

PROTOCOLS THAT WORK

Nick Adams

Berkeley Institute for Data Science, D-Lab

What I am NOT talking about

- [TCP/IP](#), [HTTP](#), or [FTP](#). -- a set of rules or procedures for transmitting data between computers
- A set of rules governing affairs between nation-states or their representatives
- Social rules of etiquette more generally
- A full research proposal document (as used in medical experimental research)

Protocol: Definition

- ~~A practice: the process by which researchers implement any research practice subject to human error or human reproducibility~~
- **A document:** a detailed, exhaustive, and sequential list of the procedures by which researchers implement any research practice subject to human error or human reproducibility
- Ex: step-by-step instructions for doing the internet research to determine whether or not a city fits the criteria for inclusion in a study sample

Protocols *Do* Work

- Drafting them clarifies methods
- Drafting them clarifies theory
- They direct research assistants even better than you can
- They remind the future-you what you did, greatly easing the task of writing up results
- They help colleagues and critics appreciate your methods and theory

When do Soc. Scientists Need Protocols?

- Statistical analyses of secondary data
- Meta-Analyses
- Surveys, especially phone surveys
- Interviews, especially with multiple interviewees
- Ethnography
- Participant Observation
- Ethnomethodology
- Experiments
- **Content Analysis** ← our focal example today
- **Always!**
- **Especially when working with teams, assistants!**

Content Analysis is Suspect

- Biernacki's critique from *Reinventing Evidence in Social Inquiry*:
 - 1. There can be no inter-rater reliability, so we should not try.
 - 2. Researchers are not transparent enough. They lose their text classification protocols, or they are shoddy. Then, they can't explain what they did years later.
 - 3. Text classification protocols are biased in ways that are hard to review when the protocols are lost or inadequately specified.
- **I agree with 2 and 3.**

The Deciding Force Project

- ~Content Analysis
- Trying to understand dynamics of (forceful) interaction between protesters and governments – how force is chosen and how it is decisive
- Extraction of Event data (and Interaction w/in Event) data from **over 8,000** news accounts about Occupy movements in the U.S.
- 8 research assistant
- Scores/hundreds of crowd workers
- **An incredible opportunity to produce unreliable/un-reproducible research!**

The RCA Approach

- **Researchers to Crowds to Algorithms**
- Designed to scale-up content analysis projects with the help of crowd workers
- Utterly reliant on reproducible methods since we are asking crowd workers to do much of the work for us
- Starts with researcher curation of documents by TUAs
- Then, sends TUAs to crowd workers for variable/attribute extraction

TUA identification

Under cover of darkness early Tuesday, hundreds of police swept into Oakland's Occupy Wall Street protest, firing tear gas and beanbag rounds before clearing out an encampment that demonstrators had hoped would stir a revolution. In less than an hour, the two-week-old, miniature makeshift city [two-week-old, miniature makeshift city] was in ruins. Scattered across the area were overturned tents, pillows, sleeping bags, yoga mats, tarps, backpacks, food wrappers and water bottles. [Scattered across the area were overturned tents, pillows, sleeping bags, yoga mats, tarps, backpacks, food wrappers and water bottles.] Signs decrying corporations and police still hung from lampposts or lay on the ground. Protesters had stayed awake through the night, waiting for the expected raid. Officers and sheriff's deputies from across the San Francisco Bay area surrounded the plaza in front of City Hall at around 5 a.m. and closed in. Eighty-five people were arrested, mostly on suspicion of misdemeanor unlawful assembly and illegal camping [misdemeanor unlawful assembly and illegal camping], police said.

