Duke UNIVERSITY INITIATIVE ON SURVEY METHODOLOGY AT THE SOCIAL SCIENCE RESEARCH INSTITUTE

2021 Future of Survey Research Conference

March 15, 2021

Duke University

Sponsored by the National Science Foundation, Grant #2040847

Conference Report

Prepared by Gabriel Madson and Alexandra Cooper
Duke Initiative on Survey Methodology

Duke | SOCIAL SCIENCE RESEARCH INSTITUTE

Table of Contents

1. Executive Summary

Survey research is at a crossroads. For at least half a century, survey data have been essential to government agencies, policy-makers, businesses, and academics across different fields to inform a wide range of critical decisions with far-reaching consequences. Even in an era of "big data," surveys remain fundamental to understanding and shaping the economy, politics and governance, and society. Yet challenges to conducting high quality surveys are substantial and increasing. Face-to-face interviewing remains the gold standard of survey research, but the rising costs of such interviews are prohibitive. New technologies, techniques, and data sources present opportunities to improve the efficiency and speed of survey data collection and/or reduce its costs but have shortcomings that may exceed their advantages. To examine and develop strategies to address the challenges facing survey research, the Duke Initiative on Survey Methodology hosted a conference January 14th and 15th, 2021, on the Future of Survey Research. This report summarizes the proceedings and highlights key recommendations that resulted.

Held virtually, the conference attracted attendants from across the US and around the world. Panelists and presenters came from 11 universities across the US and UK and from Facebook, Google, NORC, Pew Research, and the Research Triangle Institute. The conference commenced with a welcome from Arthur Lupia, Assistant Director of the National Science Foundation. Lupia remarked on the many challenges experienced in 2020 and the innovations those challenges have fostered, and detailed a hierarchy of needs for survey research from the perspective of federal funders centralized around a key need – identifying tools to produce content that generates more numerous, accurate interpretations of actionable information.

Over the course of two days, the conference then used four panel sessions, a keynote, and a brainstorming session to examine current innovations within survey research, the needs of the field and those it serves moving forward, and ideas for improving the extant social survey infrastructure. Conference participants broadly agreed that maintaining the extant survey infrastructure is untenable given the rising costs of conducting quality research at a national level. Drawing on their experiences working with and directing academic and government surveys, they examined recent innovations in survey research and the use of "big data" for ideas about how social survey infrastructure can be improved moving forward. The following three takeaways emerged from the conference proceedings:

- There is wide variation across the field in approaches and standards. Although we speak of gathering data that are fit for a purpose, large-scale publicly funded surveys gather data made available for general use by a wide range of researchers across a diverse assortment of disciplines, using an array of methods, and requiring different levels of precision. Better understanding and more thoughtful consideration of the implications of such factors as sample design and other sources of survey error are essential.
- The field must develop methods for extracting value from imperfect data, and achieving this will require NSF investment. A critical resource is a large-scale national sample survey that obtains benchmark estimates of non-demographic characteristics on key dimensions, such as religiosity and social trust, making it possible to assess – and potentially adjust – for the representativeness of other, less high-end surveys.
- Survey professionals recognize that academic and the traditional NSF-funded infrastructure surveys require and achieve a higher standard than do other surveys such as (e.g.) pre-election polls. However, many members of the general public do not. As the general public is often the subject of surveys and a key audience for their outcomes, the prevalence of low-quality surveys damages all survey research, including the gold-standard surveys conducted by or funded through the US government. The entire survey industry must be brought into this discussion and the solutions it generates, because every bad survey endangers all surveys.

2. Conference Motivation

The NSF has invested heavily in survey research for decades, thereby generating the highest standards for the highest quality of data from survey research in the world. And it has recognized the need for systematic, institutionalized means for gathering a variety of different types of this high-quality survey data. The NSF social survey infrastructure includes three main recurring surveys: the American National Election Study (ANES), the General Social Survey (GSS), and the Panel Survey on Income Dynamics (PSID). The NSF has also supported a variety of survey projects that have become something of a set of institutions in their own right, including the Comparative Study of Electoral Systems (CSES), the Cooperative Congressional Election Study (CCES) and Time-sharing Experiments for the Social Sciences (TESS). Together, these surveys provide invaluable insight into the social sciences and serve the needs of a diverse group of academics, policy-makers, businesses, and educators.

Yet, the challenges facing survey research are driving up the costs NSF faces to maintain its extensive survey infrastructure, and the rate at which those costs are rising is increasing. New technologies and data collection strategies offer the possibility of reducing these costs – but their impact on data quality and thus, value, is unknown. And though its long-term impacts on society are uncertain, the COVID-19 pandemic has both created new challenges for and prompted extensive innovations in survey research (as in so many areas). These challenges and innovations are certainly not the main driver of the issues facing survey research, which long precede the pandemic. However, they both intensify the problems the field is facing and offer the possibility of novel and previously unachievable solutions.

In recent decades the established gold standard of in-person, face-to-face long-form interviews has become increasingly costly to run and increasingly difficult to achieve. Survey research has also been critiqued for its reliance on a large range of self-reported responses provided to a very large number of questions asked at one or a very few points in time. Equally, the increased availability of "big data" together with innovations in means of combining survey responses with different forms of such data offer advances on their own terms, as well as possible enhancements to the quality of respondent data. New modes of surveying, and new tools in survey research more generally, present real advantages but also pose risks. It is essential – and far from easy – to assure that new survey tools are implemented in ways that ensure data quality and also that their technological features are both exploited effectively and integrated into rigorous theories about opinion and belief, choice and behavior.

