\alpha) | \alpha \in X\}$

If $X \subseteq Y$ and s is rational relative to X then it is rational relative to Y.

The less you know, the more rational you are!

A strategy profile $(s_1, ..., s_n)$ is rational for the agents relative to a tuple of knowledge $(X_1, ..., X_n)$ if each s_i is rational relative to X_i.

However, not all n-tuples $(X_1, ..., X_n)$ are possible. For instance if i knows that j knows P then j cannot be playing a strategy t which is dominated when P is true. And i herself cannot be playing a strategy which is dominated when j is not playing t.

So there are connections among the X_i which we have yet to fullly investigate.

Campaigning

We prove an abstract theorem which shows that under certain circumstances, a candidate running for political office should be as explicit as possible in order to improve her impression among the voters.

But this result conflicts with the perceived reality that candidates are often cagey and reluctant to take stances except when absolutely necessary. Why this hesitation on the part of the candidates? We offer some explanations.

Introduction

In [3] Dean and Parikh considered a political candidate campaigning to be elected. The candidate's chances of being elected depend on how various groups of voters perceive her, and how they perceive her depends on what she has said. Different groups of voters may have different preferences and a statement preferred by one group of voters may be disliked by another. They consider three types of voters:

- Optimistic: those who are willing to think the best of the candidate,
- Pessimistic: those who are inclined to expect the worse, and
- Expected value voters: those who average over various possibilities which may come about if the candidate is elected.

They show that if the voters are expected value voters, then the candidate is best off being as explicit as possible.
While interesting, this result is counter intuitive in that politicians are often cagey and avoid committing themselves on issues. What explains this?

In this paper, we extend the previous work by Dean and Parikh in two ways.

- We use the Fubini theorem to provide a very general, abstract version of their *(best to be explicit)* result which applies to the case where we consider a single candidate who merely wants to improve her status among the voters.
- Later we introduce a belief set for the candidate and impose the condition that she would not make any statements against her honest beliefs. We also take into account the scenarios where there may be optimistic voters, or where other problems, like voters staying at home, enter. Such scenarios offer a possible explanation for the cageyness of some candidates.

A General Theorem

This section presumes that the candidate has made some statements in the past and as a consequence the voters can assume that if she is elected then the state of the world (or nation) will belong to a certain set Z of (possible) states, those which agree with the statements she has made.[1]

A voter may have different views about the different states in Z, finding some good and others not so good. Also, different voters may have different views about the same state. If the candidate reveals more of her views on some issues, then the set of states compatible with her views will shrink. Some voters may be displeased, finding that their favorite states are no longer compatible with her new stance. Other voters may be pleased seeing that some states they feared are no longer in the running.

[1] If the statements which she has already made constitute a set T, then Z is the set of those states ω which satisfy T.

How should the candidate speak so as to improve her overall position given these forces pulling in different directions?
Let V denote the set of voters, and Ω the set of the states of the world.[2] Both V, Ω are assumed to be compact (i.e., closed and bounded) subsets of some Euclidean space $\mathbb{R}^n$.
Let the satisfaction function $s : V \times \Omega \mapsto \mathbb{R}$ represent the extent to which voter $v \in V$ *likes* the state $\omega \in \Omega$.

[2]Note that we are not making assumptions like single peak preference. The model we use allows for a candidate who is a social conservative and an economic liberal (e.g. Carter), or a socially liberal candidate who is hawkish on foreign policy (Johnson).

What is the candidate's current average degree of satisfaction among all voters? It is given by the following (Lebesgue) integral over all voters and all states in Z.

$$\alpha = \frac{\int_{\omega \in Z} \int_{v \in V} s(v, \omega) \mathrm{d}v \mathrm{d}\omega}{\mu(Z)} \tag{1}$$

namely the average value of $s(v, \omega)$ over all voters v and all states ω in Z. [3]
Here $\int_{v \in V} s(v, \omega) \mathrm{d}v$ is the extent to which a particular state ω is liked by the average voter.
Alternately $\int_{\omega \in Z} s(v, \omega) \mathrm{d}\omega$ is the extent to which a particular voter v likes the set Z.
α is the average over all voters in V and states in Z.

[3] We assume that the measure of the set of voters is normalized to be 1.

