Interests and Evidence

Brian Weatherson

June, 2017

University of Michigan, Ann Arbor and Arché, University of St Andrews

Encroachment, Reduction and Explanation

Red-Blue Game

Rules of the game:

1. Two sentences will be written on the board, one in red, one in blue.
2. You get two choices.
3. First, you pick a colour, red or blue.
4. Second, you say whether the sentence in that colour is true or false.
5. If you are right, you win. If not, you lose.
6. Let's imagine that if you win, you get $50, and if you lose, you get nothing.

Assume that you know the rules of the game, and nothing else relevant.

Red-Blue Game

An instance of the Red-Blue Game

- Two plus two equals four.
- *Knowledge and Lotteries* was published before *Knowledge and Practical Interests*.

Intuitions:

- In ordinary circumstances, I know the blue sentence is true.
- The only rational play for me is Red-True.

Explaining the Intuitions

- Pragmatic encroachment theories can easily explain this; I lose knowledge about publication dates when playing the game.
- Non-pragmatic theories have a harder time explaining it.

The Super-Knowledge Hypothesis

- Pragmatic encroachment cases are cases where knowledge is preserved, but knowledge doesn't suffice for action.
- These are cases where super-knowledge is required.

Three Objections

1. If super-knowledge is required, then not clear why Red-True is so well motivated.

Three Objections

1. If super-knowledge is required, then not clear why Red-True is so well motivated.
2. Not clear player has any different attitude towards *Red-True will win $50* and *Blue-True will win $50*, since player doesn't super-know the rules of the game.

Three Objections

1. If super-knowledge is required, then not clear why Red-True is so well motivated.
2. Not clear player has any different attitude towards *Red-True will win $50* and *Blue-True will win $50*, since player doesn't super-know the rules of the game.
3. This ends up treating like cases not alike.

The Red-Purple-Green game

A sentence is written in Red on the board. Player has three options.

1. Say Red, then say the truth value of the sentence, and get $50 if correct, nothing otherwise.
2. Say Purple, then get $50.
3. Say Green, then get nothing.

The sentence is 2+2=4. Player knows what the sentence is, the rules, and nothing else.

Comparing the Games

- It is always OK to play Purple in the Red-Purple-Green game.
- If player knows the blue sentence is true, then an instance of the Red-Blue game just is an instance of the Red-Purple-Green game.
- So in any Red-Blue game where it isn’t OK to play Blue-True, the player doesn’t know the blue sentence is true.

Comparing the Games

- It is always OK to play Purple in the Red-Purple-Green game.
- If player knows the blue sentence is true, then an instance of the Red-Blue game just is an instance of the Red-Purple-Green game.
- So in any Red-Blue game where it isn’t OK to play Blue-True, the player doesn’t know the blue sentence is true.
- This comparative argument doesn’t make any knowledge-action links.

Pragmatic Encroachment, or Scepticism

- Could just deny I know the claim about publication dates.
- But the game generates really widely.
- Any blue sentence you couldn’t bet on, when placed alongside 2+2=4, you don’t know.
- That’s a really sceptical conclusion.

Odds not Stakes

- This is a low stakes situation - it’s just $50.
- But it is a long odds bet.
- More precisely, Blue-True is rational only if it is at least as probable that Blue is true as that 2+2=4.
- And that probability claim isn’t very plausible.

The Conditional Principle

I endorse these principles as constraints on knowledge:

- If the agent knows that *p*, then for any question they have an interest in, the answer to that question is identical to the answer to that question conditional on *p*.
- When an agent is considering the choice between two options, the question of which option has a higher expected utility given their evidence is a question they have an interest in.

Reduction and Explanation

- Those principles are meant to not just be extensionally adequate.
- They are meant to explain why agents lose knowledge when considering some sets of options, like in the Red-Blue game.
- In some sense, they are meant to be part of reductive explanations.

Inputs to the Explanation

These reductive explanations take as primitive inputs

- Evidential Probability
- Evidence

I’m not going to worry about evidential probability here, but I am going to worry a lot about evidence.

The Problems with Evidence

The Red-Blue Game and Evidence

Consider a version of the game where

- The red sentence is two plus two equals four.
- The blue sentence is something that, if known, would be part of the agent's evidence.

Hypothesis:

- We can get situations where the only rational play is Red-True, but in ordinary circumstances, the agent would know the blue sentence is true, and it would be part of their evidence.

