NETWORKS

444 Lecture 25

Brian Weatherson

2024-04-11

EXAM

EXAM PREP

We’ve posted a superset of the exam questions on Canvas as `sample_exam.pdf`.

This is *longer* than what the exam will be; the real exam will be 6 questions.

The numbers will change in the numerical example.

We’ll go over some of the questions next Tuesday, but I wanted to talk about one of them now.

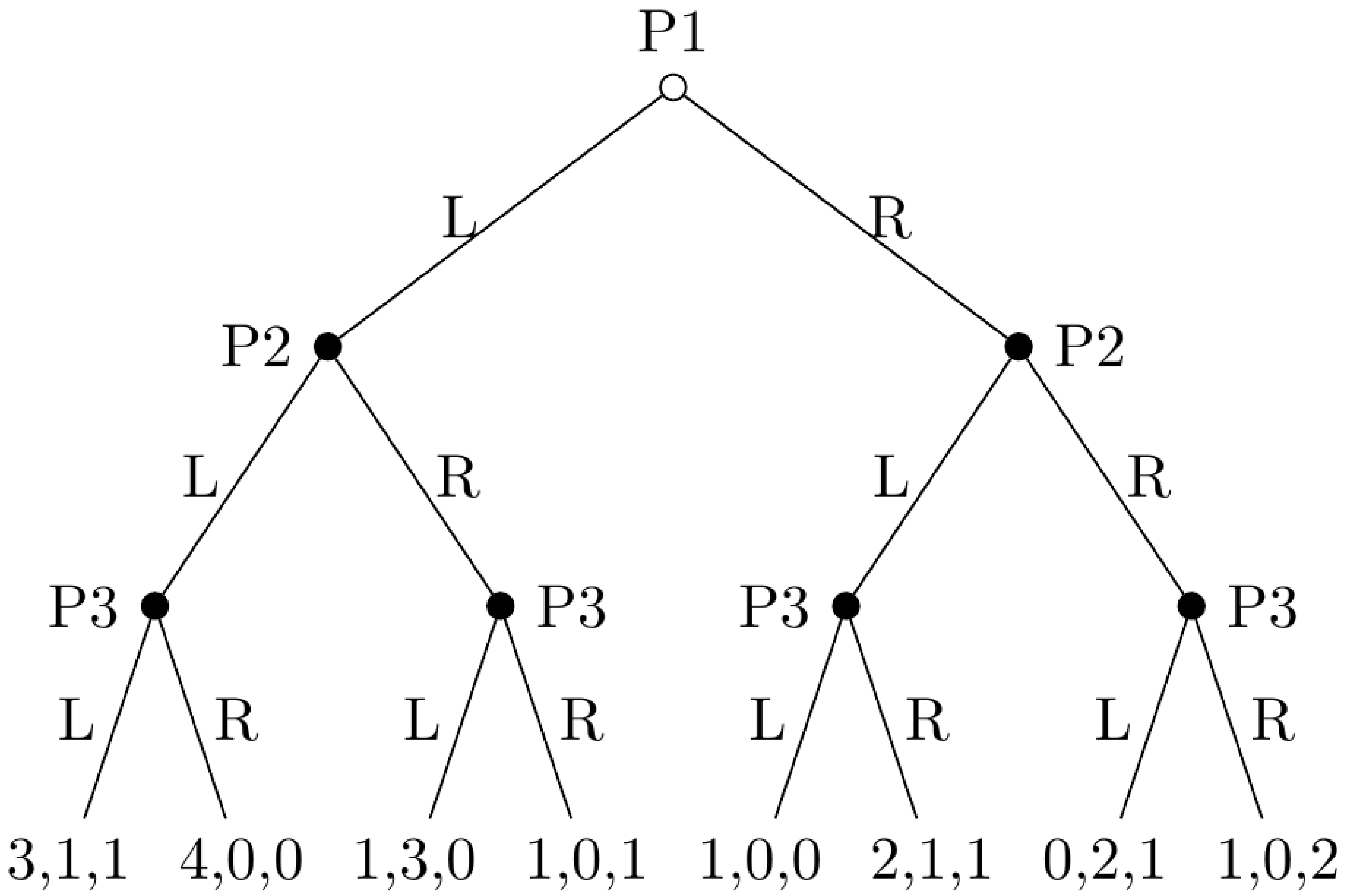

Figure 1: Tree for Q8

TESTIMONY

BACKGROUND

Two views on testimony.

1. Testimony is a basic way we know about the world, like perception, memory, inference.
2. Testimony works because people are useful signals about the world, like tree rings, and we have background evidence that they are in fact useful signals.

BACKGROUND

There are in between views, lots of possible nuance, etc, etc, but it’s very helpful to start with these two.

This is oddly not a widely discussed topic in ‘Western’ epistemology until very recently, but it’s an absolutely foundational question in Indian epistemology.

PUZZLES

1. The ‘basic way of knowing’ approach seems to be a recipe for gullibility.
2. The ‘signals’ way seems to require a massive amount of **work**, keeping track of all sorts of background. Every conversational exchange becomes something like a scientific investigation.

