# Making Causal Critiques

## Day 5 - Constructive Critiques

Jonathan Phillips

January 25, 2019

# Being Constructive

- Effective critiques are essential to learning
  - We have a scholarly obligation to point out errors in reasoning
  - We learn collectively by collaborating
  - We learn individually by thinking critically about others' work
- There is no research project that cannot be criticised

# Being Constructive

- But criticism can also be used as a weapon
  - To compete for attention/jobs
  - To discourage colleagues
  - To assert status/hierarchy/superiority
  - To destroy valuable research
  - To release our own frustrations

# Being Constructive

- To avoid these risks, we must make our criticisms **constructive**
  1. In terms of content
  2. In terms of style

# Styles of Critique

- Your task is to convince the author to *improve* their work not to *abandon* it
- So they have to:
  - Understand your comment
  - Not take it as a personal attack/become defensive
  - Have options for how to respond

# Styles of Critique

- Always remember your comment might be wrong!
  - You always know the data less well than the author
  - Recognize the inherent challenges and constraints of implementing the research
- So phrase your comment in terms of ’as I understand your argument’
- Or ’Could it be that something else is also happening?’

# Styles of Critique

- Be specific! Which part of the research design is problematic?
- Be concrete! Use an example/counterexample to communicate the risk
- Be objective! We care about the research quality, not your personal opinion
- Suggest an alternative

# Styles of Critique

- Depersonalize criticism
  - Instead of "you", refer to "in this type of research..."
  - "I feel like there might be some readers who..."
- If in doubt, use the feedback sandwich:
  - Something positive/encouraging
  - Critique
  - Something positive/encouraging

# Styles of Critique

- Finally:
  - Is the comment really necessary?
  - If it is a minor issue, is there a better way to communciate it?
  - If you have not fully understood, take time to invest in understanding it before commenting

# Strengthening Causal Arguments

1. Multiple tests
2. Multiple methods
3. Uncovering ’hidden’ units
4. Heterogeneity tests
5. Placebo tests
6. Confirming Mechanisms

# Multiple Tests

- Learning requires the interaction of theory and evidence
- Competing theories have multiple distinct implications
- We should test *all* of these implications

# Multiple Tests

- For example, Deaton argues poor health causes low economic status, but observational data cannot rule out reverse causation. Additional tests to prove his claim include:
  1. Whether the relationship falls after retirement
  2. Whether the relationship is weaker among women, who on average work fewer hours
  3. Whether the relationship holds even for diseases which could easily be cured with more income

## Multiple Methods

- Our methodologies' assumptions are often impossible to test in quantitative data
- But qualitative evidence can help justify our assumptions
- This means using multiple (mixed) methods
  - Was randomization of a field experiment successful? We can interview people about the process
  - Are their spillovers (violations of SUTVA)? We can conduct a survey and find out where people got their information from
  - Can people migrate? We can use administrative data on migration rates to assess if these differences are large enough to explain our results
  - Is a regression discontinuity threshold enforced neutrally? Or was the threshold chosen to make sure a particular unit passed?
  - Is a geographic border discontinuity 'exogenous'? We can look in the archive to know how and why it was created
- These are all "causal process observations" (Collier et al 2010)