An Example

- I see someone, call them Rahul, across the room in a restaurant in Ann Arbor.
- Rahul is someone I know well, and can recognise, but I had no idea he was in town.
- Still, the ordinary situation is that I know Rahul is here.
- Indeed, the ordinary situation is that Rahul being in this restaurant is part of my evidence.

Now play a version of the game with:

- Two plus two equals four.
- Rahul is in this restaurant.

The Challenge

- This doesn't threaten the extensional adequacy of the conditional principle.
- This set of views is consistent: E=K, and I don't know Rahul is here, so it's not part of my evidence that Rahul is here, so the evidential probability of Rahul being in Ann Arbor is not high enough to choose Blue.
- But this explanation is not a reductive explanation of why I don't know Rahul is here.
- It reasons from the lack to knowledge to the lack of evidence, and I want an explanation that goes the other way around.

Some Ways Out

1. Insist that evidence is only ever phenomenological, and the red-blue game never defeats phenomenological knowledge.
2. Give up on the project of providing reductive explanations for why changing practical circumstances lead to loss of knowledge.

Neither seems particularly plausible.

Multiple Solutions

One cost of the explanation being non-reductive is that the following position is also consistent:

- E=K
- Agents loses knowledge that *p* when the evidential probability of *p* is not close enough to one.
- Since *p* is part of my evidence, its evidential probability is 1, so it is close enough to 1.
- So there is no threat from pragmatic encroachment to knowledge here.

A non-reductive account of when pragmatic effects matter is, in this case, a non-predictive account.

A Solution

The Old Solution in Symbols

Let K be the agent's evidence, and A, B be relevant choices, E the expected value function, and p something the agent may or may not know. Then the agent knows p only if:

$$E(A|K) \geq E(B|K) \leftrightarrow E(A|K \wedge p) \geq E(B|K \wedge p)$$

Towards a New Solution

- The problem is that we don't know whether K includes p or not.
- So let K now include only the uncontroversial part of the agent's evidence. So possibly the evidence is $K \wedge p$, possibly it is K.

Split the Difference

For any action *X*, define a new function *V* as follows.

$$V(X) = \frac{E(X|K)+E(X|K\wedge p)}{2}$$

It's the average of the expected values of *X* with and without *p* in the evidence.

The New Pragmatic Constraint

If *p* is something that might be known, and is part of evidence if known, then the pragmatic constraint is that for any relevant *A*, *B*:

$$V(A) \geq V(B) \leftrightarrow E(A|p) \geq E(B|p)$$

If *A* beats *B* given *p*, then *A* must also do better than *B* on this ‘split the difference’ criteria.

Good Features of Solution

- It gets the obvious cases (like Rahul in the restaurant), right.
- It does not presuppose that we know what evidence the agent has in order to apply the rule.

Bad Features of Solution

- It isn’t obvious how it is going to generalise to cases where there are multiple propositions that might or might not be part of evidence.

Bad Features of Solution

- It isn’t obvious how it is going to generalise to cases where there are multiple propositions that might or might not be part of evidence.
- It looks completely ad hoc.

Gamifying the Problem

Newcomb’s Problem as a Game

- It is interesting to think of some philosophical problems as games, especially when they involve interactions of rational agents.
- Here, for example, is the game table for Newcomb’s problem, with the familiar human as Row, and the demon as Column.

	Predict 1 Box	Predict 2 Boxes
Choose 1 Box	1000, 1	0,0
Choose 2 Boxes	1001, 0	1, 1

Newcomb’s Problem as a Game

- It is interesting to think of some philosophical problems as games, especially when they involve interactions of rational agents.
- Here, for example, is the game table for Newcomb’s problem, with the familiar human as Row, and the demon as Column.

	Predict 1 Box	Predict 2 Boxes
Choose 1 Box	1000, 1	0,0
Choose 2 Boxes	1001, 0	1, 1

Note that the unique Nash equilibrium of the game is the bottom right corner.

The Interpretation Game

There are two players:

1. Human
2. The Radical Interpreter

Here are their goals:

- The Radical Interpreter assigns mental states (including evidence) to human in such a way as to correctly predict human's actions (assuming human is rational).
- Human acts so as to maximise evidential expected utility, where the evidence is what the radical interpreter says the evidence is.