Bringing together experts to discuss the problems facing the gold standard of survey design, as well as the challenges facing the social survey infrastructure in general, motivated the conference on The Future of Survey Research. Survey methodologists from around the world came together at the conference to educate each other about innovations at the forefront of survey research today and to discuss what steps and collaborations will make it possible to sustain and to improve social survey infrastructure. By convening a diverse group of researchers from different disciplines and from across academia government, industry, and non-governmental organizations, the conference provided a venue to share research findings, network, and encourage new collaborations to bridge disciplinary gaps and catalyze innovations. The conference also helped to foster a greater understanding of the challenges and opportunities for survey research in the future. The next step will be to build on the discussions that arose from and goals identified through the conference to plan and undertake actions to improve upon existing infrastructure while implementing useful additions to the social survey enterprise.

3. Conference Structure and Findings

The Future of Survey Research conference was hosted January 14th and 15th, 2021 at Duke University with support from the National Science Foundation (#2040847) and the Duke Initiative on Survey Methodology (DISM). Due to the ongoing COVID-19 pandemic, the conference took place virtually. Panelists and presenters included top survey methodologists and practitioners interested in advancement

of the social survey infrastructure; these 17 speakers came from 11 universities across the US and UK and from Facebook, Google, NORC, Pew Research, and the Research Triangle Institute. The conference was well attended, attracting over 800 registrants from academia, government, industry, and non-profits across the US – 36 states – and the world – 21 countries – and had at least 200 people in attendance at each one of its 4 panels and its keynote address.

Based on the difficulties facing the future of the social survey infrastructure discussed in the previous section, the conference was organized around panels examining the following four topics:

1) Data quality and transparency, to establish the standards by which data collections are evaluated. Chaired by D. Sunshine Hillygus (Duke University), with presentations from Courtney Kennedy (Pew Research), Nicholas Valentino (University of Michigan), and David Vannette (Facebook).
2) Survey methods innovations, to learn more about the most cutting-edge advances in the field. Chaired by Stanley Presser (University of Maryland), with presentations by Fred Conrad (University of Michigan), Rachel Gibson (University of Manchester), Stephanie Eckman (RTI International), and Nada Ganesh (National Opinion Research Center).
3) Looking beyond survey responses in the world of Big Data, to consider the integration of auxiliary data, such as administrative records and other so-called Big Data. Chaired by Craig Hill (RTI International), with presentations by Frauke Kreuter (University of Maryland), Mario Callegaro (Google), Ted Enamorado (Washington University-St. Louis), and Amy O'Hara (Georgetown University).
4) Imagining new institutions, to build on previous topics with consideration of implementing new frameworks for conducting survey research in the social sciences. Chaired by John Aldrich (Duke University), with presentations by Barbara Entwisle (University of North Carolina-Chapel Hill), Jeremy Freese (Stanford University), and Shanto Iyengar (Stanford University).

The conference also included a group brainstorming session and a lunch presentation on the United Kingdom Household Longitudinal Study made by Olena Kaminska (University of Essex). The full conference schedule appears in Appendix A, and information about all participants including brief biographies in Appendix B. Recordings of each panel are available at https://sites.duke.edu/surveyresearch/presentations/ using the password surveyresearch2021.

Arthur Lupia, director of the Social, Behavioral, and Economic Sciences Directorate at the NSF, started the conference with brief opening remarks that motivated the structure and questions of the conference. In his remarks he raised four questions:

- How do sampling decisions affect the ability to generalize; in terms of the quality and usability of the resulting data is there something between a nationally representative and a convenience sample, that is useful?
- How does mixing data types affect the evolution of what the data mean?
- How does questionnaire design (and the psychology of survey response) affect the possibility of accurate interpretation?
- How do categorization decisions (e.g. concepts, weights, indices) affect opportunities for accurate inference?

The panels and conversations that followed highlighted the contemporary relevance and importance of these questions and identified strategies that will help answer them.

The **first** panel focused specifically on issues related to data quality and transparency with the goal of ensuring that there is a shared standard by which potential innovations will be evaluated. In an environment of declining survey response rate and rising costs of in-person surveys, researchers have increasingly turned to alternative approaches to survey public opinion – online non-probability panels,

IVR polls, and text message surveys on mobile devices. There is wide variation in the design and quality of these various survey approaches. For example, online surveys choose from a range of different sampling designs, recruitment strategies, and implementation procedures, all of which have implications for data quality; this increased variability in survey design makes it more difficult, but also more important, to assess data quality. Doing so requires common, meaningful standards for evaluating data quality as well as sufficient disclosure about the methods used for data collection.

It needs to be more widely understood that data that may be of sufficient quality for answering one question could nevertheless not be of sufficient quality for answering other questions. Data quality is colloquially defined as "fitness for use," but large-scale publicly funded surveys produce data for *general* use and thus, decisions about survey design and administration are made without full – or even much – knowledge of the research questions the data will be used to examine. An extensive survey methodology literature and set of standard practices exist within the industry. Neither the methods nor the practices have yet to completely permeate the full range of those who use survey generated data in the social and behavioral science community (e.g., Biemer et al. 2017) or those in the more general polling industry. The difficulties associated with inaccuracies in the 2016 and 2020 election polls demonstrate the case. Establishing standards on data quality provides valuable guidance on the types of innovations that survey science should prioritize moving forward, recognizing that maximizing not the quantity but the quality of the data must be the goal.

Panelists focused on determining a set of standards for survey infrastructure worth investing in. They discussed ideal features for the future of surveys with respect to data collection, interview process, and questionnaire design. These ideals were then compared to sometimes harsh realities. For example, panelists noted the cost efficiency of transitioning from face-to-face interviews to web-based or online interview surveys, while also providing evidence of lower data quality measures and possibilities of selection bias in the alternative mode formats.