The candidate is now wondering whether she should make a statement A or its negation $\neg A$.
At the moment we are assuming that she has no restrictions as to what she can say with popularity being her only concern, so she is free to say A or its negation. Later on we will consider *restrictions* on what she can say. Whatever she says will have the effect of changing the set Z.
Note that the set V is fixed but Z depends on what the candidate has said so far and may say in the future. [4]

[4]However, fights between Democrats and Republicans over "illegal voters" or "voter suppression" are over the precise makeup of the set V. A candidate may well seek to increase her average satisfaction by seeking to include in V some members who like her present positions or to exclude those who dislike these positions.

Let X and Y be the two disjoint subsets of Z where X is the set of states where A is true and Y the set where A is false, i.e., where $\neg A$ is true. Z is $X \cup Y$.

Then the average satisfaction on $X \cup Y$ could be rewritten as

$$\alpha = \frac{\int_{\omega \in X \cup Y} \int_{v \in V} s(v, \omega) \mathrm{d}v \mathrm{d}\omega}{\mu(X \cup Y)} \tag{2}$$

where $\mu(X \cup Y)$ is the measure of the set $X \cup Y$.

We could rewrite (2) as

$$\alpha = \frac{\beta_x + \beta_y}{\mu_x + \mu_y} \tag{3}$$

where
$\beta_x = \int_{\omega \in X} \int_{v \in V} s(v, \omega) \mathrm{d}v \mathrm{d}\omega$,
$\beta_y = \int_{\omega \in Y} \int_{v \in V} s(v, \omega) \mathrm{d}v \mathrm{d}\omega$,
μ_x is the measure of X, and
μ_y is the measure of Y.

Then either $\frac{\beta_x}{\mu_x} = \frac{\beta_y}{\mu_y} = \frac{\beta_x + \beta_y}{\mu_x + \mu_y}$, or one of $\frac{\beta_x}{\mu_x}$ and $\frac{\beta_y}{\mu_y}$ is greater than $\frac{\beta_x + \beta_y}{\mu_x + \mu_y}$. To see this, note that

$$\frac{\beta_x + \beta_y}{\mu_x + \mu_y} = \alpha \tag{4}$$

Now suppose that

$$\frac{\beta_x}{\mu_x} = \alpha_1 \leq \alpha \tag{5}$$

$$\frac{\beta_y}{\mu_y} = \alpha_2 < \alpha \tag{6}$$

Then, $\beta_x + \beta_y = \alpha * \mu_x + \alpha * \mu_y$ by (4), and $\beta_x + \beta_y = \alpha_1 * \mu_x + \alpha_2 * \mu_y$ by (5) and (6), but $\alpha_1 \leq \alpha$ and $\alpha_2 < \alpha$, so we have a contradiction. So at least one of the ratios is greater than or equal to α, her current level of satisfaction.

Say $\alpha_1 = \dfrac{\beta_x}{\mu_x}$ is greater than α. then by uttering the statement A she will move from α to α_1 and benefit overall. Some voters may dislike A but they will be outweighed by those who like it.
Thus at least one of the statements A and $\neg A$ will either benefit her (raise her level of satisfaction) or at least leave her level the same. $\Box$

Corollary: A candidate is best off being as explicit as she can.
Proof: Consider all possible theories (sets of statements) which she could utter, and let T be the one that is best for her. If T is not complete, i.e. leaves some question A open, then there is an extension of T which includes either A or its negation and is no worse than T. It follows that among her best theories there is one which is complete, i.e., as explicit as possible. $\Box$.

Ambiguity and Pessimism

Pessimistic voters are voters who assume the worst of all the states which are currently possible. Thus if Z is the current set of possible states of the world then a pessimistic voter will value Z as $v(Z) = min\{s(v, \omega) : \omega \in Z\}$. It is obvious that being more explicit with pessimistic voters can only help a candidate since it might well eliminate states which the voter dislikes and at worst it will leave things the same way.

What if the voters are a mixture of expected value voters and pessimistic voters? Then given the choice of saying A and $\neg A$, at least one of the two, say A will help her with the expected value voters and can do no harm with the pessimists. Now it could be that saying $\neg A$ will help her more with the pessimists than it hurts with the expected value voters, but at least one of A and $\neg A$ is safe for her to say.

