A New Model for Industry–Academic Partnerships

Gary King, *Harvard University*
Nathaniel Persily, *Stanford University*

ABSTRACT   The mission of the social sciences is to understand and ameliorate society's greatest challenges. The data held by private companies, collected for different purposes, hold vast potential to further this mission. Yet, because of consumer privacy, trade secrets, proprietary content, and political sensitivities, these datasets are often inaccessible to scholars. We propose a novel organizational model to address these problems. We also report on the first partnership under this model, to study the incendiary issues surrounding the impact of social media on elections and democracy: Facebook provides (privacy-preserving) data access; eight ideologically and substantively diverse charitable foundations provide initial funding; an organization of academics we created, Social Science One, leads the project; and the Institute for Quantitative Social Science at Harvard and the Social Science Research Council provide logistical help.

To deliver and improve their popular products, modern internet technology firms collect enormous quantities of data about human behavior. This data enables companies to improve decisions by building on research and methods from the social sciences and other fields. However, the same data holds the potential to advance social scientific discovery, generate social good, and provide insight for understanding and ameliorating important problems afflicting human societies. Successful industry–academic collaboration can thus simultaneously advance social science, benefit society, and help industry.

However, progress in data sharing for social good will occur only if all incentives are aligned—if individual privacy is protected, company trade secrets and related proprietary information are respected, and the standards and independence of the scientific process are secured. Although "academics provide the creative fuel for much early-stage research that leads to industrial innovation," achieving all these goals simultaneously is difficult in any area (Jasny et al. 2017), especially for the novel data types collected by internet technology firms.

We propose a new model of industry–academic partnerships designed to span this divide among the needs of industry, academia, and the public, even in the most highly politicized environments. This article describes the problem, our proposed two-part organizational innovation, and a specific implementation we developed for Facebook data. The appendices provide details about the incentives of each actor to participate and ethical process standards that we developed.

THE PROBLEM

For most of their history, the academic social sciences collected or purchased their own data and so have had only occasional need for relationships with industry. When they had a need, they often followed the traditional model in the natural and physical sciences, in which private firms donate funds, data, or expertise to a university lab for a specific research program—often in return for considerations such as the right of first refusal to license any resulting patents, along with prepublication review (but not prepublication approval) (Corzo and Eastman 2015). Although this model does not always solve problems for political scientists (Voss, Gelman, and King 1995), it can work when a company funds or details a few employees to work at a university lab for a specified period. The academic researchers then operate independently and have unfettered control over their research agenda, methodological choices, and publication options. This time-honored partnership model has generated enormous value for academic researchers, private firms, the scientific community, and society at large.

Unfortunately, the difficulties of collaboration with academia in the era of big data about human behavior means that this traditional model often fails. Not long ago, much social science

Gary King ⓘ *is Albert J. Weatherhead III University Professor and director of the Institute for Quantitative Social Science at Harvard University; GaryKing.org. He may be reached at king@harvard.edu.*

Nathaniel Persily *is James B. McClatchy Professor of Law at Stanford Law School and codirector of the Stanford Cyber Policy Center; Persily.com. He may be reached at npersily@stanford.edu.*

research could be completed without industry, because most data was created by academics or accessible from governments and firms making data public. Today, big data collected by firms about individuals and human societies is more informative than ever, which means it has increasing scientific value but also more potential to violate individual privacy or help a firm's competitors. Although many types of social science research require industry partnership even to begin a study, firms are understandably more leery than ever about sharing data. In other words, social scientists have access to more data than ever before to study human society, but a far smaller proportion of existing data than at any time in history.

products that affect millions of people, but they can only work on projects the company deems a high priority. As a result, their work is not necessarily aligned with questions of the greatest interest to the scholarly community or society at large. And they cannot publish without constraint.

