Qualitative Research in the Age of Transparency

The Ohio State University, July 23, 2018

Sebastian Karcher
QDR, Syracuse University

Everyone Talks about Open Data and Transparency

BITSS

Berkeley Initiative for
Transparency in the Social Sciences

ipa
INNOVATIONS FOR POVERTY ACTION

Research Transparency

the WHITE HOUSE
PRESIDENT BARACK OBAMA

BRIEFING ROOM | ISSUES | THE ADMINISTRATION | 1600 PENN

HOME · BLOG

Increasing Access to the Results of Federally Funded Science

RESEARCH DATA SHARING WITHOUT BARRIERS

COS
CENTER FOR OPEN SCIENCE

About COS | Our Products | Our Services | Our Communities | Blog | Contact Us | Donate

OPEN, TRANSPARENT, AND REPRODUCIBLE SCIENCE
IS STRONGER SCIENCE.

RDA
RESEARCH DATA ALLIANCE

ABOUT RDA

DART

DATA ACCESS & RESEARCH TRANSPARENCY

PLOS' New Data Policy: Public Access to Data

Transparency: Journals

Data Availability

The following policy applies to all PLOS journals, unless otherwise noted.

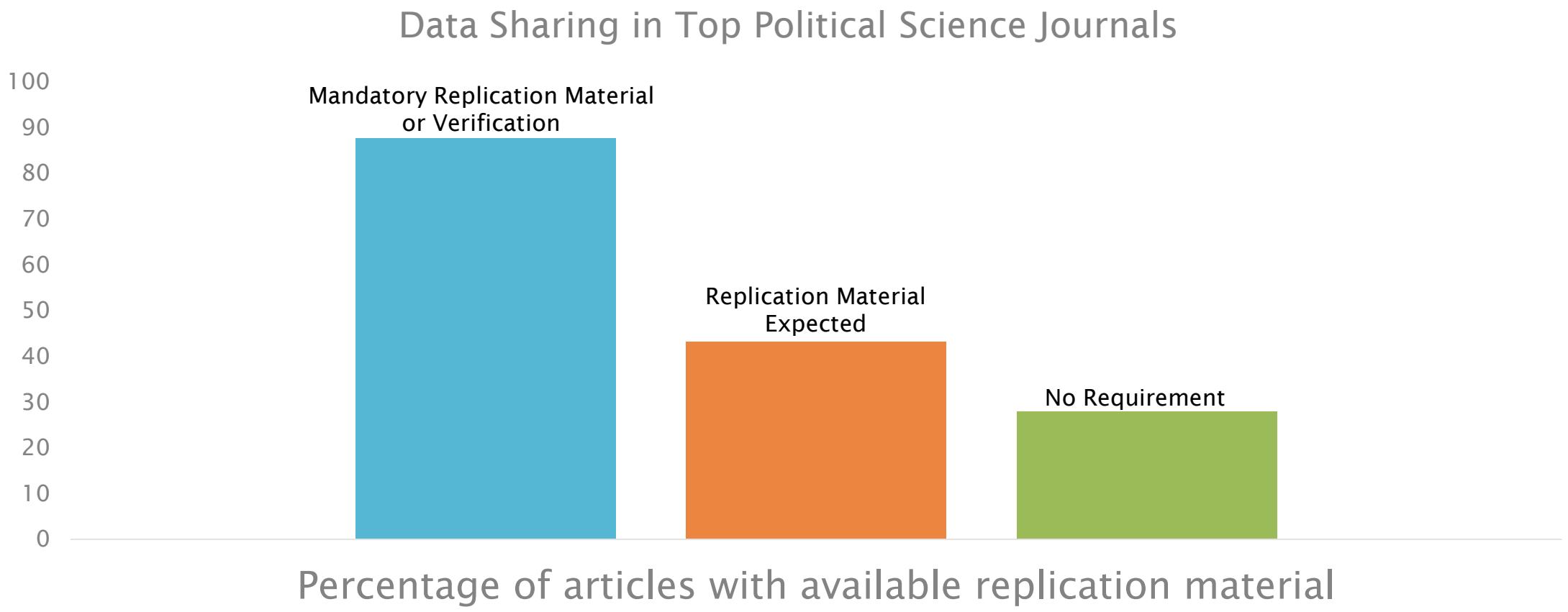
PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.

