Methodological Appendix

Beyond publicly available sources (e.g., congressional testimony, GAO reports, policy documents, and press reports), empirical material for this article comes from a series of interviews with policymakers, former policymakers, private sector managers, and other individuals involved in cybersecurity for critical infrastructure.[1] Because our use of source documents is not novel, we focus here on the interviews. The goal of this Appendix is to meet or exceed standards of transparency in qualitative work (*Comparative Politics Newsletter* Editorial Board 2016; Kapiszewski et al. 2015, Moravcsik 2014, Malterud 2001), in order to allow other scholars to judge the persuasiveness of our approach and, if they wish, to reproduce the study.

This Appendix first discusses our study design. It then addresses the mechanics of our interviews (e.g., how we used our interview guide and conducted the interviews). The third section discusses potential challenges to inference. The fourth section addresses ethical issues.

1. Study design

Role of the interviews

We conducted two types of interviews: (1) in-depth, semi-structured discussions about the key issues at stake with subjects deeply familiar with the issues, and (2) shorter conversations with peripherally involved individuals who could confirm specific details. Almost all of the interviews were of the first type. (See Annex 1, which contains descriptions of each interview and interviewee.)

In conducting and interpreting our interviews, we drew upon a review of relevant government documents and other published material, as well as our own experience with the homeland security and intelligence communities. These other sources provided us with sufficient background to formulate the proper questions of our interviewees, to engage efficiently with them on this policy domain, and to interrogate their responses when appropriate.

The focus of the interviews gradually evolved. The earliest conversations were used primarily to confirm our understanding of the landscape and to refine our research question. Later interviews were focused more explicitly on how well the quasi-voluntary regime operated in different sectors and identifying which factors explained cross-sector variation. Many of the later interviews, especially those that focused on the financial sector, were also used to link certain variables causally to changes in the effectiveness of the partnership within specific sectors and understanding the interactions among different explanatory variables – i.e., "process tracing" (Waldner 2015, Mahoney 2015). As is the case in most qualitative research, there was not a clear differentiation between the earlier and later interviews, and the "logic of discovery" (Curd 1980, Schickore 2018, Kapiszewski et al. 2015, Freedman 2008) was iterative. In other words, some interviews helped both to refine the research question and to provide empirical material for our conclusions about the state of play in a sector.

[1] We use "subject", "informant", "interviewee", "source", "respondent", and (reflecting a new trend in Psychology) "study participant" interchangeably to describe the individuals with whom we spoke.

Inappropriateness of survey research and large-N quantitative analysis

One natural question is why surveys could not be used to obtain some of the information we sought to collect – such as perceptions of how well information-sharing forums operated – in more standardized form. We hope that future work will successfully utilize surveys to follow up on some of our findings. However, we believed (and still believe) that closed-ended survey instruments would not have revealed the causal mechanisms at work as well as our in-depth conversations. In fact, we are even skeptical that surveys would have obtained useful factual information about the current state of play in the sectors we examined, for three reasons. First, as a practical matter, the number of individuals who are best able to compare systematically across sectors – i.e., the sources whose opinions are most important to capture – is rather limited, meaning that a survey might not allow us to draw statistically significant inferences (which is one of the main benefits of survey research). Second, many of these high-value sources are quite prominent and would be unlikely to respond to a standard survey. Third, closed-ended surveys would not permit the sort of interrogation of respondents' answers needed to elucidate what was actually happening in each sector.

Another approach would have been to develop "objective" quantitative indicators of variables like "capability", "degree of information sharing", "perceived threat to business model", "regulatory authorities", and the like, for the purposes of statistical analysis. Although we explored this approach, we ultimately found it unworkable.

For instance, one possible metric of capability might be the average rate of adoption of patches by firms in an industry once a vulnerability becomes known. However, the same patches do not apply (or apply equally) to all industries. We also considered attempting to collect data on "dwell time" for malware, number of intrusions, and response times. One fatal flaw in measures is that they presume that firms are equally successful in detecting intrusions in the first place, whereas our informants agreed that poor-performing organizations may fail to become aware of them in the first place.

Another metric might be rates of adoption of the Center for Internet Security's Control Standards (CIS 2019), which include 20 categories with approximately 170 specific items. For instance, Item 4.5 is "Use multi-factor authentication and encrypted channels for all administrative account access". Data on adoption of these metrics across industries – or even firms within an industry – do not yet exist. Furthermore, not all of the items are equally critical; a good deal of analysis would need to be done to determine whether two firms that both reported 90% compliance with these measures had done the same 90%. For instance, consider the debate over industry standards in the electricity sector:

> [P]roposed standards required the use of strong passwords and audit logs, a common practice in the office environment. But OT experts pointed out that legacy equipment did not always provide functionality for password protection or audit trails. Unlike office computers, OT was expected to run for decades; newer devices with stronger security measures did not necessarily exist, and some argued that they were not desirable. In an operational environment, where machinery could do physical harm, strong passwords posed the risk of an operator getting locked out at a dangerous time (Slayton and Clark-Ginsberg 2018: 121).

Equally important for our purposes, standard metrics represent relatively basic steps; they would not differentiate between sectors that were only practicing basic cyber hygiene and those that were engaged in continuous improvement. In a telling (albeit crude) example of what our sources thought of such measurements, one informant said: "Compliance does not mean security…When I was at the [name of an organization], we were constantly raping systems that were supposedly compliant."

Two other specific off-the-shelf measurements of security that we considered were the BitSight security ratings by sector (including the separate diligence performance metrics) and KPMG's 2018 Global Power & Utilities CEO Outlook (and related documentation) for that specific sector. We suspect that these sorts of measurements are useful for managers of private companies for understanding basic industry trends and where one's own company stands in relationship to other firms on some metrics. However, we do not believe that they would be good proxies for any of the variables of interest to us.