The dilemma is thus stark: academic independence without necessary data generates little value. Full information with dependence on the firm for prepublication approval poses severe conflicts of interest and selection bias. And the many ad hoc approaches around these problems can be extremely difficult for individual scholars to negotiate, especially on highly charged politicized issues where firms justifiably feel the need to keep

> *Although many types of social science research require industry partnership even to begin a study, firms are understandably more leery than ever about sharing data. In other words, social scientists have access to more data than ever before to study human society, but a far smaller proportion of existing data than at any time in history.*

The problem we must overcome is that neither of the two logical antipodean approaches works to solve the problem—especially for high-profile, highly politicized, sensitive issues. In the first approach, academic researchers would remain fully independent, without prepublication approval by the firm. Unfortunately, even if a large technology firm were willing to share data with researchers and individual privacy could be assured, proper inferences require the full chain of evidence from the world to the data—in this case, often requiring proprietary information about the firm's policies, practices, procedures, platforms, and sometimes even access to its computer systems. Published scholarly articles accessing industry data in this way—without legal agreements requiring nondisclosure and prepublication approval—are rare, even though some clever ad hoc compromises have been productive (Chen et al. 2017).

In the second approach, academic researchers sometimes go inside companies and become consultants. They sign nondisclosure agreements and obtain all data, information, and systems necessary to do their research. However, in addition to the well-known effects of financial and other conflicts of interest (Banaji and Greenwald 2013; Koehler 1993; Wilson and Brekke 1994), consultants have limited freedom to publish, usually requiring firm preapproval, which greatly limits the value of any resulting scholarship to the academic community. In practice, the effect of these restrictions on research independence often depends on how close the research question is to the firm's products. Although consultant contributions to the firm can be large, may be of the highest scientific quality, and sometimes study subjects sufficiently far from the firm's core products to make publication easy, their contributions to the scientific community become complicated because the firm has more ability to censor scholarly work. We thus seek a better solution, especially for firms with valuable, informative, and sensitive data.

Other models of industry–academic relations exist (Ankrah and Al-Tabbaa 2015; Perkman et al. 2013). For example, social scientists often leave the academy to conduct research in industry or work in collaboration with data scientists within firms. These researchers then have more access to the data and influence on

their heads down to avoid incoming fire from their customers, the public, advocacy groups, and government regulators. The challenge with all existing models is that they do not scale as well as they might and data remains only occasionally accessible, and only by privileged researchers, with scientific progress suffering as a result.

A PROPOSED TWO-PART SOLUTION

In this section, we first describe the core idea of our two-part model. Then we show how to extend the idea to allow for research into the most difficult, highly charged partisan or otherwise sensitive topic areas, for when external funding is available, and in other ways.

The Solution

The central problem we seek to solve is that firms are not willing to give any one researcher both (1) complete access to data and other relevant information—including people, processes, policies, platforms, and business strategies—necessary to do research; and (2) freedom to publish without prior approval by the firm. We thus have a standoff: employees and consultants have (1) but not (2); outside academics have (2) but not (1). Our main proposal, using ideas analogous to the literature on constitutional design (Lijphart 2004), is to span this divide with an organizational structure that gives one group of academics (1) and a different group (2), with specific types of communication allowed between the two groups. This way, the academic community can trust the resulting scholarship and the firm can ensure that its trade secrets are protected.

In our structure, the first group is composed of *independent academics* who apply to study specific topics, are awarded data access, and have freedom to publish without firm approval. The second group, serving as a *trusted third party*, includes senior, distinguished academics who sign nondisclosure agreements with the firm, forego the right to publish on the basis of the data in return for complete access to the data and all other necessary information from the firm. This trusted third party thus provides a public service by certifying to the academic community the legitimacy of

any data provided (or reporting publicly that the firm has reneged on its agreement). It has authority to make final decisions on all proposals from the outside academics so it can prevent the type of topics, such as ongoing secret litigation or firm trade secrets, that would put the academics or their projects at risk or lead the firm to withdraw from the agreement. In this way, the trusted third party can encourage research on topics important to new and not-yet-public developments in the industry.

(see SocialScience.One). In this organization, which we cochair, scholars make all decisions. Social Science One is being incubated at the Institute for Quantitative Social Science (IQSS) at Harvard University and is staffed by IQSS employees.

The remaining discussion in this section follows the outline in figure 1. Our process for new partnerships begins with a company agreeing to a general topic area of research (e.g., the impact of social media on elections and democracy). Then, the trusted

> *However, in some situations—such as politically sensitive topic areas, companies under attack by the media or others, or when ideologically diverse funders contribute to the project—building in additional protections may be helpful.*

Extensions to Politically Sensitive Topic Areas and Available Funding

Our core two-part solution is relatively simple and can be implemented easily in many situations without further complication. However, in some situations—such as politically sensitive topic areas, companies under attack by the media or others, and when ideologically diverse funders contribute to the project—building in additional protections may be helpful.

