Interests and Evidence

Brian Weatherson

February 26, 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 is Red-True.

Knowledge Norm of Action

- 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.
- Saying that knowledge isn't sufficient for action won't help here; then it isn't clear why knowledge of the rules licences Red-True as a rational play.
- Saying I never knew the fact about publication dates won't help; the example structure generalises widely.

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.

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.

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. Radical Interpreter

Here are their goals:

- 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 $\mathbf{Pr}(p|K) = 0.9$.
- 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.
- 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, 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

There are two Nash equilibria for the game - I've bolded them below.

	$p \in E$	$p \notin E$
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.

First Attempt

- Rational play in the Interpretation Game involves playing one part of a Nash equilibrium strategy.
- Any play, by either human or radical interpreter, is potentially part of a Nash equilibrium strategy.
- So it is indeterminate whether *p* is part of agent's evidence or not.
- That's not a terrible solution - it is reductive and interest-relative, but let's see if we can do better.

A Better Approach

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.

Dominance Principles

I think risk-dominance is a more sensible equilibrium choice rule, since Pareto-dominance can lead to selecting weakly dominated strategies, as here.

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

Solving the Interpretation Game

- The equilibria are equally preferred by the radical interpreter.
- But if human is making a choice while being 50/50 on what the radical interpreter will pick, they maximise expected utility by declining the bet.
- So the equilibrium: Decline bet, $p \notin E$ is risk-dominant.

My Theory

Evidence human has is the evidence radical interpreter says they have, assuming both players choose a risk-dominant equilibrium.

- This theory is reductive; it doesn't presuppose what human's evidence is before we say what they know.
- And it is interest-relative.

Problems with the Theory

Problem One: Too Complex

The conditional principle had two theoretical motivations:

1. Always maximise evidential expected utility
2. Conditionalizing on what you know doesn't change any relevant question

No game-theoretic story is going to be as plausible, or as simple.

Problem Two: Possibility of Evidence Matters

Change one variable in the interpretation game:

- Now the downside to losing the bet is 15 utils, not 100.
- That gives us the following game table, whose risk-dominant equilibria is the top left corner.

	$p \in E$	$p \notin E$
Take the bet	1, 1	-0.6, 0
Decline the bet	0, 0	0, 1

Problem Two: Possibility of Evidence Matters

- This is odd.
- It isn't odd because the stakes matter; that's just interest-relativity.
- It is odd because normally, if $\Pr(p) = 0.9$, and the agent was facing a bet on p at 15–1 odds, we'd say their practical interests block knowledge that p.
- It turns out that if p is evidence if known, the probabilistic threshold for it being known is higher than if it is certainly not evidence.
- That's surprising, though I'm not sure if there are clear intuitions here.

Conclusion

Summary

1. Evidence is what a radical interpreter, who thought you were rational and wanted to predict your behaviour, would think it is.
2. This will be interest-sensitive in any number of unclear cases.
3. How the radical interpreter solves this problem depends on much more contentious aspects of game theory than orthodox utility theory, but any viable solution will be interest-relative in some sense.