When submitting a manuscript online, authors must provide a *Data Availability Statement* describing compliance with PLOS's policy. If the article is accepted for publication, the data availability statement will be published as part of the final article.

Refusal to share data and related metadata and methods in accordance with this policy will be grounds for rejection. PLOS journal

http://journals.plos.org/plosone/s/data-availability

Journal Policies Matter



*Key 2017,https://doi.org/10.1017/S1049096516000184

Transparency: Funders

Dissemination and Sharing of Research Results

NSF DATA SHARING POLICY

Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. See Award & Administration Guide (AAG) Chapter VI.D.4.

NSF DATA MANAGEMENT PLAN REQUIREMENTS

Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled “Data Management Plan”. This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See Grant Proposal Guide (GPG) Chapter II.C.2.j for full policy implementation.

https://www.nsf.gov/bfa/dias/policy/dmp.jsp

Funder Requirements *Can* Matter

data have been shared and how effective the sharing was. We then evaluate the data sharing practices of researchers funded by the NSF at Oregon State University in two ways: by attempting to discover project-level research data using the associated DMP as a starting point, and by examining data sharing associated with journal articles that acknowledge NSF support. Sharing at both the project level and the journal article level was not carried out in the majority of cases, and when sharing was accomplished, the shared data were often of questionable usability due to access, documentation, and formatting issues. We close the article by offering recommendations for how data producers, journal publishers, data repositories, and funding agencies can facilitate the process of sharing data in a meaningful way.

Van Tuyl & Whitmire 2016:
https://doi.org/10.1371/journal.pone.0147942

Transparency: Norms*

Insecure researchers aren't sharing their data

Posted by Andrew on 4 November 2011, 10:14 am

Jelte Wicherts writes:

> I thought you might be interested in reading this paper that is to appear this week in PLoS ONE.
>
> In it we [Wicherts, Marjan Bakker, and Dylan Molenaar] show that the willingness to share data from published psychological research is associated both with "the strength of the evidence" (against H0) and the prevalence of errors in the reporting of p-values.
>
> The issue of data archiving will likely be put on the agenda of granting bodies and the APA/APS because of what Diederik Stapel did.

I hate hate hate hate hate when people don't share their data. In fact, that's the subject of my very first column on ethics for Chance magazine. I have a

http://andrewgelman.com/2011/11/04/insecure-researchers-arent-sharing-their-dat

* Or, don't mess with Gelman

Mandating Transparency: Academic Organizations

8.14 Sharing Research Data for Verification

(a) After research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through reanalysis and who intend to use such data only for that purpose, provided that the confidentiality of the participants can be protected and unless legal rights concerning proprietary data preclude their release. This does not preclude psychologists from requiring that such individuals or groups be responsible for costs associated with the provision of such information.

(b) Psychologists who request data from other psychologists to verify the substantive claims through reanalysis may use shared data only for the declared purpose. Requesting psychologists obtain prior written agreement for all other uses of the data.

APA Ethics Code:
http://www.apa.org/ethics/code/

The Political Science Experience

- Data access and research transparency (DA-RT) part of APSA ethics guidelines since 2013
- In 2014, editors from leading jo guidelines:
- Require authors to ensure that time of publication through a tr
- Require authors to delineate cle upon which their published clai
- Maintain a consistent data citat that data creators and suppliers

Data Access and Research Transparency (DA-RT): A Joint Statement by Political Science Journal Editors

Transparency requires making visible both the empirical foundation and the logic of inquiry of research. We agree that by January 15, 2016 we will:

- Require authors to ensure that cited data are available at the time of publication through a trusted digital repository. Journals may specify which trusted digital repository shall be used (for example if they have their own dataverse). If cited data are restricted (e.g., classified, require confidentiality protections, were obtained under a non-disclosure agreement, or have inherent logistical constraints), authors must notify the editor at the time of submission. The editor shall have full discretion to follow their journal's policy on restricted data, including declining to review the manuscript or granting an exemption with or without conditions. The editor shall inform the author of that decision prior to review.
- Require authors to delineate clearly the analytic procedures upon which their published claims rely, and where possible to provide access to all relevant analytic materials. If such materials are not published with the article, they must be shared to the greatest extent possible through institutions with demonstrated capacity to provide long-term access.
- Maintain a consistent data citation policy to increase the credit that data creators and suppliers receive for their work. These policies include using data citation practices that identify a dataset's author(s), title, date, version, and a persistent identifier. In sum, we will require authors who base their claims on data created by others to reference and cite those data as an intellectual product of value.
- Ensure that journal style guides, codes of ethics, publication manuals, and other forms of guidance are updated and expanded to include improved data access and research transparency requirements.

DA-RT and Qualitative Research

Area Studies and the Cost of Prematurely Implementing DA-RT

Put a DA-RT in it

Transparent Explanations, Yes. Public Transcripts and Fieldnotes, No: Ethnographic Research on Public Opinion

The Dark Side of DA-RT[1]

The Tyranny of Light

QDR Overview

- At Syracuse University
- Online since 2014, NSF funded
- Curated, qualitative & mixed methods data
- Small holdings (~45 datasets), growing steadily
- Some overlap with DA-RT

Whither Qualitative Transparency?

- Costs / Logistics
- Ethical / Legal
- Epistemic / Ontological

Costs / Logistics

- The costs involved with preparing data for archiving vary widely depending on the nature of the evidence. For research that does not involve machine-readable datasets, rendering the original sources or “raw data” in digital form for archiving can impose substantial financial and logistical burdens on researchers. What is the right balance between the costs and the benefits of rendering these types of data accessible? How can a balance be struck that does not systematically favor some modes of political analysis over others? Who should decide how to strike that balance in individual cases?

https://dialogueondart.org/petition/

Costs / Logistics: Ways Forward

- Funding for data preparation
- Lowering costs with good research practices
- Lowering costs with technology

Ethical / Legal

which political scientists at different programs engage. So, for example, *anyone* who does qualitative methods knows well that field notes are almost always confidential under human subjects provisions within one's university and are thus – literally – *illegal* to share.

http://www.e-ir.info/2016/11/24/methods-war-how-ideas-matter-within-political-science/

Ethical / Legal: Ways Forward

- De-identifying qualitative data is (often) possible
- Not all data can or should be shared – but
- Ongoing work with IRBs – increasing awareness
- Access restrictions
  - General
  - Distributed Credentialing (ICPSR)
  - Data Tags (IQSS / DataVerse)

Epistemic / Ontological

concerns. Political science needs greater intellectual diversity, not less. As Isaacs points out, the DA-RT protocol is suggestive of a epistemological and methodological disciplining move that will force scholars ever closer into a neopositivist straightjacket.

http://duckofminerva.com/2015/11/put-a-da-rt-in-it.html

Epistemic / Ontological: Ways Forward

- Find language appropriate for research tradition
- Consultative approach to data sharing
- Make Methods and Technology Meet Qualitative Practice
  - E.g. Annotation for Transparent Inquiry
  - E.g. QDA Software