In the end, we did not even find these sorts of metrics to be useful in interrogating informants' answers in the course of our interviews. When we did occasionally bring them up, subjects would swat them away or go into detail about why they were not particularly useful for understanding how a sector was performing (thus costing us time in the interview). We concluded that attempts to utilize existing metrics in this way tended to derail the conversation and even to make us appear less familiar with cybersecurity than we actually were.

Replicability and reproducibility

For several reasons, which are common to all qualitative or ethnographic studies, our research is not replicable in the strict sense. Our data analysis (i.e., write up based on the material we collected from the interviews) cannot be replicated because, given the ground rules we offered informants, we cannot provide full transcripts of the interviews to other scholars – which would otherwise be the optimal approach. (Note, however, that we can shared lightly redacted notes from most of the interviewers with interested scholars on an individual basis, after establishing some basic rules of the road to prevent triangulation to the subjects' identities and some discussion about what material could not be included in the interview notes.) In addition, the interviews we conducted do not fully capture the material upon which we based our conclusions, which are also a product of our priors about "how the world works" in this specialized domain (Schatz 2009).

Nevertheless, while not replicable, both our data collection and our data analysis are at least partly *reproducible*. Other scholars could theoretically gain access to the same universe of interviewees, sample from among them, deploy our interview guide or some variant thereof, and draw conclusions from what the subjects revealed. Scholars should thus be able to conduct a study that is materially the same as the one we conducted, and – if our conclusions are correct – they will produce the same findings. If scholars reach different conclusions but are equally transparent about their method, they should be able to assess whether specific features of each study were responsible for the different results. (For instance, they may have asked different questions of similar people, spoken with different specific individuals within the larger potential sample, or interviewed a new category of people altogether.)

We hope that reproducibility is enhanced by:

- A description of how our interviews were conducted (Section 2);
- As detailed a description of the interviewees as confidentiality permits (Annex 1).
- A discussion of recruitment (below), including the theoretical universe of sources from which we sampled (Table A1 of the Annex); and
- A copy of our Interview Guide, annotated to show how specific blocks were designed to elicit certain types of information (Annex 2).

Sampling and recruitment

We limited our sample to people actively involved with industry-government partnerships (at either the operational or policymaking level). Our interviewees did *not* include self-styled cybersecurity "experts", pundits, journalists, authors of popular press books on cybersecurity, academics without direct knowledge of policymaking in this domain, or senior policymakers only indirectly involved in the specific issue of cybersecurity for critical infrastructure. It also did not include individuals who, though expert on the technical vulnerabilities of specific systems, lacked knowledge of government-industry interactions in this domain.

The initial selection of respondents was opportunistic, based on existing contacts. From these early contacts, we snowballed to individuals with whom we were not previously familiar, focusing on those whose experience seemed most directly relevant and who were recommended by more than one original interviewee. As noted above, we also conducted shorter interviews with peripherally related individuals who could confirm particular facts or fill in details on which the original interviewee's recollection was fuzzy; these individuals were selected for their direct, personal knowledge of specific bits of information. A small number of people are able to directly evaluate how well the quasi-voluntary partnership functions *across* many sectors. Our "sample" approached the universe of such individuals.

To impose some rigor on our sampling process, and to allow for reproduction of the study, we attempted to define the theoretical universe of interviewees. (See Table A1.) In general, as Table A1 shows, we covered a cross-section of potential informants from the relevant categories. After we had interviewed about half of those individuals whom we eventually contacted, we reviewed Table A1 and selected the remaining subjects based on where the biggest gaps in our knowledge lay.

One danger of snowball recruitment is the possibility that informants will come from homogenous networks and thus likely to share certain views. As noted above, we attempted to address the possibility of less egregious sampling bias by creating the theoretical universe of informants and deliberately contacting a new set of individuals who fit into each category toward the end of the study. We also went through the exercise of imagining an "inconvenience sample", as proposed by Duneier (2011). (We found no obvious such constituency once we had specified the theoretical universe of informants.)

The worst sampling biases in field research occur when the act of interviewing individuals from one faction precludes researchers from interviewing individuals from a different faction (Duneier 2011). Because interviewees did not generally know with which other subjects we had spoken, this issue was not a problem for us. In any case, we did not find much evidence of "tribal" boundaries in this domain. One likely reason is that many former government officials are now employed in industry (and vice versa), meaning that we did not find a consistent cleavage between public sector and private sector informants. We also did not find a clear partisan cleavage among interviewees, which we attribute to the relatively non-politicized context of cybersecurity and to the fact that Administrations from different parties have generally practiced similar policies. For instance, one source who had worked for senior officials in both the Trump and the Obama Administrations was far more complimentary about officials in the former, despite the fact that the individual in question was almost certainly a Democrat by partisan affiliation.

We received only two declinations from those whom we attempted to interview. Two were senior executives in the private sector who appeared to have been simply too busy to participate, and they were easily replaced by others who worked in the same industry. We were also unable to arrange an interview with one senior government official with whom we had hoped to speak in a timely manner but have no reason to believe his/ her opinion would be different from that of other informants. Consequently, non-response bias is unlikely to have been a significant problem for inference.

2. Mechanics of the interviews

After several exploratory interviews by the first author alone (designated as such in the list of sources provided below), we conducted a number of interviews with both authors present. The purpose of this approach was to make sure that we would have a better sense of how to obtain similar material when subsequently conducting interviews alone – for example, how we would balance fidelity to the interview guide with pursuit of promising new avenues suggested by subjects' responses to our questions (see "Interview guide and usage", below). After this initial set of conversations, most interviews were conducted by one author alone. Altogether, about a third of the interviews were conducted with both authors, half by the first author alone, and a sixth by the second author alone.

Preparation for interviews

We normally sent informants a precis of the project and brief background on ourselves ahead of the interview. Where requested, we also sent a summary of the relevant portions of the interview guide. In addition, interviews were often preceded by short conversations or email exchanges about the project.