Thus *one* possible reason for a candidate to be cagey is the presence of a large number of optimists. Optimistic voters will put $v(Z) = max\{s(v, \omega) : \omega \in Z\}$. If a voter strongly prefers ω and another strongly prefers ω' and both are optimists then the candidate may prefer to remain ambiguous between ω and ω'.
A second reason may be that a strong candidate who expects to win may refrain from committing herself on issues in order to leave freedom of action open if and when she takes office.
Yet another reason could be that the candidate does not want to say something contrary to her beliefs even if that would help her with voters.
Later, we consider another reason - stay-at-home voters - to explain why a candidate may prefer to be cagey.

The Logical Formalism of Dean and Parikh

The last section gave an abstract presentation of the scenario without saying what the states of the world were and where the satisfaction function s came from. Now we will proceed to be more explicit.

- The candidate's views are formulated in a propositional language L containing finitely many atomic propositions $At = \{P_1, ..., P_n\}$
- Propositional valuations and states are conflated. $\omega \in 2^{At}$

- Propositional valuations are defined as follows in order to make the arithmetic simpler: $\omega[i] = 1$ if $\omega \models P_i$, and $\omega[i] = -1$ if $\omega \not\models P_i$
- Voters are characterized by their preference for a set of ideal states. This is formalized via two functions p_v and x_v.
  - $p_v(i) = 1$ if v would prefer P_i to be true,
    $p_v(i) = 0$ if v is neutral about P_i,
    $p_v(i) =$ -1 if v would prefer P_i to be false.
  - $x_v(i) : At \rightarrow [0, 1]$ the weight which voter v assigns to P_i such that $\sum x_v(i) \leq 1$
- The utility of a state ω for voter v is defined as

$$s(v, \omega) = \sum_{1 \leq i \leq n} p_v(i) \times x_v(i) \times \omega[i]$$

- We will first consider expected value voters. Their utility for a given (current) theory T of a candidate is calculated as follows:

$$ut_v(T) = \frac{\sum_{\omega \models T} s(v, \omega)}{|\{\omega : \omega \models T\}|}$$

- The value of a statement A for a given theory T is

$$val(A, T) = ut(T + A) - ut(T)$$

  where $ut(T + A)$ is the utility when a candidate's theory T is updated by statement A.

- In Dean and Parikh, a candidate chooses what to say next by calculating the *best statement* for a given theory T (this is what she has said so far) as follows:

$$best(T, X) = argmax_A val(A, T) : A \in X$$

  where X is the set of formulas from which the candidate chooses what to say.
  X could be defined differently depending on the type of the candidate. (i.e. depending on the restrictions that are imposed on the candidate as to what she can say.)

Extension of the Framework

Dean and Parikh showed that (given expected value voters) it is always to a candidate's benefit to say something on a certain issue than to remain silent and we gave a generalization of their result in Section 2. We find that this is not the case in practical cases. Here is a well known quote from the satirical newspaper *The Onion*.

NEW YORK After Sen. Barack Obama's comments last week about what he typically eats for dinner were criticized by Sen. Hillary Clinton as being offensive to both herself and the American voters, the number of acceptable phrases presidential candidates can now say are officially down to four. At the beginning of 2007 there were 38 things candidates could mention in public that wouldn't be considered damaging to their campaigns, but now they are mostly limited to 'Thank you all for coming,' and 'God bless America,' ABC News chief Washington correspondent George Stephanopoulos said on Sunday's episode of This Week.

The Onion, *May 8, 2008*

For a more scholarly source consider [5]

> *Modern U.S. candidates have proven just as willing to use ambiguity as a campaign strategy. Jimmy Carter and George H.W. Bush were renowned for taking fuzzy positions at crucial points during their successful runs for the presidency (Bartels 1988, 101), and Barack Obama captured the White House in 2008 while remaining vague on key issues.*

Candidate Beliefs

We extend the framework of Dean and Parikh by considering a belief set B for the candidate to represent her honest beliefs (this set need not be complete, allowing the candidate to form her opinions on certain issues later on), and a theory T to represent the statements she has uttered so far.

We will impose the requirement that the candidate will not make statements that are against her beliefs. This requires B and T to be consistent with each other, and every possible statement A for this candidate to be consistent with $B \cup T$. We could call this type of candidate *tactically honest*. In this case it is easy to see that sometimes such a candidate might want to remain silent.