TUA Identification Hand-Coding Scheme

⊕ | Police-Initiated EVENT

⊕ **EVENT TYPE

- ⇒ Raid of camp
- ⇒ Showdown at the camp (i.e. police show up creating a tense situation that might lead to arrests)
- ⇒ Warnings or Threats at the camp or through the media (i.e. police inform campers about impending enforcement of curfew orders with intention of coercing a change in behavior)
- ⇒ Surveillance of camp
- ⇒ Surveillance of individual protest organizers away from the camp
- ⇒ Arrests of protest organizers away from any protest-initiated EVENT or CAMP
- ⇒ Undercover infiltration
- ⇒ Propaganda camping against occupy (perhaps through social media)
- ⇒ "Snatch 'n' grab" arrests at camp – arrest of individual people (especially key organizers, at the camp, even with a warrant)
- ⇒ Space-taking (i.e. establishing of perimeter barricades or fences or lines of police for containment or strategic incapacitation purposes)
- ⇒ Statement (to public and/or occupy) of REFUSAL to enforce an eviction order or permit expiration or some other law coming from city
- ⇒ **New!** Notable Non-event e.g. intentional non-enforcement – Everyone is expecting police to do one thing, then they don't do it
- ⊕ Official statement from police chief, or high-level officer or official spokesperson
- ⇒

⊕ CROWD COMPOSITION (How many...?)

- ⇒ Protesters
- ⊕ Protesters from different cities
- ⇒ Counterprotesters
- ⇒ Anarchists
- ⇒ Union Members
- ⇒ Media
- ⊕ General Indicator of Diversity of...

⊕ Police presence characteristics

- ⇒ Riot gear
- ⇒ Horses
- ⇒ Urban Assault Vehicles
- ⇒ Brandishing weapons
- ⇒ Skirmish lines

+ Protester-Initiated EVENT

+ Date/Time tag

- = START
- = END
- =

+ **EVENT TYPE

- = Establishing a camp (will often look like a rally)
- = Voluntary Dissolution of a camp – not directly resulting from a police-initiated raid
- + Moving a camp to a new location (will often look like a march)
- = March/Parade (unless protesters do rally/demonstration type stuff at the start point or end point, a march from A to B should just be counted as one event.)
- = Rally/Demonstration
- = **NEW!!** Disrupting an on-going event of the perceived 1% (e.g. conservative politician's speech at a convention/country club, city council meeting **ONLY** if no intention to negotiate or make a proposal, etc.)
- = Strike
- = Divestment action (e.g. moving money from banks to credit unions)
- + Blocking Action
- + Strategic violence
- + Strategic sabotage

+ PERMITTED?

- = From when
- = Till when
- = With what conditions

+ CROWD COMPOSITION (How many people...?)

- + Diversity of...
- = Police
- = Protesters

- 100

- Event with no indication of who initiated – e.g. “a clash between protesters and police occurred on Tuesday”
- SUMMARY INFORMATION (information that applies across multiple events in a way that makes it impossible for us to know what happened in each event without going back and looking at context from other articles).

- ▬ e.g. "172 people have been arrested over the course of the protest movement."
 - ▬ e.g. "In 11 nights of marches, 14 people have been injured."
 - ▬ e.g. "Between the city's two separate camps, at least 180 tents have been set up."
- ▬ not clear if the event really happened or was just rumored to have happened or what – needs to be reviewed in light of other articles
- ⊕ an event initiated by neither police, city, nor protesters that seems likely to have affected the outcome of the movement in some way
 - ▬ E.g. "a man committed suicide near the Occupy camp on Monday, leading protesters and police to fear for the safety of the camp..."
- ⊕ Details of such events to be coded:
 - ⊕ Attendance – numbers of...
 - ⊕ Police presence characteristics
 - ⊕ ACTIONS of police or protesters that occurred during the event
 - ⊕ Date/Time tag
 - ▬
- ⊕ **Camp Information**
 - ▬ # of people camping overnight (not MERELY attending camp establishment or move of camp event) (also color code statements about relative size e.g. "the camp is getting bigger")
 - ▬ # of tents (also color code statements about relative size e.g. "the camp is getting bigger")
 - ⊕ Permitted? NOTE: sometimes protesters will have a legal right to a space for a certain time based on a lease or a court's injunction or restraining order against the police. We will treat these as permits.
 - ▬ From when
 - ▬ Till when
 - ▬ With what conditions
 - ▬ Informally allowed
 - ▬
 - ⊕ COMPOSITION of CROWD (tags, estimate codes, and % estimate codes)
 - ▬ Anarchist
 - ▬ Homeless
 - ▬ Union members
 - ▬ Reinforcement campers from other cities
 - ▬ Religious leaders or communities
 - ▬ Af-Am