Kennedy presented first, discussing four ideals for the future of survey research, including the need for survey respondents to provide genuine answers and for a sample to represent all segments of its population equally well. The difficulty of achieving these goals was made clear from Valentino's presentation about survey mode effects in the 2012 and 2016 ANES, where there were major differences in personality scales between respondents who took the survey online versus those who took it in face-to-face mode. This suggests that 1) online respondents may be more likely to engage in poor data quality behaviors such as satisficing compared to the face-to-face interview context and 2) factors such as personality traits of respondents, are likely correlated with willingness to take an online survey versus a face-to-face one, a problem that sampling alone cannot fully address. Vannette concluded the first session by discussing how the pandemic led to new uses of social media data to detect trends, in this case related to outbreaks of COVID-19, before such trends are detected by large federal systems. Vanette's presentation included discussion of data quality issues and selection effects, as well as privacy considerations in the social media context.

Discussion focused on how to reconcile lowered costs and data quality when transitioning survey modes from face-to-face to online formats. One central conclusion that arose from these discussions as a means for assessing online sample quality was access to national benchmarks outside of the traditional demographics, such as religiosity, social trust, or political partisanship. Attention should be focused on benchmarking measures of core concepts that are therefore commonly employed as explanatory variables. This national benchmark survey idea launched discussions that continued throughout the conference on how such an endeavor could be accomplished.

The questions addressed during this session centered on the first and third questions Lupia raised in his welcoming remarks: (a) how do sampling decisions affect the ability to generalize, and (b) how do

questionnaire design and the psychology of survey response affect the interpretation of the resulting data? They included the following:

- discussion of the appropriate metrics and required resources (notably, data benchmarks) for evaluating the quality of data collected using nonprobability methods;
- strategies for diagnosing and discouraging undesired behaviors and trolling in online surveys (such as respondents looking up answers to knowledge questions, paying little attention, or various forms of "cheating");
- ways to improve or incentivize attention span (thereby allowing for longer, more complex questionnaires) or alter questionnaire design in a way to mitigate the effects of limited interest and attention on the part of some or all respondents;
- issues of comparability in a transition away from face-to-face interviews to less costly alternatives. For example, how do we "translate" a face-to-face and/or voice-centric phone survey into an online survey, or especially one designed for responses via cell phone?

Following this panel, Olena Kaminska from The UK Household Longitudinal Study discussed the study's Innovation Panel (IP) during a lunch **keynote**. She presented the IP as a potential guiding model for testing methodological innovations into long-running major NSF-funded survey panels such as the ANES and GSS. The UK Study, which also includes long-running mainstage panels, relies on the IP, a separate panel, to expose respondents to changes or additions being considered for eventual integrations into the mainstage panel. This technique allows for assessments of data quality and feasibility without interruption of or potential contamination in the mainstage panel – and is directly relevant to Lupia's third question, how questionnaire design and the psychology of survey response affect accurate interpretation. Discussion following the presentation focused on how treatment effects in the IP could be properly assessed for the mainstage panel and how the survey team selects experiments to prioritize for the IP. Participants also discussed the feasibility of a model like the IP for panels like the ANES and GSS that collect data over much smaller timeframe than the UK Household Longitudinal Study.

The **second** panel focused on new advances in survey research methods. Survey researchers from diverse academic fields, from government, and from industry have been continuously evaluating new approaches to improve surveys and to address challenges in the industry – surveys were being conducted via email as early as the 1980s (Callegaro et al. 2015) and online surveys now make up the majority of academic, business, and government survey work (Schaeffer and Dykema 2011). Although survey science is changing rapidly, it has been almost a decade since the last major NSF-funded event focused on survey science.[1] Given the significant advances in survey methods since 2012, the conference hosted a session devoted to discussing the accumulation of knowledge about new and emerging approaches.

This session focused on features of new technologies that have become available and what value these can offer the survey enterprise. Topics included both the collection of survey data itself and – although this topic was also addressed separately in a panel devoted to big data – the availability of and advantages to using data sources outside of traditional surveys to enrich, enhance, and expand on data collected via surveys. Panel speakers examined the following topics:

[1] In February 2012, a workshop, "The Future of Survey Research: Challenges and Opportunities," was organized by a Subcommittee on Advancing SBE Survey Research to address the challenges facing survey-based data collection at that time (e.g., falling participation rates, rising costs, coverage of frames), innovations in survey methodology, and opportunities for merging big data. The report on that workshop is available online, at this URL:
https://www.nsf.gov/sbe/AC_Materials/The_Future_of_Survey_Research.pdf .

- the promise of online two-way video interviews conducted over online platforms such as Skype, Zoom, or WebEx;
- gains realized from linking social media data with traditional survey responses;
- the feasibility of a dual sampling approach that supplements a probability-based sample with cheaper cases drawn from a non-probability panel;
- reliance on imputation and non-probability samples as a means of informing changes to questionnaires in long-standing governmental surveys;
- multi-mode, multi-media methods for conducting surveys.