Remaining Silent - An example

We consider a single candidate c whose belief set B_c and theory T_c are given below:
$B_c = \{\neg P, Q, \neg R\}$ $T_c = \{\neg R\}$
There are four assignments satisfying T_c that the voters will take into account. The assignments are given in the form of $\langle p_v(P), p_v(Q), p_v(R)\rangle$
$\omega_1 = \langle -1, -1, -1\rangle$, $\omega_2 = \langle -1, 1, -1\rangle$, $\omega_3 = \langle 1, -1, -1\rangle$, and $\omega_4 = \langle 1, 1, -1\rangle$

We will consider two groups of voters v_1 and v_2 of the same size:
v_1's preferences:
$p_1(P) = 1$, $p_1(Q) = 1$, $p_1(R) = -1$
$x_1(P) = 0.4$, $x_1(Q) = 0.2$, $x_1(R) = 0.2$
v_2's preferences:
$p_2(P) = 1$, $p_2(Q) = -1$, $p_2(R) = -1$
$x_2(P) = 0.3$, $x_2(Q) = 0.1$, $x_2(R) = 0.4$
What should c say about P?

In this initial situation, v_1 has 0.2 points for c and v_2 has 0.4 points. $v_1(T_c) = 0.2$ and $v_2(T_c) = 0.4$. c would like to say $\neg P$ since it is in her belief set.
However the popular opinion among the voters is P, so if she were to say $\neg P$ her points would go down among both groups of voters. $v_1(T_c + \neg P) = -0.2$ and $v_2(T_c + \neg P) = 0.1$.
She also cannot say P as it contradicts her opinions i.e. her belief set. So she might choose to remain silent in regard to P.

Revealing Partial Truths - An Example

In Example 1 we showed that remaining silent about P would be a reasonable option for the candidate if she doesn't want to lie about her honest beliefs. Another option is that she could say $(P \vee Q)$ which allows her not to directly contradict her belief set, but also to increase her points at the same time by revealing only a partial truth. In this case the voters will remove the possible state $\omega_1 = \langle -1, -1, -1 \rangle$ from their calculations, and her points will go up among both groups of voters. $v_1(T_c + (P \vee Q)) = 0.4$ and $v_2(T_c + (P \vee Q)) = 0.47$.

This example shows that a candidate who does not want her statements and her beliefs to conflict can achieve more by remaining silent on a certain issue than by voicing her opinion. However she may occasionally achieve even more by making a vague statement (i.e. by revealing partial truth (and not lying)) than by remaining silent.

Dishonest Candidates

If we allow a candidate to be dishonest (i.e. to make statements that can contradict her belief set), then we can ignore the candidate's belief set altogether. This candidate will choose to be as explicit as possible given that the voters are expected-value voters as per Dean and Parikh's result. It is easily seen then the candidate will have more leeway against the candidates who choose to be honest. Moreover, even a candidate who is honest, but has some leeway is better off by being less explicit at the start of a campaign.

Let T_1 and T_2 be two positions to which she could commit at the start of the campaign. Assume moreover that $T_1 \subset T_2$. Then the number of permissible extensions of T_2 is less than the number of permissible extensions of T_1. So the best she could gain by starting only with T_1 is higher than the best she could gain by starting with T_2.

One could ask, but if T_2 was something she was going to say anyway, then why not say it at the start of the campaign? The answer is that the function $s(v, \omega)$, the extent to which voter v likes the state ω is not constant and may vary as time passes. The tastes and preferences of voters do change. So it may be wiser to wait. But when further changes in s are unlikely then being explicit can be helpful.

Stay-At-Home Voters

For another explanation as to why a candidate might want to remain silent, we will introduce a threshold value t for voters. We will stipulate that a voter's utility for a candidate c must be greater than t for that voter to vote for c.

Consider a candidate c with the following belief set and theory.