- ▬ e.g. "172 people have been arrested over the course of the protest movement."
 - ▬ e.g. "In 11 nights of marches, 14 people have been injured."
 - ▬ e.g. "Between the city's two separate camps, at least 180 tents have been set up."
- ▬ not clear if the event really happened or was just rumored to have happened or what – needs to be reviewed in light of other articles
- ⊕ an event initiated by neither police, city, nor protesters that seems likely to have affected the outcome of the movement in some way
 - ▬ E.g. "a man committed suicide near the Occupy camp on Monday, leading protesters and police to fear for the safety of the camp..."
- ⊕ Details of such events to be coded:
 - ⊕ Attendance – numbers of...
 - ⊕ Police presence characteristics
 - ⊕ ACTIONS of police or protesters that occurred during the event
 - ⊕ Date/Time tag
 - ▬
- ⊕ **Camp Information**
 - ▬ # of people camping overnight (not MERELY attending camp establishment or move of camp event) (also color code statements about relative size e.g. "the camp is getting bigger")
 - ▬ # of tents (also color code statements about relative size e.g. "the camp is getting bigger")
 - ⊕ Permitted? NOTE: sometimes protesters will have a legal right to a space for a certain time based on a lease or a court's injunction or restraining order against the police. We will treat these as permits.
 - ▬ From when
 - ▬ Till when
 - ▬ With what conditions
 - ▬ Informally allowed
 - ▬
 - ⊕ COMPOSITION of CROWD (tags, estimate codes, and % estimate codes)
 - ▬ Anarchist
 - ▬ Homeless
 - ▬ Union members
 - ▬ Reinforcement campers from other cities
 - ▬ Religious leaders or communities
 - ▬ Af-Am

- Facilities
 - Security team
 - Code of conduct
 - Libraries
 - Kitchens
 - Legal services
 - Drug harm reduction centers/ needle exchanges
 - Health clinic
 - Meditation tent
 - Social media (Twitter, Facebook, number of users, etc.)
 - ...
 -
- Official statement from camp or official spokesperson
 - Code the act of speaking, writing an open letter, op-ed or press release in brick along with the name and or title of the official spokesperson
- Visibility to the rest of the public
 - High (e.g. near city hall)
 - Medium
 - Low (away from the public)
 - Social media (Twitter, Facebook, number of users, etc.)
 -
- GENERAL ASSEMBLY ACTIVITY
 - # People attending general assemblies
 - Frequency of meetings
 - Votes
- Civilian Government Action
 - Order or threat of enforcement or legal consequences, or statement or intention not to enforce
 - Order to disperse
 - New Curfew order (might be bracketed in brick, too)
 - Announcement of intention to enforce or not enforce regulation (not permits) relating to encampment, marches, demonstrations, etc.
 - New Regulation
 - Warning
 - **New!** Inspections (Fire, Safety, etc.)
 - Indication of intention NOT to enforce

- - ⇒ Dispersal order
 - ⇒ Refusal of basic needs like sanitation services, water, electricity, (port-a-jon, porta-potty, etc.)
 - ⊕ Other official statement from city council member or higher city officer or spokesperson for such an officer
 - ⊕ Evidence of city ordinances
 - ⇒ Against semi-permanent structures
 - ⇒ Against overnight camping
 - ⇒ ETC...
 - ⊕ COSTS to City
 - ⇒ Estimates of damage, cost of extra personnel, etc.
 - ⊕ Meeting
 - ⊕ Attendees
 - ⊕ Subjects discussed
 - ⇒ Date/Time tag
 - ⊕ Relevant Actors
 - ⇒ Governor
 - ⇒ Mayor
 - ⇒ City Administrator
 - ⇒ Vice Mayor
 - ⇒ City Council (as a body)
 - ⇒ ...
 - ⇒ Mayor
 - ⇒ City Administrator
 - ⇒ City Council/Assembly
 - ⇒ Governor
 - ⇒ Attorney General of State
 - ⇒ City attorney (if speaking as advisor or legal expert on city ordinances and not as representative of city in an official court proceeding or as negotiator)
 - ⊕ City parks and recreation office (or analogous office)

TUA identification complete... Then...