In terms of our own preparation for the interviews, we attempted to review as much public information as was available about the subjects. Where appropriate, we also attempted to read any material they had written on cybersecurity or critical infrastructure. In one case, the informant was sufficiently prolific that we were only able to read a portion of what they had written on the subject in the time between when the interview was scheduled and when it was conducted. This fact caused us to lose about half the time in the interview covering the unread material (which we later read), but we were still able to extract a good deal of other information from the conversation.

Approach

Most initial interviews with sources lasted approximately an hour, the amount of time we normally requested. The shortest substantive interview was about half an hour; the longest was approximately three hours (65 minutes of formal conversation in an office followed by a lengthy tour of a government facility to describe how things worked in practice). We went back to many sources two or more times, either because the interview ran over the allotted time or because we sought to clarify specific points. These follow-up conversations varied in length from a few minutes to an hour, with the median being about half an hour.

In general, we presented the interviews as conversations. Given that our subjects were senior enough to be experienced with being interviewed by journalists, giving testimony before legislators, and the like, formal informed consent was replaced with an informal discussion of ground rules that would be familiar to all informants (as approved by our university's Institutional Review Board). In order to elicit the most candid answers possible, we offered study participants generous terms: their comments could be on the record, off the record, or on background; study participants could also designate certain specific comments as off the record at any time regardless of the larger context of the conversation, and they could change how they wished us to treat a comment *ex post*. They were also given the option of reviewing a transcript of their interview if they wished, and we offered to send them sections of the article in which they were mentioned (by moniker) before it was published to make sure they agreed with how their comments were used. We left it to them to determine whether they wished to be identified by name in the Acknowledgements.

Interview guide

In-depth interviews were based on a guide (Annex 1), which was modularized to reflect the fact that not all interviewees would have the same information to impart. Specifically, the guide focused on four main topics: (a) the rationale for and purpose of government action in cybersecurity policy, (b) the definition of "success" in this domain, (c) the empirical effectiveness of the current regime in achieving this success across sectors, and (d) the causal factors that account for variation in success across sectors. We updated the guide in minor ways throughout the project, particularly if it became apparent that interviewees were interpreting the same question differently from each other or from what we intended.

The relevant sections of the interview guide were deployed as thoroughly as possible in each interview, but we did not use the guide mechanically. Questions served as a launching point for discussion, and we always erred on the side of following up on interesting points rather than completing the guide. In other words, the interviews were semi-structured and conversational rather than formal and fully standardized (as would be the case in a survey). As discussed above, we believe that this method was the best way of eliciting the information we sought.

We avoided using terms that would effectively put words in our sources' mouths or be interpreted differently by different respondents. We also tried to eschew academic jargon (e.g., by saying "spillovers" or "effects on other sectors", instead of "externalities"). In general, we believe we succeeded. The one exception concerned the use of the term "regulation", which many informants interpreted in the Washington sense of the formal agency rule-making process rather than in the academic sense of "sustained and focused attempts to change the behavior of others in order to address a collective problem or attain an identified end or ends, usually through a combination of rules or norms and some means for their implementation and enforcement, which can be legal or non-legal" (Black 2008). As this confusion gradually became clear to us, we began prefacing questions about "regulation" by explaining our broader understanding of the term or by asking about "what the government might do" without using the term "regulation".

Rapport

We placed a great deal of emphasis on establishing rapport with the subjects. Where one or both interviewers already knew the subject, as in about a third of the cases, rapport was implicit or rapidly established, and the substance of the conversation could begin immediately. In other cases, establishing rapport involved discussions – sometimes jocular – of one another's personal background, professional experiences, shared acquaintances, and the like. This discussion was generally only 5-10 minutes, but on occasion it consumed as much as 20 minutes. Even where lengthy, we regarded these preliminaries as useful investments to establish rapport.

We generally mirrored the conversational style of our interviewees (see Leech 2002, Brinkman 2013, Weiss 1994). Where informants seemed to prefer erudite or formal language, we used that. Where they were casual (or even profane), we followed suit. In several of the interviews, the conversation was free-wheeling and energetic. In others, the cadence of the conversation was slower or more composed. Rarely, though, was the conversation stiff.

In our estimation, a high level of rapport was established in almost all interviews. None of the interviews was cut short, and many went longer than planned. (In fact, at least two interviewees rescheduled previously planned appointments on the fly in order to continue the conversation with us.) Almost all of the interviewees with whom we requested a second conversation agreed to be re-interviewed, and most were willing to provide introductions to other individuals previously unknown to us. All told, we were struck by their willingness to give generously of their time (including time that in many cases could have been billed out at very high hourly rates). Several interviewees spontaneously stated that they found the conversation stimulating, that it made them think about the issues differently, that they would like to "circle back" later in the project to see what we had found, or offered similar indications of engagement.

There were two partial exceptions to our success in establishing rapport with our study participants. The first was a senior official (Qtech), whom neither author had previously met in an interview of approximately 40 minutes conducted by telephone. The tone of the interview was collegial, growing more so over the course of the conversation, and we did not detect any lack of honesty; furthermore, the interview provided useful information. However, we perceived significant reserve, especially during the first half of the interview, which was manifested in a tendency to speak in generalities or abstractions even when pressed for specific examples. We also detected time pressure. We attributed the reserve to Qtech's status as a serving senior official, which may have evoked concerns about how candid to be on a phone call with unfamiliar interviewers. Another possibility is that Qtech may have feared (incorrectly) that providing too many specifics could have inadvertently taken the conversation into the classified realm.