To implement our proposal, to negotiate new industry–academic partnerships, and to lead the individual collaborations that result, we formed an organization called Social Science One

third party is a commission at Social Science One composed of respected scholars, who receive access to the company's operations and necessary data under confidentiality agreements. The commission receives answers to all relevant questions about any company platform, product, policy, practice, systems, or data that can help achieve its goals. The commission works closely with the firm's data scientists who, in any organization, typically have a wealth of information not available in any formal documentation.

For legal and privacy reasons, not all of the information shared with the commission at Social Science One is made public. This is the innovation underlying the two-part structure:

Figure 1
Outline of Industry–Academic Partnership Model
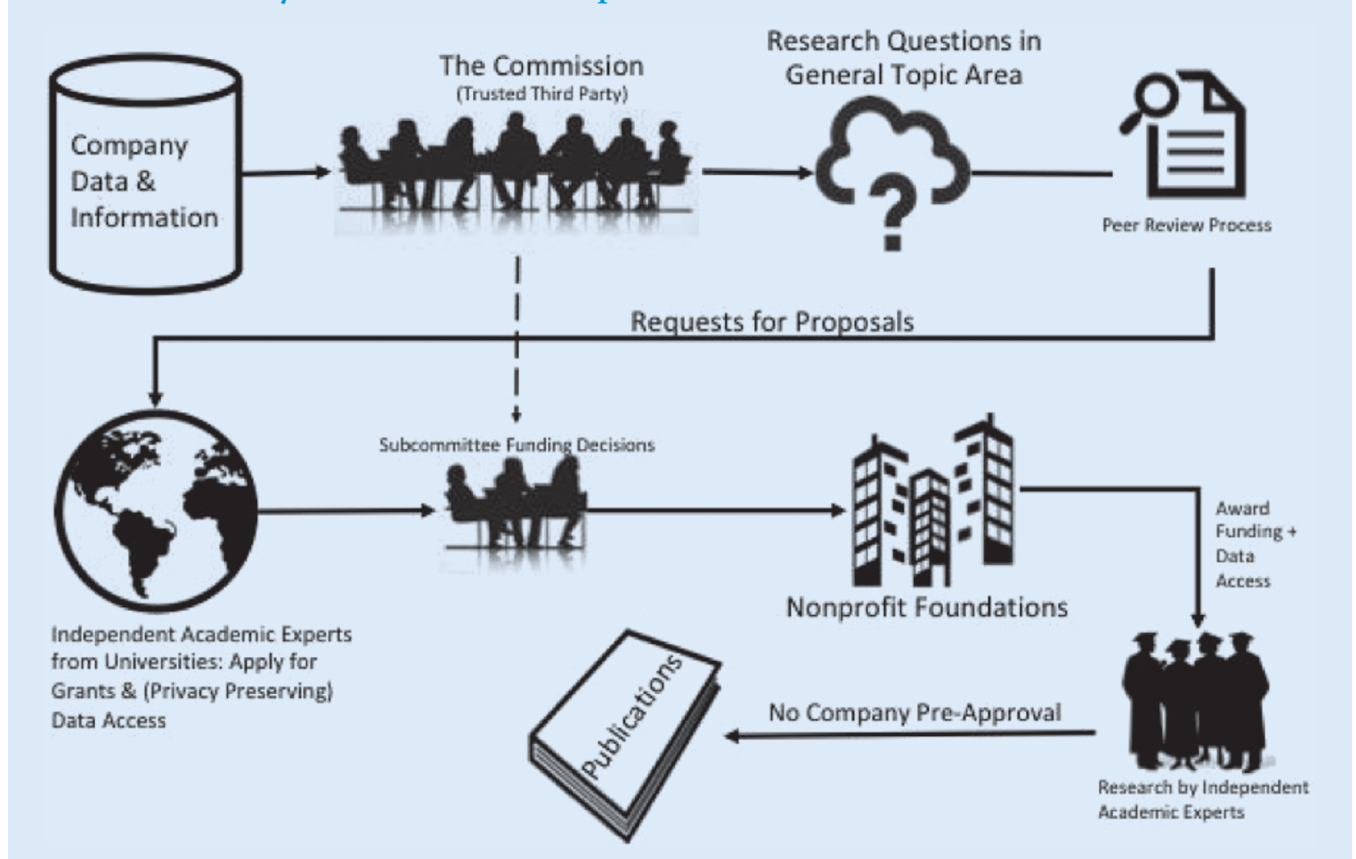
the commission has access to all information needed to make informed recommendations, and it filters this information to the broader research community on its own accord, omitting proprietary and other specifically delineated information, following agreed upon rules.