$B_c = \{P, \neg Q, \neg R\}$

$T_c = \{(P \vee Q), \neg R\}$

Assume two groups of voters v_1 and v_2 of the same size with the following preferences.

v_1's preferences:

$p_1(P) = 1$, $p_1(Q) = -1$, $p_1(R) = -1$

$x_1(P) = 0.4$, $x_1(Q) = 0.1$, $x_1(R) = 0.4$

v_2's preferences:

$p_2(P) = -1$, $p_2(Q) = 1$, $p_2(R) = -1$

$x_2(P) = 0.1$, $x_2(Q) = 0.4$, $x_2(R) = 0.5$

If we assume a threshold value of $t = 0.5$, in this initial situation both groups of voters would vote for c as shown in the left column. If c wanted to say P, which is in her belief set, the sum of her points would go up (from 1.1 to 1.2). However in this case v_2 becomes a stay-at-home voter and c is actually hurt by saying P. Considering this, she might want to remain silent regarding P.

	Initial scores	Updated scores if c says P
v_1	0.5	0.8
v_2	0.6	0.4

Note that even if the candidate was dishonest and wanted to say $\neg P$, this time she would lose v_1's votes as shown below.

	Initial scores	Updated scores if c says $\neg P$
v_1	0.5	-0.1
v_2	0.6	1

Conclusions and Future Work

We have developed a model which explains why a candidate might wish to become explicit about issues, as well as situations where a candidate may prefer to remain silent on some issues. Our model differs in significant ways from the models developed by [1], [4], [5] in that we are including a semantics for the language in which candidates speak and raising questions about when they would speak and when they would remain quiet. See also [2] for a comprehensive discussion of elections.

- Alesina, A. F., Cukierman A.: “The politics of ambiguity.” The Quarterly Journal of Economics, 105 (4) 829-850 (1987)
- Brams, S.: *The Presidential Election Game*, New Haven, CT: Yale University Press, 1978. Rev. ed., 2008 (A K Peters).
- Dean, W., Parikh R.: The Logic of Campaigning, in M. Banerjee and A. Seth (Eds.): *ICLA 2011*, LNAI 6521, pp. 38–49. Springer, Heidelberg (2011)
- Shepsle, K. A.: “The strategy of ambiguity: Uncertainty and electoral competition.” American Political Science Review 66.02 555-568 (1972)
- Tomz, M., Van Houweling R. P.: “The electoral implications of candidate ambiguity.” American Political Science Review 103.01 83-98 (2009)
- Carlos Alchourron, Peter Gärdenfors, David Makinson, “On the Logic of Theory Change: Partial Meet Contraction and Revision Functions” *The Journal of Symbolic Logic*, Vol. 50, No. 2 (Jun., 1985), pp. 510-530

- R. Aumann, "Agreeing to disagree", *Annals of Statistics*, **4** (1976) 1236-1239.
- Steven Brams, *The Presidential Election Game*, Yale University Press, 2007.
- Brandenburger, Adam , The Language of Game Theory, World Scientific, 2014.
- Steven Brams and Alan Taylor, *The Win-Win Solution: guaranteeing fair shares to everybody*, Norton 1999.
- Fagin, R., J. Halpern, Y. Moses and M. Vardi, *Reasoning about Knowledge*, MIT Press 1995.
- J. Geanakoplos and H. Polemarchakis, "We can't disagree forever," *J. Economic Theory*.
- Halpern, J., and M. Vardi, The complexity of reasoning about knowledge and time, in *J. Computer and System Sciences*, **38** 195-237, 1989.

- Jaakko Hintikka, *Knowledge and Belief: an introduction to the logic of the two notions*, Cornell University press, 1962.
- D. Lewis, *Convention, a Philosophical Study*, Harvard U. Press, 1969.
- Pearce, D. (1984) Rationalizable Strategic Behavior and the Problem of Perfection. *Econometrica* 52: 1029-1050.
- Parikh, Rohit, and Ramaswamy Ramanujam. "A knowledge based semantics of messages." Journal of Logic, Language and Information 12.4 (2003): 453-467.
- Parikh, Rohit, Çağıl Taşdemir, and Andreas Witzel. "The power of knowledge in games." International Game Theory Review 15.04 (2013): 1340030.
- Plaza, Jan. "Logics of public communications." Synthese 158.2 (2007): 165-179.
- Premack, David, and Guy Woodruff. "Does the chimpanzee have a theory of mind?." Behavioral and brain sciences 1.04 (1978): 515-526.

Rubinstein, Ariel, Lecture notes in microeconomic theory: the economic agent. Princeton University Press, 2012.

Savage, Leonard J. The foundations of statistics. Courier Corporation, 1972.