The second partial exception concerned a former congressional staffer involved in legislative discussions about cybersecurity for critical infrastructure (Mica), who was interviewed by the second author. Mica was fully forthcoming during the conversation but reluctant to provide contact information for potential interviewees who might be able to provide additional detail or confirm specific details. We were left unsure about the source of this reluctance, as the conversation was collegial and Mica was happy to continue it beyond the originally scheduled one-hour block. One possibility is that some of the individuals for whom contact information was requested had gone on to work in the Intelligence Community, and providing contact information would therefore have been inappropriate or infeasible. Another possibility is that Mica's relationship with those individuals was no longer sufficiently close to provide an introduction. It is also possible, of course, that Mica was simply too busy to follow up or insufficiently interested in the project.

One final issue related to conversational style concerns status hierarchies. Field researchers often encounter situations in which they are perceived to have a higher status than their subjects (which can create ethical dilemmas) or a lower status than their subjects (which can lead them to not challenge an interviewee's claims). In most cases, we and our interviewees had roughly similar status. However, the first author was conscious of the fact that his protocol rank in government service was lower than that of some informants and initially felt obliged to acknowledge that fact in direct address. Fortunately, we detected no attempts by interviewees to "pull rank", nor resistance to being interviewed by someone who could be perceived as junior to them (even among the small number of military interviewees). In any case, we discussed this issue after the first few interviews and agreed that acknowledgment of rank was unnecessary in this setting; all subsequent interviews except those between the first author and serving members of the uniformed services were conducted on an egalitarian basis.

Follow-up questions

As noted, we deployed the guide flexibly, often using multiple follow-ups on specific questions. An example of a series of follow-up questions from one interview (reconstructed from short-hand notes and memory) runs:

> Second author: Has there been any economic analysis that would support claims of spillovers in the […] sector?
>
> Respondent: No. [Name of entity] did some modeling, which could be useful. But the inputs in their model were made up. They are just based on guesses. The value of that exercise was forcing people to write down their assumptions, but it wasn't real analysis.
>
> Second author: Anything else out there?
>
> Respondent: Not that I can think of.
>
> Second author: Would it be fair to say that, if there was something out there, you would have encountered it?

> Respondent: I think that's likely. It's not a hundred percent. There could be something I haven't seen, but nothing has come through …[description of why the respondent would have seen any such material]. I haven't taken a look at that question in the last few months. But there's a very good chance that if there was something out there I would know about it.

An example of a series of follow-ups (verbatim from the recorded transcript of the interview) conducted by the both authors is:

> First author: This relates to a point made in our previous conversation and in conversations we've had with others: this idea of sector-wide awareness and coherence. How does this relate to the kind of model you're describing?
>
> Respondent: Above all else, this is about getting the industry partners together. Maybe the government provides some cover for them to be able to engage each other, but almost as a silent partner. Not that there isn't value in partnership with the government, but to have sector-wide coherence, the industry partners need to come together to create the bathtub of information that they can make sense of together. In my experience, this is more likely to happen amongst each other. This builds not just a common awareness but a better understanding of what's happening when they're free to do this together [….]
>
> [After several minutes of discussion]
>
> First author: You made a distinction between the information people get and understanding the situation fully. Can you give me an example?
>
> Respondent: There's the basic level of sharing that we have done for a number of years where we might have basic indicators of compromise and we'll share them with companies. The companies will ingest that and figure out how it applies to them and address it individually. Then there is being able to identify some commonality among some of the indicators that indicate a very sophisticated approach at compromising a system. Until you thread those together to tell the story of what each of those indicators -- each of the individual activities -- actually mean in aggregate, you don't realize that what's happening is that the entire system is being compromised in a way that's much greater and more sophisticated than just a couple of data breaches [….]
>
> Second author: But the realization comes from the different firms sharing with each other, not from the government ingesting all this information and telling them what the commonalities are?
>
> Respondent: Exactly, because the government can't see inside their networks. So it doesn't know what's happening inside their networks, it just sees what's on the outside.

Interrogating informants' responses

We were rather assertive about challenging claims made by interviewees. Almost always we were able to present these interrogations as requests for additional information (e.g., "That's so interesting -- could you explain why it works that way?"). Where such probes were not sufficient, we would employ material from other interviews to push informants ("That's so interesting, because we have heard other people say something that seems different…"). Alternatively, we might rely on public documents for the same purpose ("So, does this mean that the GAO has it wrong? I mean, we're not surprised to hear that they do, but…"). We also used hypothetical scenarios ("So what would you say to someone who argued that…"). We avoided probes that were phrased as challenges from ourselves directly.

In most cases, we needed to use only one or two probes to properly interrogate the information we received. The longest exchange along these lines consisted of a half-dozen politely worded challenges from us and responses from the interviewee on the same question – specifically, what level of formal regulation would be required to ensure appropriate levels of private sector investment in cybersecurity in a particular sector with a private sector interviewee who strongly opposed regulatory approaches. This lengthy back-and-forth required some finesse but did not become contentious.

Handling deflections

One common challenge in field work occurs when an interviewee deflects a question or appears to be made uneasy by it. It is not always clear at the time whether an informant is deliberately deflecting or has simply misunderstood the question. In following up on possible deflections, we generally rephrased questions in a way that would presumably be less problematic for a reticent interviewee to answer. For instance, we might switch from the second person to the third person (e.g., from "Would you…" to "I realize this probably doesn't apply to you personally, but would a different person in your position ever…").

In general, we probed enough to make sure that a deflection was intentional and then pushed no further. The two main instances of deflection that we recall came (a) when Qtech was asked to provide specific examples, as discussed above, and (b) when a private sector interviewee was asked about details of his interactions with regulators. We later found out that punitive regulatory action – which had not been made public – was pending and assume that this fact accounted for the interviewee's reticence.

Interrupting

Another common challenge in field work is what to do when informants fail to provide succinct answers or go off on tangents. Among our interviewees, fortunately, this tendency was rare. On the occasions when it did happen, we were relatively assertive about politely redirecting the conversation. (For instance, we might interrupt by saying "Wait – sorry to interrupt but what you said was so interesting and I don't want to forget to ask…") Otherwise, we generally declined to interrupt subjects, preferring to let them finish delivering their considered opinion on an issue and then circling back later to points where we might otherwise have interrupted to clarify or follow up on something they said (as in the second example above).