The process works only if Social Science One has the trust of the broader academic community and general public. Its commission is therefore designed to be composed of well-known, respected, distinguished senior scholars who represent the scientific community across important dimensions of demographic, regional, ideological, substantive, and methodological diversity. Like editors-in-chief at a scholarly journal, ultimate responsibility for Social Science One falls to its cochairs, with input from its members, the peer review process, and others. Commission members act as advisers in their areas of expertise, like an editorial board. Members of the commission, organized into committees, stay in contact with their academic communities by soliciting suggestions and ideas and by helping to publicize commission activities.

To limit security risks and the burdens of service, only certain commission members in what we call the "core group" sign confidentiality agreements, when necessary to their area of expertise. Because these core group commission members are treated as firm insiders, and given full access to relevant sensitive information, they give up the right to publish on the basis of the data access provided (at least without prepublication approval), and are prohibited from responding to requests for proposals or receiving funding as described below (which is another reason why only senior scholars are recruited to participate in this capacity). Other commission members may apply for grants as part of commission activities, as our peer review process includes procedures to manage conflicts of interest, and participate as peer reviewers.

The cochairs of Social Science One are initially appointed by the firm and the nonprofit foundations. The cochairs appoint other members with input from, but no decision-making authority by, the firm and foundations. The core group members, who sign restrictive confidentiality agreements, are compensated at fair market rates and, in highly charged partisan or otherwise sensitive environments, are paid by nonprofit foundations independent of the firm.

After the commission gets up to speed on internal data systems, policies, platforms, and practices at the company, it identifies a series of research questions, each of which appear to be answerable with access to a specific subset of the firm's data and systems, or with a scientifically appropriate and ethically designed data collection procedure such as a randomized experiment. If answers to these questions can be ascertained from research on the platform, and no legal or other agreed upon barriers to the research exist, then Social Science One follows standard academic procedures and announces an open grant competition for independent academic experts to receive data access to take on this work. This competition includes formal, public requests for proposals, peer review, and "revise and resubmit" processes.

The peer review process reflects the model's two-part organizational structure. At the start, we follow procedures similar to that which the US National Science Foundation uses for peer review, with multiple independent reviews by scholars and then panels of academics discussing each proposal and its reviews. Because the commission has sensitive confidential information about areas researchers are not permitted to pursue, such as due to ongoing litigation, trade secrets, and new developments in social media, it makes final decisions about which researchers receive data access. The firm and the charitable foundations have no role in choosing reviewers or making funding decisions.

When grants are awarded, the independent academic experts receive funding through standard university procedures for sponsored research and data access from the company. Although the overall process is complicated for Social Science One, the process for researchers is familiar and far simpler than most other industry–academic partnerships: they simply apply for and receive a grant from a nonprofit foundation via transparent processes. Even the formal "data use agreement" signed by the company and researcher's university is pre-negotiated by Social Science One so researchers do not need to begin this often arduous process anew each time and should have an easier job getting their institutions' approval.

We developed our model of industry–academic partnerships in the context of building a partnership with Facebook in the highly charged partisan atmosphere surrounding the issue of foreign influence through social media in the 2016 US presidential elections and the UK Brexit referendum—a few days after the Cambridge Analytica scandal.

A key aspect of this idea is that Social Science One has the obligation to report to the public—without permission from the company—about whether the company is keeping its end of the bargain, providing the commission full access, and answering all relevant questions. To be specific, if the commission concludes that the company has violated its agreement and prevented it from providing any piece of information it needs to address the general topic area, it will report this in a visible public statement. Social Science One has a responsibility to regularly report on its and the firm's activities, including decision-making criteria guiding the research agenda, scholar selection, and overall progress.

