

Censorship and Self-Censorship in Research on the Iraq War

By

Michael Spagat

Department of Economics

Royal Holloway, University of London

Abstract

I wrote a paper that found evidence of fabrication in public opinion surveys fielded in Iraq by D3 Systems and KA Research Limited. I sent the paper out for comments to various knowledgeable people including at D3 and commenced litigation against me in an attempt to suppress the research. I describe this experience and draw lessons about censorship, self-censorship, academic freedom and democracy.

Preliminary Draft – Comments welcome

Paper Presented at the International Studies Association 57th annual convention, Atlanta, GA, March 16, 2016 in the panel on “Censorship, self-censorship and prudent restraint.”

1. INTRODUCTION

There have been numerous public opinion polls conducted in Iraq since the 2003 invasion. Many were fielded by [D3 Systems](#) working with [KA Research Limited](#) (hereafter “D3/KA”). These companies have impressive client lists that include the BBC, The US Department of State, the World Bank and the University of Michigan, to name just four out of dozens.

Much D3/KA survey work is conducted in challenging environments such as Iraq, Afghanistan and Yemen. The D3 website states that

“Over the past ten years, we have become best known for our work in conflict and post-conflict environments, such as Afghanistan and Iraq.”

Indeed, in 2008 [ABC News won an Emmy Award](#) for an opinion poll fielded by D3/KA in Iraq.¹

The prospect of retrieving high-quality polling data from war zones is tantalizing and it is reasonable to accept some quality compromises in the interest of creating such interesting data. However, any such concessions should be made explicitly amidst an open and honest discussion of data quality. Purveyors of polling data have a responsibility to proactively inform their clients of any important weaknesses that may be present in the polling data they deliver to these clients.

In practice, both individuals and organizations tend to hide their weaknesses. Humans are not generally eager to draw attention to their own errors. Moreover, corporations may fear losing market share if they admit to flaws in past work. Thus, we often depend on outsiders to expose problems that individuals or corporations endeavour to cover up. Academic freedom is an important asset within the arena of ideas because it can enable freewheeling debate while limiting the opportunities for powerful actors to prevail through intimidation.

But what happens when corporations resort to legal threats to censor discussion of their work?

I received just such a threat from D3 Systems and it did silence me for several years. My experience raises serious questions about censorship, self-censorship, academic freedom and even democracy. In section 2 of this paper I describe my experience. I situate this experience within a broader context in section 3 and then draw some conclusions in section 4.

¹ The poll was sponsored jointly by ABC, the BBC, USA Today and ARD of Germany.

2. THE STORY

2a. Background

In 2011 I analysed five public opinion polls that were fielded in Iraq by D3/KA between December 2005 and November 2008.² Two were sponsored by the [Program on International Policy Attitudes](#) (PIPA, now the Program for Public Consultation) of the University of Maryland. Three were sponsored by the [Broadcasting Board of Governors](#) (BBG), a US government agency.

I found evidence of data fabrication in the surveys which I wrote up in a paper. Here I give just a brief sketch of this evidence before describing how D3 Systems tried to censor these findings.

2b. Evidence of fabrication

The data are rich in micro information providing, in particular, identification numbers for the person conducting each interview as well as for his/her supervisor. Two supervisors preside over suspicious data throughout all five surveys. Five other supervisors appear for some of the surveys and always preside over suspicious data when present. I will refer to these supervisors as “focal supervisors”. The percentage of interviews they supervise rises steadily from around 16% in 2005 to around 44% in 2008 while the number of their interviews ranges between 332 and 687.

There are two broad forms of evidence suggesting that most or all of the interviews of the focal supervisors are fabricated. First, the data of the focal supervisors are not credible in their own right. Second, the data look incredible in comparison to the data of the other supervisors.