Mode of interview

Interviews were conducted by a mix of in-person, videoconference, and telephone, depending on the interviewee's availability. Our preference was for in-person interviews, and each author traveled to Washington, D.C. (separately) to conduct some of the in-person interviews; however, scheduling conflicts sometimes made in-person interviews impractical. When the interview was conducted remotely, we advocated for videoconference. However, sometimes technical difficulties or other limitations (e.g., the interviewee wanted to talk while on a long drive or from an airport) forced us to conduct telephone interviews instead.

To our surprise, we found that the mode of interview was not consistently related to the level of rapport, duration of the interview, or fruitfulness of the conversation; two of the richest interviews (e.g., Thailand and Tico) were audio-only telephone conversations. We suspect that the effect of mode on the quality of an interview is mediated by (a) our prior acquaintance with the subject, and (b) the subject's personality. In other words, telephone interviews worked well with a naturally candid interviewee or one who was already known to one or both authors; however, telephone interviews were potentially problematic when the interviewee was reticent by disposition *and* was someone with whom we had no previous connection.

Transcription and reconstruction

Most interviews were recorded and transcribed using software; the output was then edited by a professional assistant and reviewed by us. When interviewees declined to be recorded or when we deemed that requesting a recorded interview would be problematic, shorthand notes were taken during the interview and subsequently typed up in long form as soon as possible afterward. In some cases, interviewees asked that recorders be turned off for part of the conversation; in two cases, an interviewee asked not only that the recorder not be used but also that the interviewer take no notes at all. In the transcript / reconstruction, any such periods were noted only as "[Off-the-record conversation on general topic of ___]" or simply as ["Off-the-record discussion"].

We often reorganized the notes topically after the fact rather than retain their original chronological format. Final notes from a typical interview ran to approximately 2,500 words; the shortest write-ups (Suriname, Brookline, and Ireland) were approximately 300 words, while the longest (Thailand) ran to 8,600 words. In total, the interviews produced approximately 120 single-spaced pages of notes.

Capture of contextual information

Final notes included discussion of the context and background for the interview, in case these might later prove to be correlated with the content. For instance a description of an interview by the first author with two senior government officials went as follows [with redactions to protect confidentiality in ellipses and brackets]:

> Location: the interview was conducted in person in XX's office. After the interview concluded, YY brought me… to their new office space to meet and chat with the rest of the team.
>
> Context: In previous conversations over the phone with XX and YY, it was clear that they were under increasing demands due to current events that limited the time available to spend on non-critical tasks. This time pressure was not reflected

> during the meeting. They were gracious with their time, spending a little more than the allotted hour, and remaining attentive and engaged throughout the conversation. Additionally, when I walked in, XX was finishing reviewing and editing an assessment of one of the elements of the current government-industry partnership in the… sector. It was a negative assessment, reflecting their current frustrations and challenges. The fact that this setback was fresh in mind for YY and XX may have been responsible for a general sense of pessimism during the conversations.
>
> Setup Discussion: I arrived a few minutes early and XX broke from his work early to chat about personal life developments since …[mention of previous connection to interviewer]. YY arrived on time, and I began by providing them an overview of the research project and its intentions. The overview generated immediate interest from YY, who wanted to jump into a conversation on what "good" looks like, which was the second question I had listed.

3. Challenges to inference

Informant credibility

One important consideration in field work is the credibility of informants. In general, we found our subjects to be candid, knowledgeable, and careful in clarifying the degree to which they had direct or only indirect knowledge of a situation. In drawing inferences about situations where some subjects were less familiar but still able to offer an opinion, we down-weighted their comments relative to those made by sources with direct knowledge, as described in footnotes in the text. (We provide basic information on informants in Annex 2, so that readers can judge whether we are relying excessively on reports from interviewees with only indirect knowledge or potential biases.) We were also prepared to down-weight the claims of subjects who seemed reticent or untrustworthy. However, in striking contrast to certain other projects in which the second author has been involved, none of our sources struck us as remotely deceptive.

One concern is that experts on cybersecurity may have incentive to exaggerate threats (*per* Masnick 2015). Indeed, there is a good deal of "hype" in popular press accounts and comments from pundits. A number of our informants did express significant concern about cybersecurity threats to critical industry in certain sectors. However, we did not find their tone strident or hyperbolic; rather they provided a measured sense of the situation that sometimes downplayed commonly discussed risks while calling attention to others that were less frequently discussed.

Interviewer-specific effects and inter-coder reliability

Because of the way interviews were divided between the two authors and subsequently discussed, it was not possible to produce formal measures of inter-coder reliability (i.e., to measure statistically whether we recorded or interpreted informants' answers the same way). Informally, however, our discussions after interviews which we conducted jointly revealed that we almost always "heard" the same thing from each source, assessed informants' candor similarly, and judged the degree of rapport we established in the same way. After each interview conducted jointly, we caucused about what the interviewee had said, discussed any unclear points, and agreed on which points (if any) required follow-up. For interviews conducted separately, we caucused more extensively, sometimes recapitulating the bulk of the conversation in order to determine what follow-up with the informant was indicated.

One indicator of whether our interviews produced reliable information is the degree to which study participants agreed on basic facts and inferences (such as characterizing sectors in more or less the same way) when interviewing was conducted by different authors. Although interviewees rated sectors differently on an absolute scale – for instance, some dwelt on the problems, whereas others emphasized the amount of progress – their relative rankings were similar regardless of who interviewed them.