Following the innovation panel presentation, the second session focused on additional up-and-coming innovations for survey research. Conrad expanded on the survey mode discussion from the first session by discussing results on mode effects for web-based, online live interviewer, and online pre-recorded interviewer surveys – noting that live interviews seemed to be most similar to the face-to-face experience whereas pre-recorded videos appear more akin to online web surveys. Interest in video tools echoed Lupia's earlier remarks about innovations fostered because of the pandemic. Pandemic-related shutdowns have necessitated the use of video technology for myriad activities previously conducted principally or entirely in person, thereby increasing the penetration of this technology among the nation's residents, and their familiarity with these tools. Gibson then discussed two methods for linking social media data with survey responses as a method for obtaining key information about survey respondents while reducing the dependence on survey questions. One major drawback she noted was the lack of an industry standard weighting procedure to deal with selection bias among those with social media profiles who consent to their data being accessed. Next, Eckman presented an innovative technique relying on multiple imputation to predict how changes to government surveys in aspects such as their question wording would affect established trends. She presented this method, which can rely on cheaper non-probability samples, as a workaround to government rigidity when it comes to making changes to long-running government surveys. The final presentation in the session, by Ganesh, covered how the Associated Press's VoteCast system is able to predict election contests with a high level of accuracy without using in-person exit interviews. The system provided highly accurate predictions in the 2018 US Senate races via both a probability and non-probability sample.

Discussion questions for the presenters covered the ins and outs of each of the innovative techniques they presented. The biggest question tackled by presenters was in contemplating what the role of survey data will be in the future when we are even better able to rely on auxiliary data like social media posts and administrative records. Panelists agreed that, just like surveys, these alternative data also suffer from their own forms of measurement and coverage error and noted that moving forward survey researchers should consider mixed-method approaches that can account for the possibility of error on both sides.

The **third** panel centered on Lupia's third question, how mixing data types affect the evolution of what the data mean. It focused on the ways in which so-called Big Data – administrative records or social media data as well as census data at the individual or aggregate level – can augment or otherwise supplement survey data. While there is widespread interest in identifying ways in which Big Data can replace, supplant, or complement survey data, to leverage these data sources successfully requires careful integration of data science and survey science to have high-quality and useful information. There are also many challenges to using "Big Data" – among them, incomplete and missing information, measurement issues, issues related to privacy and confidentiality, difficulty in accessing proprietary datasets, and inadequate training on analysis among potential users (Foster et al. 2016; Japec et al. 2015).

Themes in this session tackled many of these challenges. What are the benefits and limitations of using voter files as a sample frame from which to draw probability samples? What are the best practices to obtain consent for linking survey and administrative data? What are the obstacles – technical and ethical –

to linking social media data with individual survey responses? How should we reconcile conflicting information in administrative records and surveys responses, when neither can be considered a "truth benchmark"? For example, a comparison of administrative records from the Social Security Administration with responses in the American Community Survey (ACS) revealed that more than 40% of individuals identified as noncitizens in administrative records reported themselves to be citizens in their self-response to the ACS (Brown et al. 2019). As there are well-documented ways in which the administrative records are known to be outdated and inaccurate (Bond et al. 2014), those cannot simply replace the data collected through the ACS. Yet conversely, respondent error, social desirability, and concerns about the risks associated with revealing a status other than "citizen" surely mean that some non-citizens report that they are citizens when responding to the ACS.

Kreuter kicked off the presentations in this panel by discussing the principal use of big data as a compliment to survey data – obtaining data that is difficult or impossible to measure using survey responses. She contrasted these two approaches as involving big data gathering big quantities of somewhat fuzzy information, whereas surveys obtain small but precise information. Callegaro and Enamorado then each presented on survey projects that are improved by big data. Callegaro discussed how Google has used search-based data as a regional tool to forecast where COVID-19 peaks will likely pop up as an alternative to large federal surveys that typically cannot determine outbreaks until after they have occurred. Enamorado then presented on probabilistic record linkage – a technique for quickly matching survey respondents to administrative records using common sets of variables. This method was demonstrated using the ANES and administrative voter files to determine the percent of over-reporting on voter turnout in the ANES. Finally, O'Hara concluded the presentation portion of the session by discussing the ethical considerations associated with linking large governmental data sources as well as the institutions that need to be developed to address these ethical concerns.

Panelists and audience members spent the remainder of this panel session discussing many of the high barriers for entry researchers face when it comes to working with big data. Many big data projects are quickly overwhelmed by the technical management associated with transforming what can be giant, messy datasets into a usable product. Discussion on this portion of the topic stressed the importance of educational tools and training at universities and businesses to establish strong practices of working with very large, complex datasets. Presenters and audience members also discussed the need for collaboration across the profession when it comes to accessing and working with big data projects. Much of the current work using big data in survey research is subject to private industry restrictions that often make it inaccessible for anyone outside of the most prominent academic institutions.

Finally, the **fourth** panel covered the future of survey research in the social and behavioral sciences, including future directions for the ANES and GSS as well as discussion of the future of federally funded surveys more generally. It focused on potential institutions and partnerships; collaborative partnerships in survey research are not new (e.g., Kreuter et al., 2020), and this session sought to identify the structure and framework most effective for advancing knowledge based on what we have learned from previous such efforts. Conversations drew on a range of experiences with the collaborative CSES, CCES, and TESS as well as with industry partnerships and the UK Innovation Panel Competition, part of the UK Household Longitudinal Panel that allows researchers to submit ideas for methodological or substantive experiments to be run on an "Innovation" panel comprised of 1,500 households.

The panel both examined possible new directions and ways to foster innovation and considered ways to maximize the value of extant federally funded survey infrastructure. Panelists noted how much has been invested in data already gathered – currently archived in disparate places, generally with archiving given a lower priority than the implementation of new science – and considered how this remarkable history of public opinion and behavior might best be maintained for effective use. They also considered the different extant gold standard surveys themselves and ways that we could leverage particularly the ANES and GSS

to magnify their impact. Are there ways that they might be integrated (whether in design, implementation, or post-data-acquisition structures) to the benefit of various user communities? The planned ANES-GSS collaboration in 2020 was one small step in that direction, but it is essential to consider other ways that these infrastructure surveys might be a platform for substantive and methodological innovation. Their high-quality probability sample is a key cause of the high expense involved in conducting them, but is also central to their value. Panelists suggested that we consider ways to make the sample frame available as infrastructure.