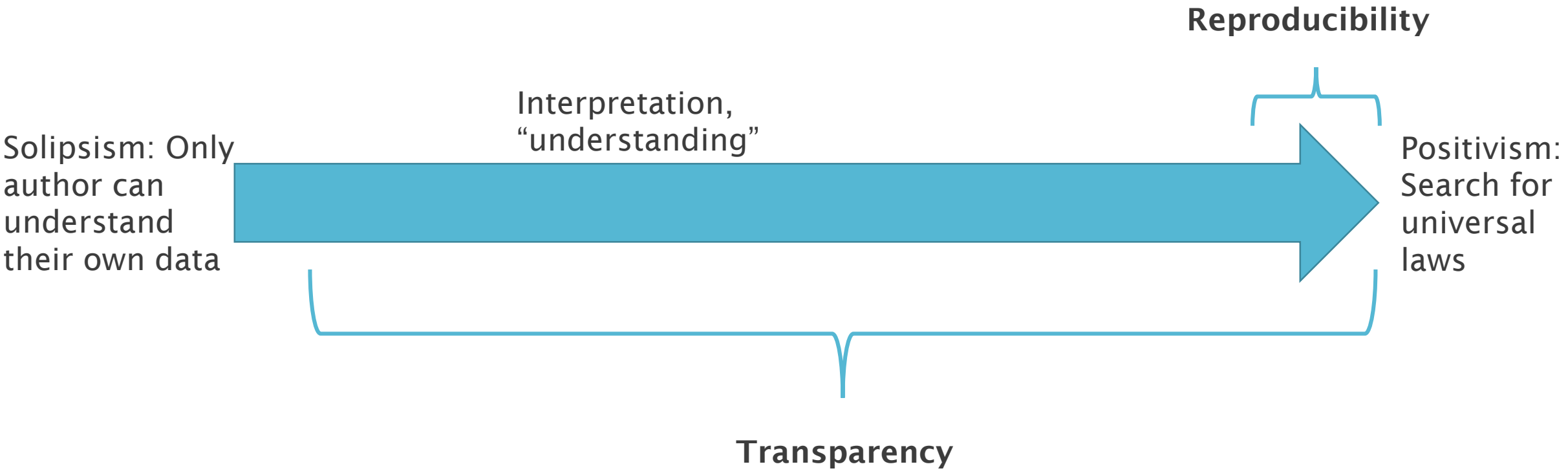
Language: Transparency vs. Reproducibility

Transparency: Show data and methods used to arrive at conclusions

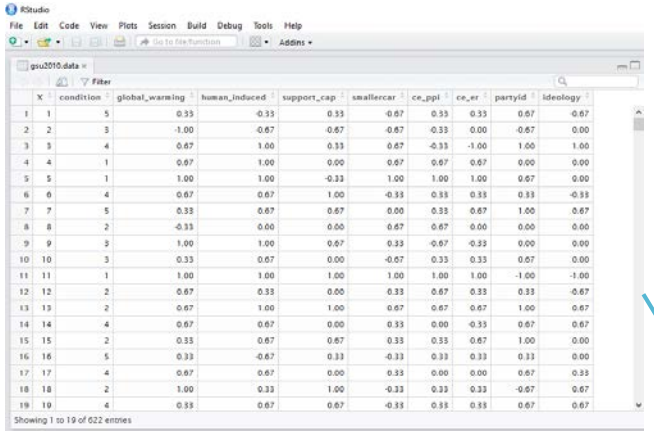
Reproducibility: Same data, same methods → same results