Transparency in write up

In the text, we opted not to use hyperlinks recommended by Moravcik (2012), which remains a controversial approach even among scholars dedicated to transparency in qualitative research and has not been adopted as a standard in political science. However, we did employ a similar approach. First, we refer (by moniker) to the interviewee who provided each quote we present and the basis for each claim we make in the text. Second, as noted above, we provide brief sketches of interviewees (Annex 2), so that readers can assess their credibility and level of knowledge with respect to each specific claim. Third, wherever interviewees were asked the same question and provided different or contrary answers, we discuss in a footnote why we reached the conclusion that we did.

Reflexivity

Reflexivity is arguably not as essential in interview-based research as it is in more in-depth fieldwork (e.g., ethnography). Even with interview-based research, however, there is a danger that interviewers may ask questions in a way that tends to confirm their own priors. Interviewers may also simply misinterpret apparently straightforward statements from informants based on preconceived notions of the way the world works. Finally, there is the danger that researchers, in writing up the results of their interviews, may attend disproportionately to confirmatory comments or claims. Researchers' awareness of and conscious attempts to correct for their potential biases – in other words, their reflexivity – reduce the chance that these biases will adversely affect their conclusions. (See Hsiung 2008, Nicolson 2008, Koch and Harrington 1998.)

Throughout the field research (i.e., interviews) and the write-up, we considered the following sources of potential bias:

- *Legacy protection*: Although both authors have worked in government, neither was implicated in policy decisions on this domain. We thus had no particular concern for reputation or validation of our own prior policy decisions.
- *Partisan coloration*: The second author had served as a political appointee in the Obama Administration, whereas currently serving political appointees were naturally from the Trump Administration, and some informants had previously served in the George W. Bush Administration. Partisan differences could lead to two sources of bias: (1) the second author's inclination to denigrate the policies of Republican administrations and (2) declination by Trump Administration officials to be interviewed or to respond truthfully. Although such partisan differences could well matter in other policy domains (e.g., environmental regulation), we did not find them problematic here. None of the interviewees

brought up partisan considerations, and we found mention of prior government experience (even with a different Administration) conducive to rapport. One reason may be that there have been few substantive differences in the policies that were actually followed by the last four Administrations.

- *Priors about the effectiveness of the quasi-voluntary regime*: The first author began with the prior that more assertive government action was needed in this domain. The second author was equally inclined to believe that the government was doing either too much (i.e., excessively burdening the private sector with security mandates) or too little (allowing owner-operators of critical infrastructure in this domain to ignore vulnerabilities that created potentially catastrophic risks). This produced a fruitful dialogue and, in our impression, better prepared us to challenge interviewees' implicit assumptions about the role of government (regardless of the answers they gave).
- *Priors about specific sectors*: Based on casual conversations with policymakers before the study, we began with the suspicion that there was heterogeneity across the sectors. Based on scholarly and popular press accounts from the last decade about the electricity sector, we also began with a prior that this subsector was characterized by significant vulnerabilities. This impression was reinforced in one early interview. However, as our written output makes clear, we rapidly modified our assessment of how well the quasi-voluntary partnership works in that sector based on reports from other informants.
- *Priors about the topic as a whole:* Our strongest prior was that the central questions in our field work were of great importance. We did not view this as a liability or source of bias; to the contrary, we believe it conveyed seriousness of purpose to the study participants and made them more likely to engage with us. However, we inevitably found ourselves thinking through the normative implications of our findings – that is, considering what the "right" set of policies in this domain would actually be. We were cognizant of this tendency and attempted to make sure that it did not inadvertently influence the way we presented questions to interviewees or interrogated their answers.
- *Priors about the way the world works:* Our efforts at reflexivity were most useful in revealing several unstated assumptions about how interactions between industry and government were likely to work. These included: the potential for capture by business of regulatory agencies, the political influence of the oil and gas industry, the motivations of large financial sector companies, the possibility of collusion among large firms, etc. It also revealed several unstated assumptions we shared about how government agencies were likely to work: that military and Intelligence Community tendencies to see regulatory problems through the prism of national security, that bureaucratic politics may affect how different agencies approach cybersecurity, that certain agencies were more or less competent, etc. For instance, we discovered that we shared a belief that DHS entities were not likely to be particularly expert in this domain. We thus made sure to interrogate informant claims that accorded with this prior, rather than simply assume that they were correct because they accorded with our sense of the world. We also made sure to interrogate our private sector sources assertively whenever the spoke up against regulation and to ensure that they were using the term in the same way an academic would. (In fact, as noted above, they generally were not; most scholars have a broad view of what regulation means, while private sector representatives tended to imagine highly prescriptive, checklist-style regulation uninformed by industry knowledge).

As we finalized the manuscript, we went through the exercise suggested by Duneier (2011) of an "ethnographic trial" – essentially, an academic red-teaming exercise. To our surprise, we did not find flaws in our interpretation of informants' reports. One reason may be that we had emphasized reflexivity at earlier stages. Another reason may be that we went back to many of our sources more than once, so the chance that we misperceived their comments was lower. A third reason may be that we recorded many interviews and kept notes about the context in which the interviews occurred; therefore, notes during the write-up stage of the project, we were able to base our interpretations of what our sources meant not only on transcribed words but also tone of voice, inflection, and contextual factors.

4. Ethical considerations

Human subjects

The project was deemed minimal risk and thus exempted from full review by our Institutional Review Board (IRB). As noted above, informed consent was oral and informal rather than written and described in terms eminently comprehensible to our subjects.

To protect respondent confidentiality, files containing transcripts of the interviews were encrypted. Personal identifiers and other information about the context of the interview were then extracted from that document and stored in a separate encrypted document from the text of the interview. The only information in common between the two documents was the moniker we assigned the interviewee. All recordings of interviews were deleted after the preparation of the draft manuscript.

