

Singer et al. 2019: Polarization and Memory

Kevin Dorst

24.223 Rationality

I. Polarizing Deliberation

Key feature of polarization: arises from *group discussions*.

Group polarization effect: opinions tend to become *more homogeneous* and *more extreme* after discussions with like-minded others.

It's easy to see how this could be *practically* rational.

- "Social-comparison theory": want to signal commitment to group cause.
- Identity-based cognition: your beliefs won't affect policy, but will affect your friendships.

Q: *Can* we believe for purely practical reasons?

Singer et al.: Could group polarization be *epistemically rational*?

II. Modeling Deliberation

Pieces of the model:

- Fixed proposition $q = \text{the defendant is guilty}$.
- Things you know are *reasons for/against belief*.
- Rational: strength of belief = (summed) *weight of reasons*.
 - Build in Uniqueness Thesis: weights are same for all people.
 - Believe if $\text{sum} > 0$; disbelieve if $\text{sum} < 0$
- Generate (random) fixed pool of relevant reasons; each person starts with random reasons from that set.
- Discussion proceeds by random person sharing random reason.
- **Memory limits:** can only remember 7 reasons; when get an 8th, take it into account to decide which reason to forget.

Defendant has motive: +2
 Witness claimed defendant was with him at the time: -2
 Witness has lied in past: +1
 Etc....

In some versions of models, they also investigate individually; ignore this.

Strategies for forgetting:

- Simple-minded: forget randomly.
- Weight-minded: forget reason with smallest weight.
- Coherence-minded: forget reason of smallest weight *that tells against your belief*.

So if have $\{+2,+2,+1,+1,-2,-2,-2\}$ ($= 0$) and get new +1 reason, then simple-minded forgets randomly; weight-minded forgets +1, and coherence-minded forgets -2.

Model results:

- Both Simple- and Weight-minded converge in opinions.
- But Coherence-minded tend to polarize into two groups.

III. So What?

Singer et al. argue that coherence-minded forgetting is *epistemically rational*—and therefore that group polarization is too.

Clearly simple-minded forgetting is irrational. What about others?

Argument:

- P1** The model's agents are epistemically rational if they do the best they can to get to the truth, given their limitations.
- P2** Either weight-minded (WM) or coherence-minded (CM) forgetting is best.
- P3** Since WM-forgetting can lead them to change their beliefs on the basis of impoverished reasons (but CM cannot) the latter is better.
- C1** CM forgetting is rational for our agents.
- P4** If CM-forgetting is rational for our agents, then it is rational for real people.
- C2** Coherence-minded forgetting is a rational cause of real polarization.

E.g. limit = 4. Has {+6,+7,-6,-8} and receives new +5. Currently +4, but weight-minded forgetting will drop +5 and change to disbelieving with -1.
 → Coherence-minded agents *never change beliefs when they forget.*