APPLICATION TO FACEBOOK

We developed our model of industry–academic partnerships in the context of building a partnership with Facebook in the highly charged partisan atmosphere surrounding the issue of foreign influence through social media in the 2016 US presidential elections and the UK Brexit referendum—a few days after the Cambridge Analytica scandal.

We also benefited considerably in getting the project off the ground from the extraordinary participation of a large number of high profile, ideologically and substantively diverse nonprofit foundations. Having their endorsement, guidance, and funding—and

just the fact that they are working together with singular purpose to make this project a success—eases considerably the difficulty of forging a partnership in the highly politicized domain that begins our project. We are grateful for and proud of our collaboration with the John and Laura Arnold Foundation, the Children's Investment Fund Foundation, the Democracy Fund, the William and Flora Hewlett Foundation, the John S. and James L. Knight Foundation, the Charles Koch Foundation, the Omidyar Network, and the Alfred P. Sloan Foundation. Because major companies have considerably more money than nonprofit foundations, at some point we may have funds flow from the company into an independent trust, which will redirect the funds to grantees. Either way, the company will remain disconnected from all decisions about which researchers will receive data access to the independent academic experts.

These eight foundations, which have never before joined together to support a single project, initially fund all research approved by Social Science One, thereby removing undue financial influence from Facebook or other private firms. Although those on one side of the ideological and partisan divide may not trust those on the other, we hope everyone recognizes the advantages of having all these foundations participate together.

Although we recommend omitting these steps in other less partisan topic areas, the foundations have agreed to the following additional processes for this area:

1. The foundations have no formal role in approving grants.
2. Funds from any one foundation (or viewpoint) are not used in isolation to support any one study.
3. All funds are pooled so that researchers and their universities receive grants from a single funding pool (with the Social Science Research Council [SSRC] contracted to serve as our fiscal agent and to help us with peer review logistics). Researchers report receiving funds and data access from Social Science One through SSRC, not from any one foundation.
4. All of this project's activities, including grant giving, decision making, committee appointments, etc., are supported by the collective views of the foundations so no one can dominate on any issue. The foundations advise, peer reviewers recommend, and the academics on the Social Science One commission decide with full academic freedom.

We are grateful to the foundations for their generosity and for supporting these principles.

By mutual agreement among the foundations, Facebook, and Social Science One, the general topic area for our project with Facebook is research on the implications of social media and digital technologies in the world—starting with democracy and elections. Here is Facebook's description:

> The focus will be entirely prospective, with the goals of understanding Facebook's impact on upcoming elections…and informing future decisions. The research sponsored by this effort will have benefits both for our understanding of social media's effects on democracy and for Facebook to better understand whether it has the right systems in place (i.e., Are we effectively able to fight the spread of misinformation and foreign interference?). Specific topics may include misinformation; polarizing content; promoting freedom of expression and association; protecting domestic elections from foreign interference; and civic engagement.

Although this is a broad scope for research, many other topics are of interest to the social scientific community and to the public. We are working to expand to those topics as well as others.

Finally, SSRC and Social Science One have established collaborations with some researchers at the Pervasive Data Ethics for Computational Research group (see pervade.umd.edu), a six-institution, NSF-funded research project on data ethics. They assist in conducting research on our processes and reporting progress so we can continually improve ethical compliance. (We subsidize costs of applicants from developing countries who must obtain certification from an external Institutional Review Board to ensure Common Law compliance.) Appendix B discusses these and other ethical precautions.

The datasets that we are in the process of making available are highly informative and immense, with trillions of numbers using petabytes of storage. Up-to-date progress reports are available at SocialScience.One.

CONCLUDING REMARKS

One might reasonably wonder whether now is, in fact, the time to discuss a data sharing program between internet companies and academics. Concerns about privacy are rightly at the forefront of everyone's mind in the wake of recent revelations. After all, the notorious Cambridge Analytica scandal began with a breach by an academic (acting as a developer) of a developer's agreement with Facebook, which barred his sale of the data to a for-profit company. That public scandal is a scandal for academia as well.