Other ethical issues

Compliance with IRB requirements is a necessary but not sufficient condition for ethical research; the ethnical obligations of field researchers frequently go well beyond what IRBs consider. For instance, field work frequently raises challenges related to beneficence, explicit or implicit deception, the expectations of subjects, the safety of non-subjects involved in field work, and the role responsibilities of academics (Kapiszewski et al. 2015, Dewalt and Dewalt 2011).

In this study, the central ethical issue we encountered was ensuring the confidentiality of our subjects, which we believe we were able to adequately address for most purposes by anonymizing responses and encrypting files. However, such measures protect sources only to the extent that the law allows (Palys and Lowman 2012), and subjects may misunderstand their risk as a result (van Maanen 1983, Palys and Lowman 2012). To this end, we have deleted any information that *we* believe could be to the detriment of subjects, even though that information was provided to us voluntarily by sophisticated subjects and even though some of that information might be of use in subsequent research.

One ethical issue that did not materialize concerns potential vulnerabilities we identified in the course of the study. Although we did become aware of certain general vulnerabilities, on only two occasions did we hear of a vulnerability specific enough to be of value to an adversary. Naturally, we left any mention of these vulnerabilities out of our write-up.

A related concern that did not materialize had to do with government or industry responses to potential threats. Although we learned of certain new initiatives in the course of the study, the basic contours of these initiatives had already been mentioned in publicly available documents. In our write-up, we kept our discussion of these initiatives at a general level.

Conflict of interest

We identified no financial conflicts of interest for ourselves or our informants. However, we note that COIs could well arise in a study like this one if the researchers have any interest in private sector consulting related to cybersecurity or seek future employment in the private sector. (This was not a concern in our case.)

Potentially classified material

Almost all interviews were conducted in unclassified settings, and naturally no classified material was discussed in these settings. The first author happened to conduct a handful of interviews in a SCIF (though classified material was not discussed). Any time information was mentioned in these settings that might potentially be classified, no notes were taken, nor was the conversation reconstructed after the fact.

The first author exchanged some emails with some study participants from a government email account, which can be used to send classified information. These emails were not forwarded to any non-government account used by either author, and the second author never saw these emails.

As legally required, all papers related to the project went through prepublication review by one or more government agencies before submission to journals.

Table A1: Theoretical universe of interviewees

Category	Sub-category	Theoretical example	Theoretical universe of individuals in sub-category
Current and former government officials knowledgeable about cybersecurity for critical infrastructure, mainly at the Undersecretary (U/S), Assistant Secretary (A/S), or Deputy Assistant Secretary (DAS or D/A) level, as well as their policy staff. Also includes White House staff and Cabinet officer staff.	DHS: NPPD / CISA[2]	Assistant Director, Cybersecurity Division, CISA	~10
	DHS: sector-specific agencies	D/A Administrator, Surface Transportation (TSA)[3]	~20
	Other sector-specific agencies	Director, Office of Critical Infrastructure Protection and Compliance Policy, Department of the Treasury	~50
	Intelligence Community	Various officials at the National Security Agency	~10
	Department of Defense	Commander, Cyber Command	~10
Current and former private sector executives in relevant sectors	Chief Information Security Officers in key sectors	Chief Information Security Officer, Goldman Sachs	Hundreds at larger firms across all industries; thousands at all firms
Congress (from the relevant committees)	Members	Sen. Joseph Lieberman	~6
	Staff	Legislative Director for Sen. Susan Collins	~12

[2] National Policy and Programs Directorate (NPPD), now known as the Cybersecurity and Infrastructure Security Agency (CISA).

[3] Transportation Security Administration.

Annex 1: Final interview guide for in-depth interviews

Research Question:
What causes variation in the success of government-industry partnerships and achievement of their aims in critical infrastructure cybersecurity?

Information Needs:

A. What is the rationale for purposive government action and engagement with industry on critical infrastructure cybersecurity?
B. How is success defined?
C. How effective is the government's efforts across critical infrastructure sectors? Three dynamics considered:
    a. Ability of the sector to defend against sector-wide cyber attacks
    b. Degree of collaboration between firms within the sector
    c. Degree of collaboration between the government and industry within the sector
D. What causal factors account for variation in the degree of success across sectors?

Sequence
Interviews begin with rapport building questions aimed at warming up the conversation and orienting the interviewee toward the subject. In order to avoid narrowing interviewee thoughts early onto any particular hypothesis, the interviews next transition to broad questions to draw out interviewee's sense of how well things are working and what "good" looks like. These questions also start to specify the dependent variables, intending to draw out interviewee thoughts on what influences each (identifying new hypotheses and independent variables). Following this, the interviews progress into questions regarding independent variables. These aim to both measure and discern how each is working to influence the dependent variables. Finally, the interview closes with a wrap-up, a request to remain engaged during the study, a request for introductions to others who might be interested and useful to talk with, and a thank you.

Questions

1. Could you describe the relationship between government and industry within [SECTOR X] on cybersecurity? What does it look like and how does it operate? (warm-up orienting question that may being to reveal influential components and interactions)
2. What are the government and industry each trying to address through their partnership? What gaps are being addressed? (Information need A: Rationale)
3. What does success look like? What would you say the sector is pushing toward in terms of the relationship between government and industry in addressing cybersecurity in [SECTOR X]? (Information need B: Defining success)
    a. How would you know when it has been achieved?
4. How well equipped is the sector to defend against cyber attacks across the sector? (Information need A: Effectiveness)
    a. How would you rate the sector overall on a scale of 1 to 10? Why?
    b. How would you rate, from 1 to 10 again, each of the sub-sectors? Why?
        i. How do you explain the differences between sub-sectors? (Information need D: Causal Factors)
    c. [If time] How would you rate the top firms within the sector? Why?

      i. How do you explain the differences between firms? (Information need D: Causal Factors)