Under cover of darkness early Tuesday, hundreds of police swept into Oakland's Occupy Wall Street protest, firing tear gas and beanbag rounds before clearing out an encampment that demonstrators had hoped would stir a revolution. In less than an hour, the two-week-old, miniature makeshift city [two-week-old, miniature makeshift city] was in ruins. Scattered across the area were overturned tents, pillows, sleeping bags, yoga mats, tarps, backpacks, food wrappers and water bottles. [Scattered across the area were overturned tents, pillows, sleeping bags, yoga mats, tarps, backpacks, food wrappers and water bottles.] Signs decrying corporations and police still hung from lampposts or lay on the ground. Protesters had stayed awake through the night, waiting for the expected raid. Officers and sheriff's deputies from across the San Francisco Bay area surrounded the plaza in front of City Hall at around 5 a.m. and closed in. Eighty-five people were arrested, mostly on suspicion of misdemeanor unlawful assembly and illegal camping [misdemeanor unlawful assembly and illegal camping], police said.

Variable/Attribute Extraction from TUAs

txt-thrshr

with Justice and the AFL-CIO, at Downtown Crossing on October 13, 2011. On November 17, 2011, hundreds of Occupy Boston and MassUniting marchers (labor unions, community organizers; SEUI, Local Ironworkers, Jobs for Justice, and more) took to the city streets again in solidarity with another global day of action.

Show full article

Event type

Event setting and composition (location, attendance, time/duration)

Arrests/injuries

Permit

Community response

Hide Ontology

Event type

- What type of event is described in the highlighted text?
 - What did the crowd block?
 - What type of event did the protesters disrupt?
 - What is the position or title of the 1% who is being protested against?
 - What type of strategic violence occurred?
 - What type of strategic sabotage occurred?

Event setting and composition (location, attendance, time/duration)

- Which of the following information does the text give about the event's location?
 - Exactly where did the event take place?
 - Approximately how many protesters were in attendance?
 - Who estimated the number of protesters in attendance?

Arrests/injuries

- Which of the following information does the text give about arrests that occurred during the event?
 - Exactly how many arrests occurred during the event? (if the text specifically states, "no arrests were made," then type in the number 0)
 - Approximately how many arrests occurred?
 - Who was arrested?

Permit

- Was the event permitted?
 - When does the event permit start?
 - When does the event permit expire?
 - Does the event permit come with conditions or requirements that the protesters must follow?
 - Which are conditions or requirements of the event permit? Select all that apply:

Community response

- Which of the following parties responded to the event?
 - What was the party's level of support?

Variable/Attribute Extraction from TUAs

txt-thrshr

< Rewind

What type of event is described in the highlighted text?

March/Parade

Rally/Demonstration

Strike

Blocking Action

Establishing a Camp

Moving a camp to a new location

Disrupting an on-going event of the perceived 1%

Divestment Action

Voluntary Dissolution of a camp

Strategic Violence

Strategic Sabotage

with Justice and the AFL-CIO, at Downtown Crossing on October 13, 2011. On November 17, 2011, hundreds of Occupy Boston and MassUniting marchers (labor unions, community organizers; SEUI, Local Ironworkers, Jobs for Justice, and more) took to the city streets again in solidarity with another global day of action.

Show full article

Hide Ontology

Event type

- What type of event is described in the highlighted text?
 - What did the crowd block?
 - What type of event did the protesters disrupt?
 - What is the position or title of the 1% who is being protested against?
 - What type of strategic violence occurred?
 - What type of strategic sabotage occurred?

Event setting and composition (location, attendance, time/duration)

- Which of the following information does the text give about the event's location?
 - Exactly where did the event take place?
 - Approximately how many protesters were in attendance?
 - Who estimated the number of protesters in attendance?

Arrests/injuries

- Which of the following information does the text give about arrests that occurred during the event?
 - Exactly how many arrests occurred during the event? (If the text specifically states, "no arrests were made," then type in the number 0)
 - Approximately how many arrests occurred?
 - Who was arrested?

Permit

- Was the event permitted?
 - When does the event permit start?
 - When does the event permit expire?
 - Does the event permit come with conditions or requirements that the protesters must follow?
 - Which are conditions or requirements of the event permit? Select all that apply:

support?