For example, analysis of categorical variables uncovered dozens of questions for which *none* of the respondents of the focal supervisors used one of the offered responses.³ Responses to surveys contain substantial random components⁴ so such widespread failure to use some categories is not credible in its own right without reference to the data generated by other supervisors. At the same time, this evidence of fabrication is strengthened by the fact that the other supervisors have no such empty categories. There is a similar phenomenon for the responses of “don’t know” and “refused”; the respondents of the focal supervisors do not use these

² I did this work originally with Steven Koczela of MassINC Polling Group but, under pressure, he signed a legal agreement not carry the work forward. So I proceed now on my own and will write the present paper in the first person.

³ In fact, for many of these questions there are multiple categories that are not used by any respondents for the focal supervisors.

⁴ See [Tourangeau, Rips and Rasinsky \(2000\)](#).

categories in dozens of questions whereas both categories always get some use for the other supervisors.

Each survey has a question that appears early in the questionnaire and is then, essentially, repeated late in the questionnaire. The respondents for the focal supervisors display correlations close to 0 on these questions. Such a strong form of response randomness is not credible in its own right. Moreover, the respondents for the other supervisors display large positive correlations in their responses to these questions, further strengthening the evidence of fabrication for the focal supervisors.

Television viewing times display remarkable diagonal patterns for the focal supervisors. In particular, there are blocks within which one household of a focal supervisor turns its television on at 19:00 and off at 19:30, the next household for this supervisor turns on at 19:30 and off at 20:00, the next goes on at 20:00 and off and 20:30, etc..

The above description just highlights some of the evidence of fabrication. Nevertheless, it should be sufficient to establish at least the minimal point that there is a serious issue with these surveys that must be addressed.

2c. Potential depth of the problem

The BBG surveys were obtained through a Freedom of Information request to the US government made by Steven Koczela.⁵ The PIPA surveys are publicly available except for the fact that the Roper Center charges a fee for providing one of them.

There are other surveys that were fielded in Iraq by D3/KA between December 2005 and November 2008 but I have been unable to obtain any of them. Steven Koczela made a request more than four years ago to the US State Department for their surveys but has received no response. It is likely that this work informed US diplomatic and military policy in Iraq, particularly during the “Surge” period.

There were other surveys sponsored by consortia including the BBC and ABC news during this time period for which I do not have the data. All these polls must be considered to be under a cloud until their data are made available and analysed. In fact, there is no reason to suppose that the problem ended in 2008. These surveys also probably influenced government policies in the US, UK and beyond. Moreover, they fed into public debates and elections.

⁵ He obtained a fourth survey (from March of 2005) in addition to the surveys described above. However, this survey did not contain any of the supervisor numbers that caused problems in the later surveys so I did not discuss it in the original paper.

2d. Circulation of the findings

I wrote the findings up in a paper, submitted it to a journal and sent copies to anyone I could think of with a direct interest in the work, including a researcher at D3.

The main reason for distributing the paper was to receive comments for use when revising the paper for the journal after, hopefully, receiving a “revise and resubmit”. People close to the surveys I analysed are in a good position to spot possible errors in the paper and suggest improvements. It is, obviously, better for me to identify and address weaknesses before publication rather than afterwards.

Courtesy was a second important motivation to share the first draft with researchers involved in the work. This way none of these people would need to improvise a response after the publication of the paper. Moreover, by sharing I placed them in a position to influence the direction of the paper, something which was as beneficial for them as it should have been for me.

Third, the findings in the paper suggest that D3/KA should investigate their Iraq operations. Thus, distributing my paper was of potential use to D3/KA to improve the quality of their work.

2e. Censorship

Shortly after sending the paper out for comments and refereeing I received a threatening letter from a lawyer under the subject heading “Re: D3 Systems, Inc. v Michael Spagat and Steve Koczela”:

“This firm represents D3 Systems, Inc. (“Our Client” or “D3”). Our client has retained us to commence litigation against you and any entity with which you are affiliated (including....) seeking compensation for, and equitable relief to terminate, your distribution and publication of false and defamatory statements about D3 to its clients and others.”

