This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

October 8, 2010

This work is licensed under the Creative Commons Attribution-NoDerivs 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nd/3.0/ or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

Measurement and Experimentation in the Social Sciences

Arie Kapteyn

RAND

Abstract

I propose to build an advanced data collection environment for the social sciences that maximizes opportunities for innovation, and is fast, cost effective, and easy for everyone in the scientific community to use. The core of this "laboratory" is a representative panel of households in the United States who have agreed to be available for regular interviews over the Internet. The Internet panel is representative in the sense that respondent recruitment is based on a probability sample. Internet access will not be a prerequisite for participation in the panel. If a respondent does not have Internet access at the time of recruitment into the panel, he or she will be provided with a laptop and broadband access. The laboratory will incorporate and pioneer new forms of data collection— including, but not limited to, smartphones, self-administered measurement devices for the collection of biomarkers, experience sampling, web cameras, Global Positioning System (GPS) devices, accelerometers to measure physical activity, and eye tracking equipment.

The challenge of measurement

Social scientists use many sources of information to construct their models of human behavior in a social and societal context. These sources include introspection, participatory observation, surveys, physical measurements and biomarkers, administrative data, laboratory experiments, field experiments, and natural experiments.[1]

Every source of information has strengths and weaknesses. Which source is used may depend on the research question at hand, but in many instances also reflects the personal preferences and skills of researchers and the size of research budgets. More importantly, measurement is typically limited to one domain or at most a few. Thus numerous studies are conducted each addressing one of a variety of domains of human life but mostly ignoring the relationship with other domains. This situation can partly be attributed to budgetary limitations: After all, who has money to study "everything"? Yet in fact, resources are wasted because different studies often collect overlapping information

[1] This list is not an exhaustive typology and several finer distinctions can be made, e.g. between experiments with real stakes and hypothetical experiments, panel surveys and cross sectional surveys, etc.

This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

while missing opportunities to capture a more complete understanding of human behavior in its various aspects. In short, empirical social sciences tend to be fragmented, not only because of disciplinary differences in approach and the communication issues associated with these differences, but also because data themselves are fragmented.

Admirable exceptions exist: The US Health and Retirement Study (HRS, http://hrsonline.isr.umich.edu/) is the most prominent example[2]. Since 1992, the HRS has collected information biennially on individuals 50 and older about income, work, assets, pension plans, health insurance, disability, physical health and functioning, cognitive functioning, and health care expenditures. Additionally, off-year surveys cover topics such as the Consumption and Activities Mail-Out survey (CAMS) and one-off surveys on topics such as Medicare part D and time use. Recently (since 2006), HRS has started to collect biomarkers such as grip strengths, breathing tests, saliva (to extract DNA), dried blood spots (for Hemoglobin A1c, total cholesterol and HDL cholesterol), etc. In addition, administrative data (e.g. social security earnings records) have been added, subject to consent by respondents. The HRS is such a successful scientific model that it has been reproduced in some 20 countries (England, several continental European countries, South Korea, Mexico, China, and India), with largely similar set-ups and comparable questionnaires.

Despite the collection of additional information in off years and the growing breadth of information that is being collected (like biomarkers), there are obvious limitations to the amount of information that can be collected in any single domain and to the number of experiments one can do. Thus although the HRS has been revolutionary in its multidisciplinary approach and in its continuous incorporation of innovations, many factors continue to limit what we can be learn (including the age requirement for respondents).

What would be a next step?

Imagine a survey like the HRS that would allow researchers to recontact study participants at any time, include all ages, combine conventional surveys with physical measures and biomarkers use modern technology to monitor behavior, and allow for data collection across a broad swath of domains and over an extended time period. Naturally we would also want to link administrative data to individual records (subject to respondents' consent and with adequate data protection safeguards). This kind of survey would allow researchers across multiple disciplines to consider all relevant domains for empirical analysis, to quickly monitor the effects of major events (e.g. the financial crisis or the swine flu pandemic), and to design experiments that take advantage of a wealth of readily available background information. The format would also support in-depth studies on subsamples (e.g., conducting functional Magnetic Resonance Imaging experiments or qualitative interviews with a subset of study participants).

[2] This is not to say that other surveys don't cover material from different disciplines. For instance, the PSID, which started as primarily a socio-economic panel has added content over the years and now has a substantial health component.

This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

What is possible today?

My proposal is not to build something totally new with unproven technology, but rather to build on what has been proven to work, using technology that currently exists or is right around the corner. I propose to build a virtual laboratory, an advanced data collection environment for the social sciences that maximizes opportunities for innovation, and that is fast, cost effective, and easy for everyone in the scientific community to use. The core of this laboratory is a representative panel of households who have agreed to be available for regular interviews over the Internet. The Internet panel is population representative in the sense that respondent recruitment is based on a probability sample. Internet access is not a prerequisite for panel participation: If a respondent lacks Internet access at the time of recruitment, he or she is provided with a laptop and broadband.

The virtual laboratory will develop and test new modes of data collection and the collection of new types of data, including, but not limited to, self-administered measurement devices for the collection of biomarkers (e.g., infrared blood sugar monitors), web cams, accelerometers and heart rate monitors for measuring physical activity and physiological responses, devices for experience sampling, Day Reconstruction Methods, intensive methods to increase both unit and item response, preloading, and data quality checks. Current technology allows respondents to participate in surveys using their preferred hardware device, such as a netbook, desktop computer, iphone or other smartphone, or any other device like the browser on a game console.

It is neither possible nor useful to describe the many kinds of data that might be collected. The idea is that the laboratory we propose will be able to follow new technological and scientific developments, without committing to one particular technology *ex ante*. A few examples will illustrate the point.

