20. Nielson and Stuart, Rational polarization?

Kevin Dorst
kevindorst@pitt.edu

TOTAL: In ideal evidential scenarios (when evidence is clear and shared), ideally rational (Bayesian) agents expected to converge in opinions.

→ N&S claim that TOTAL is presupposed in many popular and social-scientific discussions of (e.g.) political disagreements.

→ But N&S claim that TOTAL is false: simple examples show that in any case of finite learning, polarization can be ideally rational.

And more subtle reasoning shows that even in cases of *infinite* and *complete* evidence, polarization is still possible.

I. Local Polarization

Let P be my (ideally rational) credence function and Q be yours.

Assume probabilistic, obey ratio formula, and update by conditioning.

P and Q *locally* polarize on a given proposition[1] A upon learning E iff

[1] Stats speak: "event"

$$P(A|E) < P(A) \leq Q(A) < Q(A|E)$$

This can totally happen!

Election. Abby and Bill are Democrats facing off in a primary; Christa and Dan and Republicans facing off in a primary. We know only one of each pair will win their primaries, and only one of the four will win the general election. I think Bill is the stronger Democrat; you think that Abby is. Precisely:

	a	b	c	d
P:	$1/6$	$1/4$	$1/3$	$1/4$
Q:	$1/2$	$1/12$	$1/4$	$1/6$

As a result, learning that Abby and Christa won their primaries ($\{a, c\}$) makes me lower my credence that a Democrat will win ($\{a, b\}$), and you raise your credence that a Democrat will win. Where $E = \{a, c\}$:

$P(\{a,b\}) \approx 0.42$ and $Q(\{a,b\}) \approx 0.58$, yet $P(\{a,b\}|E) \approx 0.33$ and $Q(\{a,b\}|E) \approx 0.66$.

	a	b	c	d
$P(\cdot\|E)$:	$1/3$	0	$2/3$	0
$Q(\cdot\|E)$:	$2/3$	0	$1/3$	0

In general, whether E polarizes P and Q on A depends on whether P and Q disagree on the likelihood ratios: is E more to-be-expected if A or if $\neg A$? Precisely:

Thm. if $0 < P(A) \leq Q(A) < 1$, then E polarizes P and Q iff

"The proof of this result uses only the probability axioms and algebra. We omit it, assured the reader can furnish it herself should she so desire."

$$\frac{P(E|A)}{P(E|\neg A)} < 1 < \frac{Q(E|A)}{Q(E|\neg A)}$$

Upshot: No reason to expect learning the same evidence to reduce disagreement. And since learning any finite stream of evidence is equivalent to learning a big conjunction, no reason to expect any finite stream

of evidence to reduce rational disagreement.

→ *So*, say N&S, there's little reason to think that observing societal polarization provides evidence for irrationality.

Q: Is this a good argument?

II. Global Polarization

We can measure the overall disagreement between P and Q using their **total variational distance**, i.e. the maximum degree to which they disagree about any proposition.

$$d(P,Q) = \max_{A \subseteq W} |P(A) - Q(A)|$$

Let H = the set of worlds that P assigns higher probability to than Q = $\{w : P(w) > Q(w)\}$.
Then $d(P,Q) = P(H) - Q(H)$.

Is this a good measure of overall disagreement?

- When P and Q agree on everything, $d(P,Q) = 0$.
- When P and Q disagree maximally on something, $d(P,Q) = 1$.
  In particular, the areas they assign positive credence to are disjoint.

Increasing d can also be perfectly rational.
E.g. now we agree that Abby's stronger than Bill and that Christa is stronger than Dan, but we disagree on how much stronger:

$d(P,Q) = P(\{b,c\}) - Q(\{b,c\})$
$= \frac{5}{8} - \frac{4}{12} = 7/24 \approx 0.29$,
while $d(P(\cdot|E), Q(\cdot|E))$
$= P(\{c\}) - Q(\{c\}) = \frac{1}{3} \approx 0.33$

	a	b	c	d
P:	$1/4$	$1/8$	$1/2$	$1/8$
Q:	$1/2$	$1/12$	$1/4$	$1/6$
$P(\cdot\|E)$:	$1/3$	0	$2/3$	0
$Q(\cdot\|E)$:	$2/3$	0	$1/3$	0

Upshot: Increasing *global* polarization can be fully rational, too.

Q: Is this a good argument?