VIGILANCE

One interesting recent version of the ‘signals’ approach is due to Hugo Mercier and colleagues.

On their model, humans are by nature **vigilant**. Everything goes through a two-step process.

1. Are there any red flags here? If not, accept.
2. If so, do something like the scientific investigation.

VIGILANCE

This seems (to me at least) promising both as a model for how things should go, and how things do in fact go.

And it might explain the challenges with various computer systems.

E.g., one problem with self-driving cars is that they don't know where to look, and so have to do something like a full analysis of all 360 degrees of their periphery.

VIGILANCE

But it suggests two problems in worlds like ours.

1. A lot of the ‘red flags’ we are innately disposed to check for (and there is some evidence that very young children do this, so it may be innate) are optimised for worlds where communication is one-on-one.
2. If disagreeing on distinct issues is a ‘red flag’, that might contribute to polarization.

POLARIZATION

We're not going to get into this, but as you might have seen, there is a bit of work on models for how polarization develops.

The big challenge here is that polarization can easily seem very rational.

- It's good to believe people who are reliable and not people who are unreliable.
- We don't have any way of telling who's reliable other than who agrees with us.

NETWORK MODELS

KINDS OF INTERACTION

- Small group
- Announcement
- Overlapping neighbors

ONE ON ONE

You mostly talk to the same people.

They mostly talk to the same people as you talk to.

You have lots of background on how reliable they are.

And you all talk about who was and wasn’t reliable, so everyone has a long run interest in telling the truth. The reputation effects push strongly towards honestly, with a little strategic lying mixed in.

ANNOUNCEMENT

Some people get to make announcements that everyone hears (and everyone hears that everyone hears etc), but you don't get to talk to them in the same way.

This breaks some of the things our ancestors learned about good practice with testifiers.

Can't ask follow up questions on oddities in the announcement, and often can't use non-verbal cues to support/undermine confidence.

OVERLAPPING NEIGHBORS

This gets to be a more interesting model.

Imagine that everyone is arranged in a circle, and that everyone only talks to people alongside them.

Then there are very different dynamics in the group.

You're probably familiar with one real world effect of this - information gets lost as it moves down the chain.

HUB AND SPOKE

Another kind of model, one we'll talk about a bit (especially on Thursday) is a hub and spoke model.

Everyone except one person is arranged in a circle, but one person is in the middle.

Everyone can talk to their neighbors, and to the person in the middle.

Different policies the person in the middle might have lead to different results.

CASCADES

1. I have a jar, some colored balls, and a coin. I flip the coin, and don’t show it to you.
2. If heads, I put 7 red balls and 4 blue balls in the jar; if tails, I put in 6 blue and 5 red. (Again, this is hidden, but you know my policy.)
3. I then invite each of you to come up one at a time.
4. You can draw one ball from the urn, look at it, not show it to anyone or tell anyone about it, then return the ball to the urn.

5. Then I’ll draw a ball.
6. Before that, you can bet on whether I draw red or blue. If you’re right, you get a prize.
7. Your *bet* is public information, but not whether it won.

THEORY

This can easily lead to a **cascade**.

The first person should bet on whatever color they see. (This isn't obvious, but we can do the math if people are interested.) And everyone knows this.

If the first person sees a red ball, everyone should bet on red no matter what they see.

If the first person sees blue, the second person should bet on whatever color they see.

If the first two bet on blue, everyone should bet on blue from then on.

REALISM?

Does this really work?

Maybe. What it’s a model of is what happens when people (a) have some private information, (b) have more public information, (c) cannot share their private information, but (d) must act publicly on the basis of their information, and (e) are perfectly rational.

In those cases, even if the information we collectively have would show us what the correct thing to do is, we might all do something less than fully useful.

In reality, people will probably over-weight their private information.

ZOLLMAN EFFECT

ZOLLMAN EFFECT

Kevin Zollman (Carnegie Mellon)

Sometimes sharing all information can lead to us getting stuck in bad equilibria.

SETUP

Imagine there’s a new drug, and as usual we’ll test it by giving it to some random patients.

Assume the drug is *on average* better than baseline.

But it’s not 100% reliable, so it gets an ‘unlucky’ group of initial test subjects, and looks bad.

NEXT STEP

Everyone sees this, we're assuming all information is public, and everyone concludes that the drug probably doesn't work.

People aren't stupid, they don't think this is the final word.

But, this is the crucial bit, experiments are expensive, both in money and potentially in health.

So no one double checks.

SOLUTIONS

You don't want *no* communication; then there is no point to doing experiments.

Apparently what works best in these models is having a little *friction*.

In real life we sort of get this. Usually a meh result about a new drug/treatment does not get as widely announced as positive findings, so in real life cases we should often avoid this trap.

THREE QUESTIONS

1. Is this just an artifact of some model? Apparently not; it seems to be a fairly resilient effect across different kinds of models.
2. How big is the effect? My sense is that it isn’t huge. You lose a few drugs that are a few percent better than baseline.
3. Are there real world instances of this? Good question!

FOR NEXT TIME

We’ll look at how playing around with the hub of hub and spoke models can make it possible to mislead people even while only ever telling them the truth.