# Multiple Methods

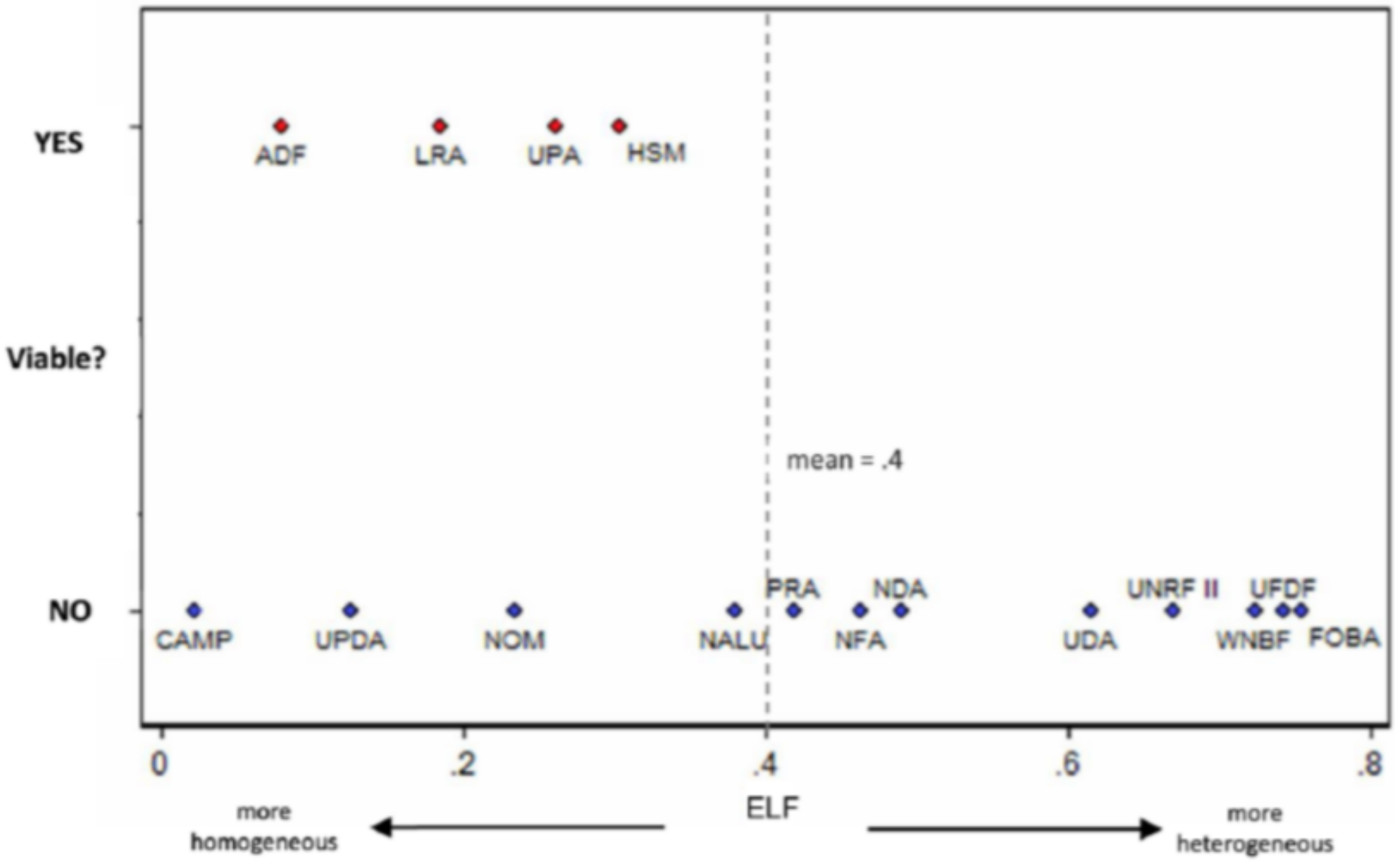
- For example, Nunn (2008) asks whether the slave trade explains underdevelopment in parts of Africa, using an instrumental variable for distance to the Americas
- Qualitative evidence helps:
  - To verify that slaves usually sailed from the nearest port, and not a different country
  - To inform the need for extra controls, eg. for legal system, natural resources
  - To identify the direction of the selection bias - reverse causation is less of a problem because he shows the richest ethnic groups were most affected by the slave trade
  - To establish the exclusion restriction for the instrumental variable: that plantations were set up in the Carribean because of the climate, not because they were near the supply of slaves in West Africa

# Uncovering ’Hidden’ Units

- If our sample size is small and good comparisons are difficult to make, one strategy is to identify more units
- Those units are often ’hidden’, either because we did not think about them or there is no data on them
- We can expand our dataset and adjust our research question
- For example, John Londregan’s talk ’uncovered’ non-trading product-country pairs to provide another source of variation to explain