The survey firm Langer Research Associates⁶ supported the D3 threat:

“D3 and Langer Research Associates (one of D3’s clients to whom you delivered the Subject Document) have exhaustively reviewed the Subject Document and conclusively determined that it is false, misleading and asserts facts and conclusions that are incorrect. These reports have concluded that you have intentionally manipulated the data in the Subject Document in order

⁶ The involvement of Langer Research Associates can be explained by the fact that its founder was the pollster at ABC news when it won its Emmy award for Iraq polling.

to create the appearance of an artefact of fabrication. They have demonstrated that the artificial grouping of supervisors into different categories and selective presentation of your findings do not present an accurate view of the data. Your selective grouping of supervisors and variables creates misleading patterns that you have falsely asserted is evidence of potential fabrication. When D3 analyzed the data, each of the so called 'tests' that you present no longer substantiate your claims."

The letter goes on at some length without specifying any particular weakness I should address. It then concludes:

"Accordingly, WE HEREBY DEMAND, on and in behalf of our client that you:

1. IMMEDIATELY CEASE AND DESIST further delivery, dissemination, distribution and publication of the Subject Document and any of its content.
2. Deliver to this office within 8 days of the date of this letter a complete and accurate list of all parties to whom you have delivered, or requested publication of, all or any portion of the Subject Document. We note that you have refused to provide this information to our client and we warn you that your continued refusal to do so is not only further evidence of your intention to interfere with our client's business relations, but also serves to exacerbate the financial damage your actions have inflicted on our client and therefore to increase the amount of compensatory and punitive monetary damages that our client will seek from you.
3. Deliver to this office within 8 days of the date of this letter a list of all individuals and organizations with whom you communicated, or from whom you received information, in connection with the preparation of the subject document.

This letter does not constitute an admission, waiver, agreement or forbearance of any kind. We hereby reserve all of our client's rights and remedies."

2f. Immediate consequences of the threat

The legal action led my employer, RHUL, to spend considerable money on legal fees. We eventually replied, saying that I want to make the work as accurate as possible and will be happy to address any specific weakness D3 can identify. There was no reply to this letter. It immediately became clear that the legal threat was empty.

Yet the censorship strategy followed by D3 and Langer Research Associates succeeded along multiple dimensions. First, we had to waste academic resources on legal consultations. Second, the threat kept this story buried for more than four years. On the other hand, I have to admit that I am partly responsible for the suppression of my own research. I will return to this issue in the next sub-section which discusses the borderline between censorship and self-censorship.

People at D3 and Langer Research Associates might reasonably have expected to damage my position with my employer although this did not happen. Still, many employers might have questioned whether it is really worth it to have an academic who is so expensive to maintain. Thus, the censorship strategy cuts to the core of academic freedom. Many universities may not be as supportive as RHUL has been. There could be some academics who have not been promoted or have been denied tenure in part because of threats received from powerful actors.

2g. Censorship and self-censorship

I allowed this story to stay buried for more than four years and realize now that this was a mistake. Hopefully, this sub-section will prevent some people from following my path in the future.

My thinking was that there were people poised to challenge my paper so I should not release it until I was absolutely certain it was 100% error free. This is a recipe for paralysis since such standards are impossible to achieve. It may have been possible to approach this goal with enough work. However, over the last four years I had to grapple with major and time consuming family issues and also with the time demands of being Head of Department. So I have continually postponed this project. Thus, I allowed censorship to cross over to self-censorship.

I believe the above dynamic plays out regularly in various guises. A researcher may have an important point to make but knows that critics will try to pick apart her paper upon release. She may then respond by striving for perfection. The result is delay, possibly infinite delay.

I believe that the solution is to recognize that all research is flawed and that there is no shame in releasing a paper that turns out to have an error in it. When this occurs the right decision is to correct the error and move on.

I should stress that the original sin of censorship lies with D3 and Langer Research Associates. It is simply unacceptable to respond to an academic critique with a legal threat. Such actions hide the truth, restrict academic freedom and

weaken democracy.⁷ But it is also important for academics to stand up for these principles when they are threatened.

