

# Rational Polarization

Kevin Dorst

January 2019

## Abstract

Groups of people have a predictable tendency to *polarize*—to divide into subgroups such that, on many topics:

- (1) Individuals in the *same* subgroup tend to *converge* in opinions; and
- (2) Individuals in *different* subgroups tend to *diverge* in opinions.

This widely-confirmed empirical tendency is standardly taken to be a hallmark of human irrationality. It need not be. I'll first show that rational, predictable polarization is possible: whenever you face *ambiguous evidence*—evidence that you should be unsure how to react to—predictable polarization can be fully epistemically rational. This claim be proven in a general Bayesian framework, as well as illustrated with a simple demonstration. I'll then argue, further, that this abstract possibility plays a role in the actual polarization we observe. One core contributor to predictable polarization is *confirmation bias*: roughly, the tendency for people to favor evidence that confirms their prior beliefs. And I'll argue that—given common structures of evidential ambiguity—rational agents who care only about the truth should sometimes exhibit confirmation bias.