In quantitative research, reproducibility is a *byproduct* of transparency

Transparency vs. Reproducibility: Epistemology



Quantitative Research: Matrix Data



	X	condition	global_warming	human_induced	support_cap	smallercar	ce_ppl	ce_er	partyid	ideology
1	1	5	0.33	-0.33	0.33	-0.67	0.33	0.33	0.67	-0.67
2	2	3	-1.00	-0.67	-0.67	-0.67	-0.33	0.00	-0.67	0.00
3	3	4	0.67	1.00	0.33	0.67	-0.33	-1.00	1.00	1.00
4	4	1	0.67	1.00	0.00	0.67	0.67	0.67	0.00	0.00
5	5	1	1.00	1.00	-0.33	1.00	1.00	1.00	0.67	0.00
6	6	4	0.67	0.67	1.00	-0.33	0.33	0.33	0.33	-0.33
7	7	5	0.33	0.67	0.67	0.00	0.33	0.67	1.00	0.67
8	8	2	-0.33	0.00	0.00	0.67	0.67	0.00	0.00	0.00
9	9	3	1.00	1.00	0.67	0.33	-0.67	-0.33	0.00	0.00
10	10	3	0.33	0.67	0.00	-0.67	0.33	0.33	0.67	0.00
11	11	1	1.00	1.00	1.00	1.00	1.00	1.00	-1.00	-1.00
12	12	2	0.67	0.33	0.00	0.33	0.67	0.33	0.33	-0.67
13	13	2	0.67	1.00	1.00	0.67	0.67	0.67	1.00	0.67
14	14	4	0.67	0.67	0.00	0.33	0.00	-0.33	0.67	0.67
15	15	2	0.33	0.67	0.67	0.33	0.33	0.67	1.00	0.00
16	16	5	0.33	-0.67	0.33	-0.33	0.33	0.33	0.33	0.00
17	17	4	0.67	0.67	0.00	0.33	0.00	0.00	0.67	0.33
18	18	2	1.00	0.33	1.00	-0.33	0.33	0.33	-0.67	0.67
19	19	4	0.33	0.67	0.67	-0.33	0.33	0.33	0.67	0.67

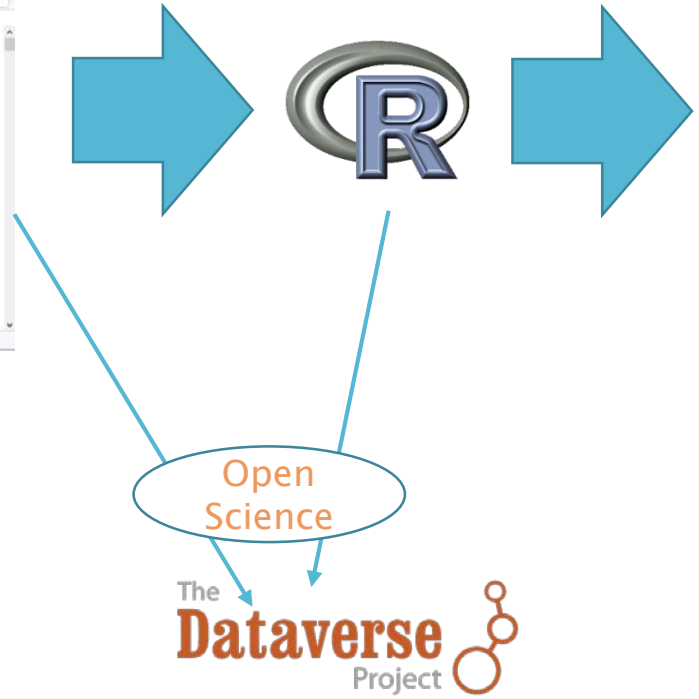


Table 2. Treatment Effects (Experiment 1)

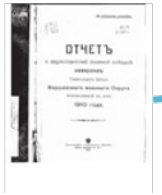
	Perceived Personal Influence	Expected Reciprocity	Believe Global Warming is Happening	Believe Global Warming is Human Induced	Support Cap on Carbon Emissions	Willingness to Take Personal Action
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.38**	0.23**	0.37**	0.25**	0.28**	0.20**
	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Pro Descriptive (D)	-0.04	0.02	0.05	-0.03	-0.06	-0.02
	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)	(0.09)
Pro Descriptive + Injunctive (D+I)	-0.08	0.07	0.09	0.06	-0.02	-0.05
	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)	(0.08)
Con D	-0.18**	-0.06	-0.06	-0.05	-0.16**	-0.17**
	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)	(0.09)
Con D+I	-0.10	0.11*	0.03	-0.06	-0.08	-0.06
	(0.08)	(0.07)	(0.09)	(0.09)	(0.09)	(0.09)
PartyID	0.21**	0.21**	0.28**	0.31**	0.21**	0.13*
	(0.08)	(0.08)	(0.11)	(0.09)	(0.10)	(0.09)
PartyID * Pro D	-0.05	-0.04	-0.10	-0.04	-0.05	0.04
	(0.11)	(0.11)	(0.14)	(0.13)	(0.13)	(0.13)
PartyID * Pro D+I	-0.10	-0.12	-0.06	-0.13	-0.14	0.04
	(0.11)	(0.11)	(0.13)	(0.12)	(0.13)	(0.11)
PartyID * Con D	-0.10	-0.19**	0.00	-0.10	0.03	0.11
	(0.11)	(0.11)	(0.14)	(0.13)	(0.13)	(0.13)
PartyID * Con D+I	-0.13	-0.21**	0.01	-0.15	0.01	0.04
	(0.11)	(0.10)	(0.13)	(0.12)	(0.13)	(0.11)
SER	.48	.47	.52	.53	.51	.54
n	612	612	612	611	612	612