# Uncovering 'Hidden' Units

- For example, Lewis (2016) wanted to ask whether ethnicity affected the formation of rebel groups, as many authors had argued
- But there is a selection/survival bias in the data - we only have data on the groups that *succeeded*
- She collected data from Uganda on *all* rebel groups
  - Expanding the sample from 1 (in most datasets) to 15
  - Showing that ethnicity does not affect rebel group formation, but may affect their success


YES
Viable?
NO
ADF
LRA
UPA
HSM
mean = .4
PRA
NDA
UNRF II
UFDF
CAMP
UPDA
NOM
NALU
NFA
UDA
WNBF
FOBA
0
.2
.4
.6
.8
ELF
more homogeneous
more heterogeneous

# Heterogeneity Tests

- Sometimes the implications of theory are very precise
- Treatment is likely to have affected subgroups to different degrees
- We can use heterogeneity tests to disaggregate the effect to each subgroup and compare

# Heterogeneity Tests

- For example, Ferraz and Finan (2008) ask how random audits affect corruption rates
- They find that audits significantly reduce corruption
- But their theory is that this is produced by 'electoral accountability'
- They provide evidence for this specific theory by:
  - Subsetting the data to only those municipalities with local radio stations which broadcast the findings of corruption and showing the effect is much *stronger*
  - Subsetting the data to only those municipalities with mayors in their first-term and face re-election incentives, showing corruption is lower
- What other theory would be consistent with *all* of these findings?

# Placebo tests

- Our theory has very precise implications, and we normally test the 'positive' version
- But we can also test the 'non-predictions' of our theory, when there should *not* be an effect
- If we found an effect where there should *not* be one, we might think something is weird in our data/methodology and have less confidence in our main result

# Placebo tests

- For example, with a regression discontinuity on close elections we expect a 'jump' effect when elections are tied (winning margin=0)
- We expect there *not* to be a 'jump' effect when winning margin=10%
  - In fact, RDDs assume continuity away from the threshold, so we need there to be no jump
- So we can apply our regression discontinuity again and see what the effect is at winning margin=10%
- If we still find an effect, there might be something wrong with our data/method

# Placebo tests

- The same with difference-in-differences
- If we were estimating the effect of a treatment that applied to some units on 5th August 2012, we expect no effect on 3rd July 2009
  - Or on 4th August 2012
  - Or on 6th August 2012
- The more tightly the data are consistent *only* with your theory, then the more credible your theory is

# Placebo tests

- Placebo tests also work for small-N studies (Glynn and Ichino 2012)
- We want to assess the effect of presidentialism on reducing party cohesion
- A good comparison is between the USA (presidential) and Canada (parliamentary)
- But we also gain confidence if we can show that other similar parliamentary systems have cohesive parties (Britain, Australia, etc.)

# Mechanisms

- Often we talk as though we are testing 'treatments'
  - But that leaves an empty black box between treatment and outcome
- Really we want to test **theories**, which include a clear logical connection between the treatment and the outcome
- To show that a specific theory is operating, we want to trace every step of the mechanism

# Mechanisms

- For example, multiple studies show a clear treatment effect: high ethnic diversity reduces public goods provision
  - But these studies had *no theory*
- Habyarimana et al (2007) asked "why?"
  - Preferences
  - Technology
  - Strategy selection
- They designed laboratory games to test exactly each mechanism
- Eg. To test if there is an ethnic 'technology' that helps co-ethnics, they asked Ugandans to find a specific person in a neighbourhood, and paid them a reward if they did
  - Co-ethnics found their target 43% of the time, non-co-ethnics only 28% of the time

# Mechanisms

- Process Tracing is one way of demonstrating which mechanism connected the treatment and the outcome

## Mechanisms

- Brady (2004) provides an example of process tracing to evaluate the plausibility of a difference-in-differences research design
- Difference-in-differences analysis suggested media announcements that Al Gore won Florida in 2000 caused 10,000 Gore voters to stay at home, allowing Bush to win.
- But:
  - There were only 10 minutes until the polling stations closed
  - Only 20% heard the announcements
  - Around half were Bush voters, who may also have stayed home
  - Voters still had a reason to vote for other offices
- Brady estimates that at most 224 people did not vote due to the media announcements