Yet, for this reason, we believe that now is precisely the time to have this conversation and to invent and implement structures that protect users' privacy while allowing independent academic analyses. If we do not develop these institutions, only the firms will have access to the data on some of society's most pressing challenges or will determine which academic results the public can see. Absent such an effort, many academics, commentators, and government regulators will continue to distrust the companies' representations that they understand the extent of the problems, have conveyed them accurately, or have implemented adequate solutions. With this new approach, however, we can take a critical step toward independent analyses of the dynamics of social media's effect on society—which will have downstream benefits for both the general public and the firms—and we can begin to tackle numerous other societal problems that these datasets can also address. We hope the specific features of our model will be helpful with other firms too, but the fact that new organizational structures can sometimes be developed to solve political problems technologically is our main message.

For any implementation of our model of industry–academic partnerships, achieving these difficult objectives requires delicate and often extensive negotiations throughout the process of structural organization, question definition, empirical research, and eventual publication. However, the questions are too important, the potential advances are too large, the range of knowledge that could be learned is too significant, and the information about the greatest challenges of society are too valuable to miss the advances that the scientific community can bring to the table.

REFERENCES

Altman, Micah, and Gary King. 2007. "A Proposed Standard for the Scholarly Citation of Quantitative Data." *D-Lib Magazine*, 13. Available at http://j.mp/2ovSzoT.

Ankrah, Samuel, and Omar Al-Tabbaa. 2015. "Universities–Industry Collaboration: A Systematic Review." *Scandinavian Journal of Management* 31 (3): 387–408.

Banaji, Mahzarin R., and Anthony G. Greenwald. 2013. *Blindspot: Hidden Biases of Good People.* New York: Delacorte Press.

Chen, M. Keith, Judith A. Chevalier, Peter E. Rossi, and Emily Oehlsen. 2017. *The Value of Flexible Work: Evidence from Uber Drivers.* National Bureau of Economic Research, No. W23296. Available at http://bit.ly/2HjgqPq.

Corzo, Janet, and Perkins Eastman. 2015. "How Academic Institutions Partner with Private Industry." *R&D*, April 20. Available at http://bit.ly/IndAc.

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. "Calibrating Noise to Sensitivity in Private Data Analysis." 2006. In *Third Theory of Cryptography Conference*, 265–84. New York, March 4–7.

Gaboardi, Marco, James Honaker, Gary King, Kobbi Nissim, Jonathan Ullman, and Salil Vadhan. 2016. Working Paper: "PSI (Ψ): A Private Data-Sharing Interface." Available at http://j.mp/2oSQ7F6.

Jasny, B. R., Nicholas Wigginton, Marcia McNutt, Tania Bubela, S. Buck, Robert Cook-Deegan, T. Gardner, et al. 2017. "Fostering Reproducibility in Industry–Academia Research." *Science* 357 (6353): 759–61. Available at http://bit.ly/2Hezzlx.

King, Gary. 1995. "Replication, Replication." *PS: Political Science & Politics* 28:444–52. Available at http://j.mp/2oSOXJL.

Koehler, Jonathan J. 1993. "The Influence of Prior Beliefs on Scientific Judgments of Evidence Quality." *Organizational Behavior and Human Decision Processes* 56 (1): 28–55.

Lijphart, Arend. 2004. "Constitutional Design for Divided Societies." *Journal of Democracy* 15 (2): 96–109.

Perkmann, Markus, Valentina Tartari, Maureen McKelvey, Erkko Autio, Anders Broström, Pablo D'Este, Riccardo Fini, et al. 2013. "Academic Engagement and Commercialisation: A Review of the Literature on University–Industry Relations." *Research Policy* 42 (2): 423–42.

Unger, Joseph M., Elise Cook, Eric Tai, and Archie Bleyer. 2016. "Role of Clinical Trial Participation in Cancer Research: Barriers, Evidence, and Strategies." In *American Society of Clinical Oncology Educational Book*, American Society of Clinical Oncology Meeting, Vol. 35, 185–98. NIH Public Access. Available at http://bit.ly/childcanc.

Voss, D. Stephen, Andrew Gelman, and Gary King. 1995. "Pre-Election Survey Methodology: Details from Nine Polling Organizations, 1988 and 1992." *Public Opinion Quarterly* 59:98–132. Available at http://j.mp/2n6au5o.

Wilson, Timothy D., and Nancy Brekke. 1994. "Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations." *Psychological Bulletin* 116 (1): 117.