*p<.10, **p<.05, one-tailed test
Note: D = Descriptive, I = Injunctive. Cell entries are unstandardized linear regression coefficients with bootstrapped standard errors in parentheses. Baseline is the control condition. Party identification is coded with seven categories from Republican (-1) to Independent (0) to Democrat (+1). SER is the Standard Error of the Regression (Beck, 2010).

While we expected norms promoting action to influence both beliefs (Hypothesis 1a) and behavioral intentions (Hypothesis 2a), we find little support our hypotheses. Table 2 reports regression results with indicators for each treatment condition (relative to a control condition), including a control for party identification. Looking at the rows for the Pro Norm conditions, we do not see any effect on any outcome. Indeed for several outcomes (perceived personal influence and support for a carbon emissions tax), we find effects in the opposite direction of our

Toby Bolsen, Thomas J. Leeper, and Matthew Shapiro. 2014. "Doing What Others Do: Norms, Science, and Collective Action on Global Warming." *American Politics Research* 42(1): 65–89.

Qualitative Research: Granular Data



Analysis

Analysis

Analysis

Analysis

DANILOV'S DEFENSIVE PLAN OF 1910

The Russian war plan of 1910 was more defensive and more oriented toward Germany than any other Russian plan made between 1880 and 1914. General Danilov, the principal draftsman of the plan, allocated 53 divisions to the German front and 19 to the Austrian front. His deployments abandoned Warsaw, the forward fortresses, the Narew River barrier, and indeed most of Poland. Forces deployed against Germany and Austria were to advance after completing their concentration, if circumstances permitted. Rail and road improvements were, however, planned with strictly defensive operations in mind. The abandoned forward zones were to be stripped of all supplies.[1]

Danilov gave two reasons for concentrating forces far from the frontier and primarily against Germany. One was Russian military weakness after the Russo-Japanese War, including the continuing lag in the speed of Russian deployments, and the other, the possibility of a large-scale German attack against Russia and the uncertainty of adequate French support.

The Russo-Japanese War and its attendant political and economic difficulties had a ruinous effect on the Russian army. Matériel stocks were drawn down and not replaced because of Russia's financial crisis, which lasted almost until 1910. Official targets for the stockpiling of equipment and supplies would not even be approached until 1913 or 1914.[2] Between 1906 and 1908 as many as one-third of Russia's soldiers were used for internal security duties, which sapped morale and interfered with training.[3] Moreover, Russia's defeats in Manchuria had called into question the fitness of existing military doctrines, organizational forms, and commanders. The first few years after the war were spent debating the doctrinal and institutional implications of the defeat. Significant reform and reorganization were delayed until 1908 and 1910, when funding was increased and policymaking powers were concentrated in the hands of the war minister.[4]

Russian military figures warned at several times during this period that the army's devastated condition made the use of force dangerous [end of 166] against even Turkey, not to mention Austria or Germany.[5] Such was the state of disorganization even as late as 1909 and 1910 that it would have been "utterly impossible to make war," according to General A. Brusilov, who commanded the Russian Eighth Army. In Brusilov's high-priority corps in the Warsaw military district, supplies were barely sufficient even for peacetime requirements, and the transport service was in chaos.[6]

