Polarization and Rationality

Kevin Dorst
kevindorst@pitt.edu

Yale Inference Colloquium Series
May 26, 2021

I. Summary

Three models of polarization:

Trust Dynamics:

- Agents decide whether to test a new option (B) or go with a known option (A), based on how confident they are that B is better.
- They learn from their neighbors—*but* they trust their neighbors reports less as their credences diverge more.
- Upshot: when trust diminishes rapidly enough, one group ends up somewhat confident that A is better, while another group ends up certain that B is better.

E.g. treatments. Antibiotics for chronic Lyme? O'Connor and Weatherall 2019.

Conformity:

- Like before, but now instead of mistrusting others, agents feel some pressure to conform to their neighbors
- When the network has a clique structure, this can prevent information about the superiority of B from breaking into A-performing (and often A-believing) cliques.

Reporting practices and Confirmation Bias:

- Model different journalistic methods for selecting what news to report: hyperbole, extreme reporting, and fair reporting.
- Agents have confirmation bias: more likely to ignore report the further it is from what they expect.
- For bi-modal reporting (extreme or fair reporting), this leads to polarization.

I.e. the less likely they previously thought it to be.

Upshot: we shouldn't see these models as in competition. Rather, each identifies a mechanism that plausibly contributes to polarization. Frustratingly, addressing one may exacerbate another.

More links? Helps with conformity; can hurt with trust.

⇒ A proper understanding of the range of models that predict polarization should make us wary of simple policy interventions.

II. Questions

Q: What is the role of simple computational models of (e.g.) polarization, in light of Jim's observations that:

1) there are no doubt *many* sufficient, interacting causes; and
2) trying to combine (or weigh) them into one "grand unified model" looks like an intractable problem.

Option 1: How-possibly explanations?

- Goal is to delineate the (plausible) ways polarization can arise in communities of various structures, so that we can use them to assess the promise and risks of various interventions.
- Problem: *far too many* (plausible) how-possibly explanations!

We need more constraints.

Option 2: Rational models?

- One constraint: showing how polarization can arise under even strong rationality assumptions.
- I have sympathy! But is that how we should see these models?

Perhaps conformity model?

Option 3: Psychological models?

- Psychologists identify a specific reasoning mechanism that they suspect will lead people to polarize.
- Models can be used to test whether that hunch is correct.
- Confirmation bias model? *But* then the model has to work hard to match the details of the psychological mechanism.
  Many modulators of confirmation bias, especially of the "selective-exposure" variety (Whittlestone 2017).

Sometimes, it turns out not. E.g. "positive test strategy" McKenzie (2004); Salow (2018).

Confirmation in competition with surprisingness (Frey 1986; Vosoughi et al. 2018).

Option 4: Further features of polarization?

Polarization is often:

a) *Predictable:* you can predict *which side* you'll end up on.

b) *Profound:* both sides are arbitrarily confident.

c) *Robust:* confidence remains upon learning how it came about.

E.g. consider "Capital punishment has a deterrent effect".

Even before you have an opinion.

Do models that predict *this type* of polarization? Maybe not:

- Confirmation-bias model has (a) and (b), but, I think, not (c).
- Conformity model has (c) and maybe (a), but not (b).
- Not sure whether trust model has any of (a)–(c).

References

Frey, Dieter, 1986. 'Recent Research on Selective Exposure to Information'. *Advances in Experimental Social Psychology*, 19:41–80.

McKenzie, Craig R M, 2004. 'Framing effects in inference tasksâĂŤand why they are normatively defensible'. *Memory & cognition*, 32(6):874–885.

O'Connor, Cailin and Weatherall, James Owen, 2019. *The Misinformation Age: How False Beliefs Spread Cailin O'Connor James Owen Weatherall*. Yale Press.

Salow, Bernhard, 2018. 'The Externalist's Guide to Fishing for Compliments'. *Mind*, 127(507):691–728.

Vosoughi, Soroush, Roy, Deb, and Aral, Sinan, 2018. 'The spread of true and false news online'. *Science*, 359(6380):1146–1151.

Whittlestone, Jess, 2017. 'The importance of making assumptions : why confirmation is not necessarily a bias'. (July).