However, panelists also considered other alternatives to the idea of building on existing infrastructure and bringing it together. One suggestion was a Center for Survey Innovations for promoting advances in survey methods, practices, and content, as a way to advance survey science most assuredly and rapidly; another suggestion was for a broader center aimed at the integration of surveys and big data. A similar (but distinct) idea was a Survey Research Laboratory to promote research and development by using RFPs to generate new ideas, assessing those through testing, and guiding suitable successes into application in some gold standard setting.

Each of the panel presentations stressed the need for innovation given a context of decreasing budgets and increasing costs. Iyengar discussed the 2020 ANES's efforts to link responses to social media posts as a way of studying political behaviors, such as campaign ads watched in news feeds or candidates mentioned in posts. The study also included a randomization of asking for consent to social media access at the beginning and end of the study, and found consent at the beginning was slightly easier to secure. Freese then discussed the GSS's transition to online data collection for the 2020 sample due to COVID-19. This transition included adding a panel study re-interview from the 2016 and 2018 samples to look at changes in opinion from the mode transition as well as a cross-sectional study to maintain cumulative cross-sectional trends. He discussed the many wording changes that needed to be implemented for studies typically conducted face-to-face (e.g., changing "I would like to ask you" to "We would like to ask you"). Entwisle concluded the presentations by discussing the concerns of survey research expressed by federal agencies and returned to some of the earlier panel discussions of leveraging alternative sources of data to supplement (or substitute) for pieces of the social survey enterprise. She argued some form of transition to relying on alternative data sources is needed to address problems of cost while still keeping ingenuity of surveys useful.

Following the presentations, discussion from panelists and audience members centered around properly conveying the usefulness of the social survey enterprise to an audience outside of the scientific community. Given the unusual era in which this conference took place, conversation also examined topics relevant to survey research during the COVID-19 pandemic such as how to properly weight data collected for time series cross-sectional studies like the ANES and GSS during this time. These surveys are typically weighted based on national benchmarks from large federal surveys (e.g., American Community Survey or Current Population Survey) that were likely affected by the pandemic, adding uncertainty to processes like weighting.

Each of the conference sessions examined aspects of current studies, tools, and approaches with promises of new directions and the potential for important innovations. Each also drew from a different set of potential participants to map out potential directions for moving survey research toward the future.

5. Conclusions and Next Steps

Three conclusions arose from the conference panels and brainstorming sessions on how to improve the social survey infrastructure. Below, each will be discussed in turn along with the next steps required to put these improvements into place.

1) Improving the current survey infrastructure, including the ANES, GSS, and PSID, through a combination of what we have learned from forced changes to survey operations in the wake of COVID-19 and the newest and most promising developments in survey methodology innovation.

2) Increasing collaboration of the survey methodology field to remove high barriers of entry and enable better access to auxiliary data collections.

3) Implementing a national benchmark survey that goes beyond basic demographic benchmarks found in large government surveys like the American Community Survey and Current Population Survey.

<u>Improving Current Infrastructure</u>

One of the main topics discussed throughout the panels and brainstorming sessions was how to improve the social survey infrastructure currently supported by the NSF, in particular the large-scale face-to-face surveys. As discussed in the Conference Motivation section, these face-to-face surveys, which represent the gold standard of survey practice, are facing higher costs per respondent every year. While the impossibility of conducting face-to-face interviews in the wake of COVID-19 necessitated speedy innovation, what has been learned from those innovations – and the widespread changes to important factors like video connectivity and familiarity with video conferencing tools among much of the general public – have also reinforced the important role these surveys play as the benchmark against which less expensive surveys are evaluated. The final session of the conference directly dealt with these two considerations, discussing the causes and consequences of rising costs to data collection for surveys like the ANES and how these surveys have attempted to innovate in the wake of COVID-19. After the final session, panelists and attendees discussed the limitations of the current survey infrastructure as well as ways in which it can be improved.

The two primary limitations discussed throughout the conference were uncertainty with mode effects and the difficulty of bringing survey innovations into long-running surveys. The data quality session in particular highlighted that survey mode has consequential impacts on our estimates of certain social and political behaviors, likely due to selection effects from underlying characteristics and attitudes associated with participation in certain survey modes. While transitioning surveys like the ANES to online formats offers significant cost savings, it is still unclear whether and how these alternative modes can meet the gold standard of face-to-face interviews. The other limitation to improving current survey infrastructure comes from the difficulty of implementing innovations without inadvertently corrupting the time series.

These challenges tie directly into the questions that Lupia raised in his welcoming remarks, and addressing them will provide opportunities to examine and answer those questions. The simple fact is that innovation is essential, whether in the form of an Innovation Panel similar to the UK's, a Survey Research Lab, a Center for Survey Innovation, and/or careful and novel use of methodological approaches. It is essential that any of these possible responses be crafted to address those key questions about the effects of sampling decision on data quality, the incorporation of new sources of data, the impacts of changes to questionnaire design (and mode of administration) and the ways in which the resulting information is coded, categorized, and used in analyses.

Increasing Collaborations

Access to restricted data from private industry surveys and auxiliary data worth linking to data collected through surveys both often involve a high barrier of entry that only prominent scholars and elite institutions can surpass. As a result, a few companies' interests and/or a few scholars' relationships can limit what collaborations emerge, while much of the survey research community is left on the outside looking in. Yet, successful examples of collaborations around big data have required collaboration by multiple, diverse entities to achieve success, such as that between the University of Maryland and Facebook for the COVID-19 Symptom Survey.