Example 1, GPS tracking: With more and more cell phones equipped with GPS, GPS tracking is becoming more sophisticated and yet more affordable. Software is now available or can be easily developed to track a respondent's GPS-enabled cell phone from the web and combine it with real-time location based information. This combination would have many possible applications, such as allowing researchers to initiate a small survey by text messaging questions to a respondent when s/he leaves the gym or visits a tax consultancy office.

Example 2, Eye tracking: In recent years, the increased sophistication and accessibility of eye tracking technologies have generated a great deal of interest in the commercial sector. Most applications focus on web or software usability, presenting a target stimulus to a sample of consumers while an eye tracker is used to record the activity of the eye. By examining fixations, saccades, pupil dilation, blinks, and a variety of other behaviors, researchers can determine a great deal about the effectiveness of the web or software interface. This technology can be easily adopted and used to test alternative interviewing techniques and to examine respondent behavior during an interview, e.g. to gauge which information on a screen is actually taken into account when answering a question.

This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

Example 3, Accelerometers to measure physical activity: An accelerometer is a device that measures proper acceleration, the acceleration experienced relative to freefall. Accelerometers are increasingly being incorporated into personal electronic devices like the iPhone and allow researchers access to objective measurements of physical activity. These measurements can be retrieved in real time or can be uploaded to a central location when the respondent has access to a computer, allowing for follow-up questions based on the measured activity.

Example 4, Telemetry: Telemetry is a technology that allows the remote measurement and reporting of information of interest to a central location for further analysis. Thus it can be used to link the output of all these new technologies.

Example 5, Integrating survey information with social network information: Having access to social networking sites like Facebook (only with a respondent's permission, of course) provides researchers with ample information about a respondent without actually asking questions. This technology can help reduce the respondent burden, gives the respondent more flexibility and a familiar interface, and allows for consistency checks based on the data retrieved from the social networking site.

What exists today?

MESS

An existing facility that comes closest to what I am proposing is the MESS project in The Netherlands (http://www.centerdata.nl/en/TopMenu/Projecten/MESS/index.html).[3] The core element of the MESS project, currently about mid-way through its first seven years of funding, is the so-called LISS panel (Longitudinal Internet Studies for the Social Sciences). The LISS panel consists of approximately 5,000 households representing the Dutch-speaking population in the Netherlands. The panel is based on a probability sample drawn from the population registers. Households without prior Internet access are provided with broadband access (and a PC) to participate. The LISS panel has been fully operational since early 2008 and has now collected two years of data.
Annual interview time is about 300 minutes. Panel members complete relatively brief (30-minute) online questionnaires monthly and are paid for each completed questionnaire. Half of the interview time is reserved for the LISS core study. This core study is repeated yearly (spread out over several months) and borrows from various national and international surveys to facilitate comparison with other data sources. The core survey covers a much broader range of topics and approaches than would be possible with other surveys using more traditional interview methods. The remaining interview time is used for experiments and innovation: Respondents can complete online questionnaires at any time during the month.

[3] For the sake of full disclosure, I am one of the principal investigators of the MESS project. I am also the director of the American Life Panel, discussed below.

This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

The application and review procedures for experiments are similar to those of TESS (see below), but there is no *a priori* restriction on the size or duration of the experiment that one can propose. In the first two years, about 40 proposals for experiments were accepted.

TESS

TESS (Time Sharing Experiments for the Social Sciences) is somewhat similar to the MESS project in its use of a standing Internet panel, Knowledge Networks. The panel is available at no charge to researchers who complete an application. The TESS web-site lists about 125 papers based on experiments conducted with the panel between 2003 and 2008. MESS and TESS do have some notable differences: TESS does not collect much core information about the panel members, except for basic demographics, and the number of items in a questionnaire as well as sample sizes are strictly limited (essentially, the more items, the smaller the sample size). Also, unlike LISS (MESS), TESS considers proposals only for experiments, not for regular surveys. Nevertheless, TESS services are clearly in demand.

American Life Panel

The RAND American Life Panel (ALP) is similar to the Knowledge Networks and the LISS panel in its reliance on a probability based sample and its ability to include respondents without prior Internet access by offering a laptop and Internet subscription. The panel currently includes approximately 3000 US households, with firm plans to increase the number to 5000 (including a Spanish language subpanel). Since 2007, some 120 experiments or surveys have been conducted. The HRS survey instrument has been programmed for the ALP and administered to the ALP respondents, so the full HRS core information on all panel members is available. Use of the panel for surveys or experiments is open to all researchers, but is not free[4].

The ALP is used intensively (approximately three surveys or experiments per month) and one might worry about survey fatigue and hence increased attrition. The annual attrition is between 5 and 6% a year. The low attrition rate may be partly due to the relatively generous incentives offered to respondents ($20 per half hour of interview time). Occasional comparisons with other surveys about similar topics show broad consistency. Data are disseminated through a web-site that allows free download of datasets, the construction of a custom made dataset by combining variables from different waves and putting them in a "shopping cart". One can download data at any time during or after the field period. (https://mmicdata.rand.org/alp/index.php/Main_Page)

[4] A substantial part of the experiments and data collection is supported by grants from the National Institute on Aging. Other major funders are the Social Security Administration and several non-profit institutions. The pricing is $3 per interview minute for the first 500 respondents, $2.50 for the next 500 and $2 per interview minute beyond 1000 respondents.

This paper was submitted to the National Science Foundation as part of its SBE 2020 planning activity (www.nsf.gov/sbe/sbe_2020/). Its inclusion does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government.

Conclusion

A laboratory as proposed will both dramatically expand opportunities for social science research and be highly cost effective. The technology exists; we only have to put it together.