Don't Worry

- Protocols are less complex than a massive hand-coding scheme.
- But, protocols make all of this possible.
- Protocols list detailed instructions for applying a hand-coding scheme to raw text.
- Protocols for variable/attribute extraction are converted into a sequence of instructions and pop-up question boxes in Text Thresher.

Our Protocols

- Determining whether or not a city should be in our dataset
- Gathering news articles (web-scraping by hand)
- Formatting news articles (could have done this in python)
- Identifying text units of analysis (TUAs)
 - 1st and 2nd pass
- Training new RAs
- Data clean-up
- Text Thresher variable-question taxonomies for each TUA-type
- Text Thresher crowd worker training
- Actor Dictionaries

Examples

[Imagine a few seconds of elevator music while I pull up a chrome window, here.]

Protocols Writing Epistemology

- **Phenomenologist/Pragmatist Epistemology:** there is no objectivity, only inter-subjectivity.
- Intersubjective agreement ranges from extremely-high-but-imperfect to non-existent.
- Our unachievable goal is to create protocols with instructions that are understood exactly the same by everyone who will ever read them.
- Our perhaps-achievable goal is to create protocols with instructions so unambiguous, so belabored and redundant that every person on our project will understand them so similarly that they perform research tasks (e.g., text classification) identically to one another.

How to Write Protocols that Work

- Step 1: Figure out a good way to do the research process the protocol will detail.
 - Find the news articles yourself
 - Observe the phenomena you want others to observe for you
- Step 2: Re-do the process, writing down everything you are doing in sequential order.
- Step 3: Forgetting that you already know how to perform the process, try to follow your own steps as a naïve robot would and see if it works. You will find that you included many assumptions.

How to Write Protocols that Work

- Step 4: Imagine alternative universes. What if the RA is working in a different browser? How would the process change? Account for (with if/then statements and/or contingency steps) or kill those alternative universes. (e.g., Google search by locale).
- Step 5: Anticipate RA's worries. Will they worry that Lexis/Nexis will uncheck boxes as they scroll through multiple pages of news results? Allay those concerns when you can. It will save them valuable cognitive effort.
- **You're not nearly done!**

How to Write Protocols that Work

- Step 6: Share your protocol with your team. Allow them to try it **AND to edit it and make comments on it!** Best to do this around a table with the protocol open in GoogleDocs or Gist.
- Step 7: **ENCOURAGE** RAs to improve the document
 - Make a hero out of that obtuse or contrarian RA!
 - Don't let the RAs who love you believe that you have a shred of pride invested in the protocol.
 - Encourage them to add examples to show their understanding and guide other protocol users
 - BUT, Ensure those examples aren't too narrow or too broad
- Step 8: Iterate Steps 3-7 with the whole team
- Step 9-N: Test the protocol by performing the task on the same set of raw data. Do your results align? If not, keep iterating until they do.

Advantages of Great Protocols

- Drafting them clarifies methods
- Drafting them clarifies theory
- They direct research assistants even better than you can
- They remind the future-you what you did, greatly easing the task of writing up results
- They help colleagues and critics appreciate your methods and theory

Fun Task

- X minutes
- Write a protocol describing the process of making a peanut butter and jelly sandwich
- Assume the person enacting the protocol is already in his/her kitchen and all ingredients (PB, Jelly, and Bread) are in the refrigerator.
- Do not create unnecessary/inefficient steps (e.g., “do the moonwalk”).
- S/he with the most detailed protocol wins!*

*Prize TBD

Protocols to Create Software

- Gherkin: a small pickle
- Also, a language for test-based development of software.
- With only a handful of logic keywords, you can specify exactly what a computer program should do in ways that programmers will understand and can test. Upshot: you can lead a software development team without even knowing much at all about how to code.
- GIVEN some conditions (e.g., user is facing the home menu)
- WHEN a user tries to do 'x'
- THEN 'y' should happen

Read more at:

<https://github.com/cucumber/cucumber/wiki/Gherkin>

Improve My Protocol on Protocols

- Shoot me an email and I will share a Google Doc with my protocol on protocols!
- nickbadams@berkeley.edu
- Or find me at BIDS in 190 Doe Library,
...or at the D-Lab – 356 Barrows Hall

THANKS!! 😊