A distinct but related problem is – bluntly stated – that there are a lot of bad surveys, and a plethora of survey industry professionals who do not apply, and may not even be fully aware of, good survey research practices. The existence of substandard surveys may not on its face appear to be something that would affect the gold-standard portions of the field. Yet the reality is that every bad survey – and each inaccurately reported or simply inaccurate result – damages survey research generally. The problem is the nature of the field: conducting surveys depends on contacting individuals, and having those individuals be willing to devote time and effort to providing information on topics selected by survey researchers using tools developed by those researchers. It also requires confidence in survey findings on the part of the ordinary people who comprise the audience for many of those findings and the policies that result from them. If ordinary people do not perceive survey research as useful and worth contributing to – that is, participating in – survey research becomes impossible, and the value of the data it provides declines or disappears. Every bad election forecast, every push poll detracts value from the field.

It is essential, then, to (a) expand and increase access to large-scale, productive collaborations between and among survey researchers and other data providers, and (b) reduce the prevalence of sloppy or misguided survey research. Both these things are needed to ensure that survey research continues to be recognized not just by scholars, but by the general public as a valuable part of the scientific toolkit. We must increase connectivity across the field, and must maintain, further develop, employ, and promulgate shared standards for practice. Achieving these goals will not by itself be sufficient to maintaining and revitalizing survey research, but is necessary to doing so.

Implementing a National Benchmark Survey

Much of the social survey infrastructure, even outside of NSF-funded surveys, bases sampling procedures and weighting adjustments on demographic benchmarks produced by federal government surveys like the American Community Survey and Current Population Survey. This adjustment process makes it possible for the samples in these surveys to reflect the demographic composition of the population of interest. However, as we saw in both the 2016 and 2020 US Presidential elections, adjusting samples to these demographic benchmarks is not sufficient to correct for survey biases. Underlying characteristics and attitudes outside of standard demographic are associated both with these social and political behaviors and with willingness to engage in certain survey modes and behaviors at all – such as taking an online survey and giving consent for access to social media.

As it stands today, there exists no timely, gold-standard attitudinal benchmark survey that survey researchers can draw on to make necessary sample and weighting adjustments for the factors that recently have proven so important to data quality and the representativeness of samples. The ANES and GSS are not conducted with sufficient frequency to serve this purpose. With the rise of alternative survey modes, such as online web surveys, and the use of big data to supplement the survey response, the lack of an attitudinal benchmark survey becomes increasingly problematic, as features of survey participation and response type in big data can be associated with attitudes outside of the standard demographic benchmarks provided by federal surveys. Across the board, the survey and polling industries need access

to benchmarks on attitudes like religiosity, political party identification, social trust, and more. Such information will make possible better assessment of sample and data attributes and quality, and will allow better adjustments through practices such as weighting, to facilitate meaningful reports and projections.

The idea of an attitudinal benchmark survey was first discussed during the Data Quality panel and much of the conversation on future directions of the field focused specifically on the implementation of this resource. There was strong agreement over the need for an attitudinal benchmark survey from virtually all sectors of the broader survey research community. Panelists and audience members discussed the possibility of a single, large-scale survey versus a variety of smaller surveys aimed at obtaining these attitudinal benchmarks. They also discussed that obtaining such a survey would likely require a collaborative effort from NSF-sponsored surveys like the ANES and GSS, as well as from reputable private polling firms. The collaborative efforts between the ANES and GSS in the wake of the COVID-19 pandemic suggest that such collaboration on an attitudinal benchmark survey is possible. Following the conference, researchers from various academic positions and survey firms plan to create a working group to discuss the details of implementation and content of an attitudinal benchmark survey.

This attitudinal benchmark survey as a new infrastructure resource for survey science – and survey practitioners – is the most concrete action items of the recommendations to come from this conference. We are pleased to spearhead developing a collaboration, and plans, for its implementation. Such a resource will prove tremendously valuable for the field generally, and will help us make significant progress toward answering the questions spelled out by Lupia. Providing those answers is essential to the future of survey research.

6. References

Biemer, Paul P., Edith D. de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars E. Lyberg, N. Clyde Tucker, and Brady T. West, eds. 2017. *Total survey error in practice*. John Wiley & Sons.

Bond, B., Brown, J. D., Luque, A. & O'Hara, A. 2014. "The nature of the bias when studying only linkable person records: Evidence from the American Community Survey." CARRA Working Paper #2014-08. Washington, D.C.: US Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-08.pdf.

Brown, J.D., Heggeness, M.L., Dorinski, S.M. et al. 2019. "Predicting the Effect of Adding a Citizenship Question to the 2020 Census." *Demography* 56: 1173–1194. https://doi.org/10.1007/s13524-019-00803-4

Callegaro, Mario, Katja Lozar Manfreda, and Vasja Vehovar. 2015. *Web survey methodology*. Sage.

Foster, Ian, Rayid Ghani, Ron S. Jarmin, Frauke Kreuter, and Julia Lane, eds. 2016. *Big data and social science: A practical guide to methods and tools*. crc Press.

Hillygus, D. Sunshine, and Jesse Lopez. 2020. "Easy as 1, 2, 3? Challenges of the 2020 Census and Implications for Political Science." Journal of Political Institutions and Political Economy 1, no. 2: 289-317.

Hox, J.J. and de Leeuw, E.D. 1994. "A Comparison of Nonresponse in Mail, Telephone, and Face-To-Face Surveys." Quality & Quantity, 28, no. 4: 329-344.