5. How would you describe the degree of collaboration between firms in [SECTOR X]? (Information need C: Effectiveness)
   a. How would you rate the sector overall on a scale of 1 to 10? Why?
   b. How would you rate, from 1 to 10 again, each of the sub-sectors? Why?
      i. How do you explain the differences between sub-sectors? (Information need D: Causal Factors)
   c. [If time] How would you rate this amongst the top firms within the sector? Why?
      i. How do you explain the differences between firms? (Information need D: Causal Factors)
6. How would you describe the degree of collaboration between the government and firms in [SECTOR X]? (Information need C: Effectiveness)
   a. How would you rate the sector overall on a scale of 1 to 10? Why?
   b. How would you rate, from 1 to 10 again, each of the sub-sectors? Why?
      i. How do you explain the differences between sub-sectors? (Information need D: Causal Factors)
   c. [If time] How would you rate this among the top firms within the sector?
      i. How do you explain the differences between firms? (Information need D: Causal Factors)
7. What are some examples of successes or steps toward success? (Information need D: Causal Factors)
   a. Can you walk me through how these came to fruition?
8. What is missing from this conversation? (catchall question offering an opportunity for interviewees to identify important factors and ideas that the interview questions may not be able to capture well)
9. [If time, ask for thoughts on already identified independent variables that were not mentioned in this conversation] (Information need D: Causal Factors)

Wrap-up

- What else do you think we should know or be considering?
- Are you interested in seeing my notes after I’ve written them up?
- Are you interested in seeing the results of this research?
- Who else do you think I should engage on this? Are you able to send an introductory email?
- Would you mind if I circled back for advice or if I find anything I thought you might be interested in?
- Thank you, very grateful for your time and thoughts.

Appendix references

Barry, CA., Britten, N., Barbar, N., Bradley, C. & Stevenson, F. (1999). "Using reflexivity to optimize teamwork in qualitative research." Qualitative Health Research, 9(1): 26-44.

Black, Julia, 2008. “Constructing and contesting legitimacy and accountability in polycentric regulatory regimes”, *Regulation and Governance* 2 (2): 137-164.

Brinkman, Svend. 2013. *Qualitative Interviewing*. Oxford: Oxford University Press, pp. 28-30.

Comparative Politics Newsletter Editorial Board. 2016. “Guidelines for Data Access and Research Transparency in Qualitative Research in Political Science,” *Comparative Politics Newsletter* 26 (1): 13-21.

Curd M.V. 1980. “The Logic of Discovery: An Analysis of Three Approaches”, in T. Nickles, ed., *Scientific Discovery, Logic, and Rationality*. Boston Studies in the Philosophy of Science, vol 56. Springer, Dordrecht.

Duneier, Mitchell. 2011. “How Not to Lie with Ethnography,” *Sociological Methodology,* 41: 1-11.

Dewalt, Kathleen and Billie R. Dewalt. 2011. *Participant Observation: A Guide for Fieldworkers.* Walnut Creek, CA: AltaMira Press. Chapter 11.

Hsiung, Ping-Chun. 2008. “Teaching Reflexivity in Qualitative Interviewing.” *Teaching Sociology* 36 (3): 211-26.

Kapiszewski, Diana, Lauren M. MacLean, and Benjamin L. Read. 2015. *Field Research in Political Science: Practices and Principles*. Cambridge University Press.

Koch, T. & Harrington, A. 1998. “Reconceptualizing rigour: The case for reflexivity”, *Journal of Advanced Nursing*, 28(4): 882-890.

Leech, Beth L. 2002. “Asking Questions: Techniques for Semistructured Interviews.” *PS: Political Science and Politics*, 35(4): 665-668.

Mahoney, James. 2015. “Process Tracing and Historical Explanation”, *Security Studies*, 24:2, 200-218.

Malterud, K. 2001. “Qualitative research: Standards, challenges and guidelines”, *The Lancet*, 358: pp. 483-488.

Masnick, Mike. 2015. “Ted Koppel Writes Entire Book About How Hackers Will Take Down Our Electric Grid... And Never Spoke To Any Experts”, *techdirt*. November 19. Accessed at https://www.techdirt.com/articles/20151117/07350332835/ted-koppel-writes-entire-book-about-how-hackers-will-take-down-our-electric-grid-never-spoke-to-any-experts.shtml, September 18, 2019.

Moravcsik, Andrew. 2014. “Transparency: The Revolution in Qualitative Research,” *PS: Political Science and Politics* (January).

Moravcsik, Andrew. 2012. “Active Citation and Qualitative Methods,” *Qualitative and Multi-Method Research* (Spring).

Palys, Ted and John Lowman. 2012. “Defending Research Confidentiality ‘To the Extent the Law Allows:’ Lessons from the Boston College Subpoenas,” *Journal of Academic Ethics* 10 (4): 271-297.

Nicolson, Paula. 2008. “Reflexivity, ‘Bias’ and the in-Depth Interview: Developing Shared Meanings” in Linda Finlay and Brendan Gough, eds., *Reflexivity: A Practical Guide for Researchers in Health and Social Sciences*, Chapter 10, pp.133-45.

Schatz, Edward, ed. 2009. *Political Ethnography: What Immersion Contributes to the Study of Power*. University of Chicago Press.

Schickore, Jutta, "Scientific Discovery", *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta, ed.

Slayton, R. and A. Clark-Ginsberg. 2018. “Beyond regulatory capture: Coproducing expertise for critical infrastructure protection” *Regulation & Governance*, 12: 115-130.

van Maanen, John. 1983. “On the Ethics of Fieldwork.” In *An Introduction to Social Research: A Handbook of Social Science Methods, Volume I*. Robert B. Smith, ed. Cambridge, MA: Ballinger Publishing Company. pp. 227-252.

Waldner, David. 2015. “Process Tracing and Qualitative Causal Inference”, *Security Studies*, 24:2, 239-250.

Weiss, Robert S. 1994. *Learning from Strangers: The Art and Method of Qualitative Interview Studies*. New York: The Free Press.