A Version of the Game

- Human faces a choice between taking and declining a bet on p.
- If bet wins, it wins 1 util, if it loses, it loses 100 utils.
- p is like the claim that Rahul is in the restaurant; it is unclear whether it is in human's evidence.
- If K is the rest of human's evidence, then $\Pr(p|K) = 0.9$.
- The Radical Interpreter has to choose whether p is part of the evidence or not.
- Human has to decide whether to take the bet or not.
- The Radical Interpreter gets what they want if human takes the bet iff p is part of their evidence.

Table for the Game

	$p \in E$	$p \notin E$
Take the bet	1, 1	-9.1, 0
Decline the bet	0, 0	0, 1

- Since the bet is rational iff p is part of evidence, The Radical Interpreter wins in the top-left and lower-right quadrants, and loses otherwise.
- In the bottom row, human gets a payout of 0, since the bet is declined.
- In the top-right, the bet is a sure winner, so it's expected return is 1.
- In the top-left, bet wins with probability 0.9, so its expected payout is −9.1.

Equilibria of the Game

This is a coordination game, and like most coordination games, it has multiple Nash equilibria.

	$p \in E$	$p \notin E$
Take the bet	**1, 1**	-9.1, 0
Decline the bet	0, 0	**0, 1**

That corresponds to the conditional principle not setting a unique solution to what the agent's evidence/knowledge is.

Solving Coordination Games

Stag Hunt

	a	b
A	5, 5	0, 4
B	4, 0	2, 2

- This game has two equilibria, Aa and Bb.
- Let’s talk about the choice between them.

Pareto-Dominant

- The *Aa* equilibrium is better for both players than the *Bb* equilibrium.
- That is, it is **Pareto-dominant**.
- Some theorists think we should select Pareto-dominant equilibria, when they are available.

Risk-Dominant

- Each player does best playing *Bb* if they think it is 50/50 which equilibrium strategy the other player will play.
- That is (simplifying a little), the *Bb* strategy is **risk-dominant**.
- Some other theorists think we should select risk-dominant equilibria, when they are available.

Looking Ahead

Here’s the quick version of the rest of the paper.

1. In the game between Human and The Radical Interpreter, there is a Pareto-dominant and a Risk-dominant equilibria.
2. Risk dominance is a better equilibrium choice rule than Pareto-dominance.
3. The risk-dominant equilibria is the low evidence equilibria.
4. So that’s what The Radical Interpreter will choose.
5. So in that case Human does not have *p* in their evidence.

The Two Equilibria

	$p \in E$	$p \notin E$
Take the bet	1, 1	-9.1, 0
Decline the bet	0, 0	0, 1

- There is literally nothing to choose between them for The Radical Interpreter.
- Human would prefer the top-left equilibrium.
- But it is very risky; the lower-right equilibrium is safer.

Why Prefer Risk-Dominance

- It does better in games where people don’t know exactly what the payoffs are.
- In epistemic games, the Human (at least) doesn’t know exactly what the payoffs are.

An Example

Assume that Row and Column are playing a version of this game, with for now unknown x.

	a	b
A	4, 4	0, x
B	x, 0	x, x

- x will be chosen at random from $[-1, 5]$.
- Column will be told the value for x
- Row will be told x with a small error margin, chosen at random from $[-\varepsilon, \varepsilon]$.

Solving this Global Game

- There is only one solution (more or less) to this game.
- Both players play *B* if they get a 'signal' of more than 2.
- Both players play *A* if they get a 'signal' of less than 2.
- Strictly speaking, we don't rule out a whole bunch of options for what to do when the signal is 2.
- We only need to use iterated deletion of strongly dominated strategies to solve this game.

Generalising

- Any coordination game can be easily embedded in a global game like this.
- One of the payoffs is replaced with an unknown variable, and the player gets a noisy signal of the payoff.
- In general, the solution to the most natural form of the global game is to play the risk-dominant equilibria.
- And that's a realistic setting for what Human faces.

Odds and Ends

Progress Made

- We have a thing to say about cases like Rahul in the restaurant.
- It doesn’t rely on figuring out antecedently what agent’s evidence is.
- It is consistent with the story that agents lose knowledge that *p* when they can’t conditionalise on *p*.

Downsides

- Much more complicated.
- It turns out to make a difference whether p is possible evidence or not.