Japec, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., Lane, J., O'Neil, C., & Usher, A. 2015. AAPOR report on big data. https://www.aapor.org/Education-Resources/Reports/Big-Data.aspx.

Kennedy, Courtney, Andrew Mercer, Scott Keeter, Nick Hatley, Kyley McGeeney, and Alejandra Gimenez. 2016. "Evaluating online nonprobability surveys." Pew Research Center 61.

Kreuter, Frauke, Neta Barkay, Alyssa Bilinski, Adrianne Bradford, Samantha Chiu, Roee Eliat, Junchuan Fan et al. 2020. "Partnering with Facebook on a university-based rapid turn-around global survey." *Survey research methods* 14, no. 2: 159-162.

Schaeffer, Nora Cate, and Jennifer Dykema. 2011. "Questions for surveys: current trends and future directions." *Public opinion quarterly* 75, no. 5: 909-961.

Appendix A: Conference Advertising and Schedule

Conference Announcement: Future of Survey Research Conference

The Future of Survey Research Conference will convene an interdisciplinary group of researchers from a wide range of academic fields and industries to discuss innovations in social, behavioral, and health surveys. Via a virtual platform, it will foster deliberation centered on renewing and augmenting the extensive survey infrastructure the US has developed over the past half century. It will focus on mapping out potential directions for moving survey research toward the future by brainstorming ideas for cutting-edge substantive and methodological innovations, with the objective of maximizing the potential of survey research and survey data to serve science, governments, and industry.

The virtual conference will bring together scholars and practitioners from diverse backgrounds and fields in the various social and behavioral science areas that employ the highest quality survey research to support their activities. The conference focuses on four subject areas over the two days:

1) Data quality and transparency, to help establish the standards by which data collections are evaluated

2) Survey methods innovations, to assemble novel advances at the research frontier

3) Survey Plus (or looking beyond survey responses), to consider integration of auxiliary data, such as administrative records and other new types of "Big Data"

4) Imagining new institutions, building on the earlier foundations to consider the potential for collaborative innovations in survey research across the social and behavioral sciences, governments, and industry

By convening a diverse group of researchers from different disciplines, this conference will provide a venue to share research findings, network, and encourage new collaborations to bridge disciplinary gaps and catalyze innovations. The conference will help to foster a greater understanding of the challenges and opportunities for survey research in the future, with the possibility of transforming future data collection for US federal agencies and private industry.

Hosted by Duke University with support from the NSF, Award #2040847.

Conference schedule:

The conference took place from Thursday to Friday, January 14th and 15th, 2021.

Thursday, January 14th, 2021

9:30 Welcome & Introduction
- Arthur Lupia, National Science Foundation

10:00 Session 1: Data Quality – Setting the Standards for Future Research Collections
- D. Sunshine Hillygus, Duke University (chair)
- Courtney Kennedy, Pew Research Center
- Nicholas Valentino, University of Michigan
- David Vannette, Facebook

12:00 Lunch Keynote
- Olena Kaminska, University of Essex

1:00 Sesion II: Recent Innovations in Survey Methods
- Stanley Presser, University of Maryland (chair)
- Fred Conrad, University of Michigan
- Rachel Gibson, University of Manchester
- Stephanie Eckman, RTI International
- Nada Ganesh, National Opinion Research Center

3:00 Brainstorming Future Oppertunities

Friday, January 15th, 2021

10:00 Session III: Big Data
- Craig Hill, RTI International (chair)
- Frauke Kreuter, University of Maryland
- Mario Callegaro, Google
- Ted Enamorado, Washington University
- Amy O'Hara, Georgetown University

12:00 Break

1:00 Session IV: Imagining the Future of Survey Research
- John Aldrich, Duke University (chair)
- Barbara Entwisle, University of North Carolina – Chapel Hill
- Jeremy Freese, Stanford University
- Shanto Iyengar, Stanford University

3:00 Preparing for Next Steps

Appendix B: Participant Bios

John Aldrich, Ph.D., Pfizer-Pratt University Professor of Political Science, Duke University, specializes in American politics and behavior, formal theory, and methodology. He has served as President of the Southern Political Science Association, Midwest Political Science Association, and the American Political Science Association.

Mario Callegaro, Ph.D., is Senior UX Survey Research Scientist at Google UK, London, in the Cloud User Experience team. He works on any survey related projects within his organization. He also consults with numerous other internal teams regarding survey design, sampling, questionnaire design and online survey programming and implementation.

Frederick Conrad, Ph.D. Professor of Psychology, University of Michigan, is the director of the Michigan Program in Survey Methodology where he is a Research Professor. His current research concerns efficiency of text message interviews, data quality in live video interviews, and the potential for social media content to supplement or even replace certain survey data.

Stephanie Eckman, Ph.D., is a Fellow in the Survey Research Division at RTI International. She conducts research into the best ways to collect and analyze survey data. She is interested in data quality and using data to answer important questions about society. Her research focuses on frame creation and sample selection methods, the use of geographic information technology in surveys, and how respondent and interviewer motivations affect the data we collect.

Ted Enamorado, Ph.D., Assistant Professor of Political Science at the Washington University in St. Louis, is a faculty affiliate at the Center for the Study of Race, Ethnicity & Equity and the Division of Computational & Data Sciences. His fields of specialization are Political Economy and Political Methodology.

Barbara Entwisle, Ph.D., Kenan Distinguished Professor of Sociology, University of North Carolina at Chapel Hill, focuses on social, natural, and built environments and their consequences for demographic and health outcomes. Her work ranges from the study of migration, residential change and health from a life course perspective in the United States to agent-based modeling of migration and other responses to environmental stress in Northeast Thailand.