APPENDICES

APPENDIX A. INCENTIVES TO COOPERATE

The structure of any industry–academic partnership must satisfy a company's legal, fiduciary, and business needs; academics' need for scientific freedom to work and publish; and the public's need for privacy and potential social goods derived. The basis for the partnership, then, is a structure, process, and purpose that is incentive compatible for all involved.

Our plan is incentive compatible for a company because it (1) decides on the commission cochairs along with the funders; (2) chooses the general topic areas that may be studied, given personnel and other resource constraints; and (3) retains the ability to exercise control in the event that a specific research project would violate a company's legal obligations, interfere with ongoing or imminent litigation, violate privacy, or compromise proprietary information or competitive standing.

As a trusted third party, the role of the commission at Social Science One is to simultaneously protect the company, the public, and the scholarly community. It must ensure that the definition of research questions and formal requests for proposals are not so narrowly stated that they predetermine the answer. In that circumstance—where no one is vulnerable to being proven wrong—nothing of value to science can be learned, and academics would be uninterested in participating. At the same time, few companies would participate in facilitating research designed solely to evaluate its own actions, at least not without the ability to learn how to improve its products and services going forward. If either outcome seems likely, no industry–academic partnership is possible. And if the company constrains questions in ways that violate this agreement, Social Science One is free to report this publicly.

Academics may prefer that companies have fewer rights to choose what can be analyzed so they can access any data they wish, but no firm will (or legally could) go forward without these rights. Insisting that they offer unrestricted rights would mean no data sharing (except with prepublication company approval), which gets us nowhere. Moreover, just as many requests for proposals from charitable foundations and governments allow for only a circumscribed set of topics they choose, companies have this right under our proposal as well. For-profit and nonprofit organizations may have different motivations for providing funding for certain questions and not others; however, requests for proposals from all are constraining to some degree.

Individual customers, whose data was gathered beginning with their decision to sign up for a technology firm's service, have the legal and ethical right to privacy. Of course, they will have agreed in the "terms of service" for their data to be used for research, but we must ensure they also benefit from the results of this research. The structure of our project is designed to reduce the risks of privacy violations and maximize the potential benefits from the research, but it is also incumbent upon Social Science One to clearly communicate the tradeoff. Fortunately, there is precedence for successful communication of this tradeoff in other areas of research (Unger et al. 2016).

The optimal way forward, then, is to find research questions of intellectual interest to the scientific community and either provide valuable knowledge to inform product, programmatic, and policy decisions, or are orthogonal to firm interests. Of course, any company participating in this process must understand that the purpose of research is to learn new things, discover answers to existing questions, and find new questions never before conceived. As a result, some bad news for the company will sometimes unavoidably surface. But even that bad news can be useful for the company as it improves its products. If successful, the results for everyone involved and many others should be a great deal of social good for the public as well.

APPENDIX B. ETHICAL PROCESS STANDARDS

Achieving consensus about ethical research principles in any area is rare, and especially difficult as societal views morph in response to

fast-moving technological developments, increasingly informative data collection, and ever-changing technology products. Instead of the impossible task of trying to select specific rules everyone agrees with, we institute the following nine rigorous ethical processes and adapt rules along the way within these processes.

First, to ensure accountability for the actions of individual researchers, proposals may only be submitted by colleges and universities, other nonprofit institutions, or commercial firms not competing with the data provider, on behalf of faculty or others with principal investigator (PI) rights. (An applicant with "PI rights" has been given permission to represent their institution; students, postdoctoral fellows, and others may participate as co-PIs on faculty-led projects.) We also require the researcher's institution to be a party to all research data agreements.

Second, to submit proposals, PIs must receive approval from their own university's Institutional Review Board (IRB), or the equivalent in other countries. Under US federal regulations, applicants are not permitted to determine whether their research is exempt or otherwise meets the rules, and so all applicants must go through this process. Applicants from countries where their IRB equivalent does not certify Common Law compliance (mostly from outside the United States and Western nations) or institutions not required to have IRBs must obtain certification from an external IRB, such as the Western IRB (see wirb.com).

Third, we ask peer reviewers to evaluate each proposal's scientific merit and its potential benefits to society, and at the same time are asked to review the ethical track record of the proposed investigator and the potential costs of the research to research subjects and others. This process is designed to ensure that only responsible researchers are granted access in appropriate circumstances.