3. THE BROADER CONTEXT

[Kuriakose and Robbins \(forthcoming\)](#) (hereafter “KR”) show that data fabrication occurs widely in survey research, especially in surveys conducted in relatively poor countries. In particular, they find widespread duplicate and near-duplicate observations in many of the more than one thousand public opinion surveys they analyse. This pattern suggests that many survey researchers are cutting and pasting observations multiple times into their datasets, sometimes with alterations.⁸

The KR work has two main implications for the D3/KA situation. First, it suggests that the Iraq work of D3/KA is far from unique in having serious quality problems so one could argue that it is unfair to single out D3/KA. Here it is instructive to contrast the responses of D3 with Arab Barometer to this issue. Both were confronted with evidence of fabricated data. Arab Barometer has been carefully reevaluating their data, making corrections and improving their data collection methods to improve their future work.⁹ D3 denied they have a problem and threatened to sue. So I believe that it is more than fair to single out D3. Certainly I have far more confidence in the data of Arab Barometer going forward than I do with D3 data.

The second implication of the KR work for the D3/KA Iraq work is that duplication could be the main fabrication mechanism at work in these surveys. I have begun investigating this question using KR’s Stata programme for detecting duplicates and near duplicates. Indeed, it looks like duplication is a big part of the story, at least for the BBG surveys.

[Faranda \(2015\)](#) describes how the State Department hired outside consultants to investigate possible fabrication in surveys fielded by two firms in one (undisclosed) country. The consultants found widespread duplication in the data of seven supervisors. I note that the D3/KA surveys I analysed also found seven such supervisors and there were also two field teams for these surveys. Nevertheless, Faranda certainly could have been describing unrelated work in another country.

Faranda (2015) goes on to describe new measures implemented in State Department surveys to address the fabrication issue. Thus, the State Department and Arab Barometer are responding positively to our expanding knowledge of the

⁷ Note that much of the work in question was funded from the public purse.

⁸ [Simmons et al. \(2016\)](#) dispute the Kuriakose and Robbins findings, at least their application to surveys done at the PEW Research Center where they work. They may have a point about some particular surveys but Kuriakose and Robbins (forthcoming) is convincing that there is a widespread general fabrication issue.

⁹ Michael Robbins works for Arab Barometer and is driving their improvements forward ([Robbins \(2015\)](#)).

fabrication issue while building confidence in their data. At the other extreme, we must question the data quality of organizations that minimize the existence of the fabrication threat, particularly those like D3 that use the legal system to suppress the issue.

Self-censorship rears its ugly head once again in this discussion in the form of a general reluctance to name surveys that have been marred by fabricated data. Commendably, Arab Barometer has been open and forthright about its data quality problems and measures it is pursuing to address these problems. However, Kuriakose and Robbins seem to be moving from an early version of their paper that named the (publicly available) surveys they analysed to a published one that will withhold this information. The State Department is ignoring a FOIA for their D3/KA Iraq data and appears to have an official policy of never naming contractors who have delivered fabricated data. Such secrecy puts us in an awkward position of knowing that we have consumed a lot of fabricated data but not knowing which data are valid and which are not.

4. CONCLUSIONS

This is a simple story of corporations derailing legitimate academic research. It is hazardous to guess at how common such situations are but it seems reasonable to suppose that my experience is far from unique. Even if actual censorship threats are rare they can still have a chilling effect if researchers distort their agendas to steer clear of potential trouble. In other words, censorship leads to self-censorship.

I considered whether exposing the D3 threat will, itself, have a chilling effect on research but have decided that coming forward is the right thing to do. I suspect that most censorship threats are probably bluffs anyway and it may be a sad truth that researchers just need to learn how to handle them. More generally, truth is almost always better than secrets.

For faulty surveys that circulate only within the government it should be sufficient for some sanctions to be imposed on contractors and for corrections to be issued to decision makers who may have been influenced by faulty data. For such surveys in the public domain there should be public corrections to repair the historical record. Unfortunately, it appears that at the moment the redresses described in this paragraph occur only rarely.