Jeremy Freese, Ph.D., Professor of Sociology, Stanford University, is interested broadly in the relationship between social differences and individual differences, and between social advantage and embodied advantage. This includes work differences in physical health, cognitive functioning, health behaviors, and the role of differential utilization of knowledge and innovations toward producing differences.

Nadarajasundaram Ganesh, Ph.D., is a Senior Statistician in the Statistics and Methodology department with NORC at the University of Chicago. Ganesh has responsibility for survey weighting, population control totals, survey data analysis, and developing statistical models and estimation methodology.

Rachel Gibson, Ph.D., Professor of Political Science, University of Manchester, focuses her research on the impact of new information and communication technologies on political parties, particularly with regard to their activities in the elections and campaigning sphere. While early accounts of the effects of the internet pointed to positive outcomes such as increased party competition, grassroots activism and more meaningful interactions with voters, twenty years on those expectations look increasingly naïve. The rise of cyber-hacking, automated attempts by foreign and domestic actors to spread misinformation

on social media platforms and the mis-use of personal data by campaigns all appear to be increasingly regular features of contemporary elections.

Craig A. Hill, Ph.D., is Senior Vice President, Survey, Computing, and Statistical Sciences at RTI International . He has more than 30 years of experience in social science research, directing research projects both large and small for a wide variety of federal, academic, and commercial clients. He has published and presented papers related to social science methods, including hospital ranking methodology, interviewer fraud, new technology for social science research, and social media in survey research.

Sunshine Hillygus, Ph.D., Professor of Political Science, Duke University, has published widely on the topics of American political behavior, campaigns and elections, survey methods, public opinion, and information technology and politics. She is director of the Duke Initiative on Survey Methodology and co-director of the Polarization Lab.

Shanto Iyengar, Ph.D., Professor of Political Science, Stanford University, is Director of the Political Communication Laboratory. Iyengar's areas of expertise include the role of mass media in democratic societies, public opinion, and political psychology.

Olena Kaminska, Ph.D., is a Research Fellow at the Institute for Social & Economic Research, University of Essex and a survey statistician for the UK Household Longitudinal Study. Her current research focuses on improving quality and efficiency of survey data, specifically within three broad themes. The first theme is related to motivation and its role in improving quality of survey answers. The second theme is concerned with tackling nonresponse through fieldwork, including through motivation and adaptive design. And the third investigates improvements of statistical estimation and correction for nonresponse in complex sample design situations, including in longitudinal and cross-sectional surveys. Her wider research interests also include work on satisficing, real-world eye-tracking in survey context, nonresponse bias, sample design, and social desirability.

Courtney Kennedy, Ph.D., is director of survey research at Pew Research Center. In this role, she serves as the chief survey methodologist for the Center, providing guidance on its survey research and overseeing the Center's national, online survey panel. Her research focuses on reducing errors in public opinion surveys. Recent projects examine how survey recruitment and weighting affect data quality.

Frauke Kreuter, Ph.D., Professor, Joint Program in Survey Methodology, University of Maryland, is Co-Director of the Social Data Science Center. She is also Professor for Data Science in the Social Science and Humanities at the Ludwig Maximilians University in Munich; and head of the Statistical Methods Research Department at the Institute for Employment Research (IAB) in Nürnberg, Germany. Her research focuses on sampling and measurement errors in complex surveys. In her work at JPSM she maintains strong ties to the Federal Statistical System, and served in advisory roles for the National Center for Educational Statistics and the Bureau of Labor Statistics.

Arthur Lupia, Ph.D., Gerald R Ford Distinguished University Professor of Political Science, University of Michigan, is Assistant Director of the National Science Foundation and leads their Social, Behavioral, and Economic Sciences Directorate. For the White House Office of Science and Technology Policy, he is co-chair of the Subcommittee on Open Science. At the University of Michigan, he is the Gerald R Ford Distinguished University Professor. His research examines how people make decisions when they lack information. He has worked with organizations around the world to improve quality of life through better management of complex information flows. His topics of expertise include information processing, persuasion, coalition building, and strategic communication.

Amy O'Hara, Ph.D., Research Professor in the Massive Data Institute, Georgetown University, is Executive Director of the Federal Statistical Research Data Center at the McCourt School for Public Policy. She also leads the Administrative Data Research Initiative, improving secure, responsible data access for research and evaluation. O'Hara addresses risks involved with data sharing by connecting practices across the social, health, computer, and data sciences.

Stanley Presser, Ph.D., Distinguished University Professor of Sociology, University of Maryland, is interested in the interface between social psychology and survey measurement. His research focuses on questionnaire design and testing, the accuracy of survey responses, nonresponse, and ethical issues stemming from the use of human subjects. His books include Questions and Answers in Attitude Surveys (with Howard Schuman), *Survey Questions* (with Jean Converse), and *Methods for Testing and Evaluating Survey Questionnaires* (chief editor). In addition to being Professor of Sociology, he teaches in the Joint Program in Survey Methodology, which he founded in 1992 with colleagues at the University of Michigan and Westat, Inc.

Nicholas Valentino, Ph.D., Professor of Political Science, University of Michigan, focuses his research on political campaigns, racial attitudes, emotions, and social group cues in news and political advertising. His current work examines the intersection between racial attitudes and emotion in predicting participation and vote choice.

David Vannette, Ph.D., is a research scientist largely focused on survey research methods, data science, and research-on-research at Facebook. His substantive research focuses on understanding social phenomena, including political communication, political psychology, and public opinion.