Fourth, we send all proposals that pass standard peer review for a separate ethics review, from either SSRC or Social Science One, to ethicists specializing in the types of data we will make available. This step is especially useful for IRBs that are not current on ethical problems involved in social science analyses. This unfortunately is a common problem, in part because IRBs were originally designed for medical research and still spend most of their time evaluating research outside of the social sciences. Ethics reviews are not pass/fail determinations, since ethical (or even normative) principles do not have unanimous support; we instead use them to help researchers (and Social Science One) think through issues in sophisticated ways that they might not have previously considered.

Fifth, the privacy of individuals represented in firm data must be protected. Any breach would damage the credibility of the researcher, their university, the process we have arranged to make data available for the academic community, the social good intended by the research, and the reputation of the company. As such, researchers' access to and use of such data are held to higher standards for privacy, confidentiality, and security than required by any existing law or university practice. Data-access rules are calibrated to the sensitivity of the dataset to be analyzed. For datasets that raise no legitimate privacy concerns, data may be released publicly or with minimal safeguards, such as via application programming interfaces, or with much aggregate level data commonly released by social media firms. For the most sensitive datasets, researchers may need to work inside an actual or virtual company facility and have their work continually monitored or subjected to full audit. (A beneficial side effect of full audits is that they help ensure replicability, about which more below.) For datasets with sensitivity between these two extremes, scholars may have to access data on highly secure, encrypted laptops provided by the firm, or via other techniques. Most of these access methods do not involve distributing the datasets to anyone. Instead, researchers mostly access data on secure servers.

Our procedures effectively change from a regime of individual responsibility, where scholars with access legally agree to follow the rules and the rest of the community hopes they comply, to one of collective responsibility, where multiple people are always checking and the risk of improper actions by any one individual is greatly limited.

Sixth, social science insights are normally about population averages and broad patterns, and for which facts about any one individual are unnecessary and not of interest. Social scientists are usually interested in patterns about everyone, not anyone in particular. As such, privacy-protected analyses of certain types of data can be facilitated through cutting edge work in computer science on "differential privacy" (Dwork et al. 2006; Gaboardi et al. 2016). For certain classes of data and related analysis, these techniques lead to modified datasets or modified statistical analysis procedures that can enable social scientists to discover aggregate patterns while at the same time being able to provide mathematical guarantees of privacy for any individual. This emerging field holds great promise for protecting the privacy of individuals and advancing scientific knowledge for all of society. Because these techniques have not yet been developed for all types of data of interest and the consequences for statistical inference are still being worked out, we are enlisting researchers in this to modify their techniques to apply to the types of data we are analyzing.

Seventh, all research generated by the data provided must be able to follow the "replication standard" and thereby produce and archive replication data files (King 1995). This means that published research completed under grants from this process will be replicable by other researchers under specialized conditions that we are developing and publishing. The privacy and confidentiality concerns of this type of research obviously complicate this process, but several procedures are available to respond to those issues. A formal citation will be established for each data subset with a "universal numeric fingerprint" that uniquely identifies the content of a dataset even if the format in which it is stored changes (Altman and King 2007) along with a persistent identifier that locates it. The code, methodological details, and metadata (but not data) will be publicly available in Dataverse (see dataverse.org). And the full replication archive, including the data and all procedures necessary to replicate the analyses will be available and accessible by academics.

Eighth, the independence of academic research from private firms and nonprofit foundations with substantive interests or ideological preferences must be protected. Having multiple foundations with differing perspectives can be especially helpful for highly charged partisan and sensitive issues. When feasible, we recommend adopting rules to prevent any one foundation or viewpoint from having too much influence over the process. In other industry–academic partnerships—such as in less politicized environments or in substantive areas where funding from nonprofit foundations is unavailable—our plan's two-part structure still works with few modifications and without any added difficulties. The only real difference would be that the company may fund the commission, consultants, and outside experts directly or through establishment of an independent trust.

Finally, because of the ever-changing nature of ethical understandings, we are enlisting researchers in ethics to study the commission's decisions, how they are viewed by other academics and the general public, and how they might be improved. Such information must be used to improve the commission's processes and decision making.