Annotation for Transparent Inquiry (ATI)

See more on qdr.syr.edu/ati

Any digitally published manuscript can be annotated using ATI (here: an article in *International Organization* published by Cambridge University Press)

Hungary, this was their only stated concern. However, many states conditioned their recognition decision on an action related to Indian troop withdrawal and gave three different types of reasons for doing so. States also differed in the extent of troop withdrawal they required before recognition. See Table 2 for a full list of states, their stated reason for conditioning recognition on withdrawal (if any can be identified), and what recognition was conditioned on (whether actual withdrawal or a proxy).

The first type of reason, opposition to condoning or legitimizing aggression, is labeled as "Non-aggression." A good example comes from Mexican Foreign Minister Emilio Óscar Rabasa who reported that the Mexican president had decided not to recognize Bangladesh because, "since the Mexicans, like many Latin Americans, refuse to condone territorial aggrandizement as a result of war, they would prefer to wait on the withdrawal of Indian troops as the sign of true independence."[90]

This statement also appeals to "true independence." Self-determination is another important value expressed by the Mexican representative and is the second type of reason commonly appealed to as justifying recognition as Bangladesh. For

88. See Figure 2.
89. A frequent concern was that states had to recognize in a group, or on the same day as multiple other states. However, even allowing for minor coordination problems, this in and of itself cannot explain the length of time taken to make recognition decisions and declarations.
90. Cable from Hope, 16 January 1972, FCO 37/1020.

QDR Jan 31
CambridgeCore/ATI

they would prefer to wait on the withdrawal of Indian troops as the sign of trueindependence."

Annotation for Transparent Inquiry (ATI) QDR

Full Citation: Sir Peter Hope, UK Ambassador to Mexico, a confidential telegram from Hope to the Foreign and Commonwealth Office, 26 January 1972. Folder 37/1020 of the FCO Archives held at the National Archives at Kew, UK.

Source Excerpt: Rabasa said that, since the Mexicans, like many Latin Americans, refuse to condone territorial aggrandizement as a result of war, they would prefer to wait on the withdrawal of Indian troops as the sign of true independence.

Analytic Note: This is a confidential telegram from UK Ambassador to Mexico Sir Peter Hope to the Foreign and Commonwealth Office of 26 January, 1972, from folder 37/1020 of the FCO Archives held at the National Archives at Kew, UK. This excerpt shows that the Mexican Foreign Minister, Emilio Oscar Rabasa, gave as a reason for the Mexican President's decision not to recognize Bangladesh, that they did not want to condone territorial aggrandizement as a result of war until Indian troops had been withdrawn. The telegram also indicates that this reason and another reason, i.e. that Mujib's assumption of several cabinet portfolios cast doubt on the fact that his government had been elected by the people, were the only two reasons cited by the Mexican government.

Data Source: https://data.beta.qdr.org/api/access/datafile/2529?key=13e4c93f-1172-4d53-8a07-6f2651e5da97

Less

ATI Annotation: Displayed alongside article. Created by author, curated by QDR, hosted and served by Hypothesis, displayed on publisher's web site

Elements of an ATI annotation: One or more of the following:

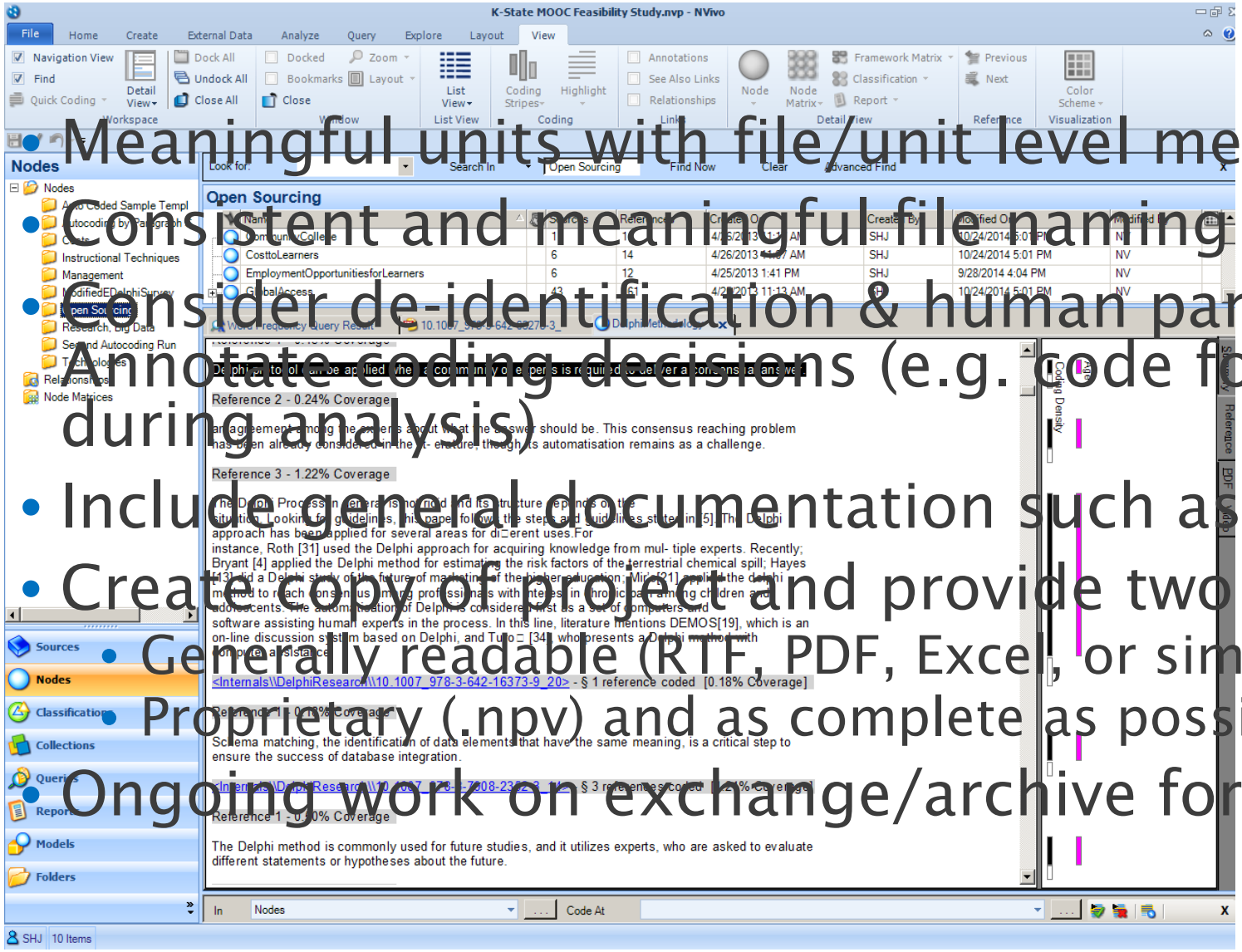
- Analytic Note
- Source Excerpt
- Source Excerpt Translation
- Link to Data Source

Link to data source housed in QDR

Any passage in the text or in notes of a manuscript can be annotated using ATI

Example: bit.do/qdr-ati-omahoney

Sharing QDA Data



- Meaningful units with file/unit level metadata
- Consistent and meaningful file naming
- Consider de-identification & human participant concerns
- Annotate coding decisions (e.g. code for de-identification during analysis)
- Include general documentation such as coding guidance
- Create copy of project and provide two exports:
  - Generally readable (RTF, PDF, Excel, or similar)
  - Proprietary (.npv) and as complete as possible
- Ongoing work on exchange/archive format