Project Implicit: Advancing theory and evidence with technical and methodological innovation

Brian A. Nosek, Ethan Sutin, Jeffrey J. Hansen, Lili Wu, N. Sriram, Frederick L. Smyth

University of Virginia

Mahzarin R. Banaji

Harvard University

Anthony G. Greenwald

University of Washington

Abstract

Just as the invention of specific technologies dramatically changed the course of the older sciences, the Internet has the potential to transform what the behavioral, social, and mind sciences can accomplish in the 21st century. Project Implicit represents a case study of a single modest effort in which, over the course of a decade, five million tests of implicit social cognition were administered to drop-in participants. From this experience, a broader interdisciplinary collaboration was initiated, providing a case example of using technical and methodological innovation for accelerating and enhancing scientific research. The project's Virtual Laboratory provides a nexus for research collaboration, a virtual workbench for developing and managing web-based research protocols, and a robust infrastructure for administering experiments. With collective effort, the Internet will transform the research process by fostering collaboration, expanding access to samples all over the globe, facilitating data integration, and enabling a common framework for scientific research.

Abstract = 150 words

Keywords = Internet, research methodology, measurement, experimental design, innovation, implicit cognition

More than researchers would care to admit, study design is guided by practical factors of what study design *can* be done, rather than how it *should* be done. In an ideal research world, a scientist would conceive a question, generate the best paradigm to test it, and then execute the study. The "perfect study," a mythical, Platonic ideal, is the source of scientific striving. Such a study would allow strong inference, randomly select from the population of interest, and have unassailable validity.

Teachers of undergraduate research methods occasionally confront study designs that approach such perfection. These examples will usually elicit the red pen response: "Excellent design and completely impractical." The undergraduate's solid grasp of validity and inference confronts the pragmatic realities of actually doing the research. Only a semester of graduate school is needed to purge the mind of such magnificent designs. Seasoned researchers' thoughts are bound by feasibility – cost, time, complexity, expertise, and availability of samples, methods, or measures.

On a collective scale, practical constraints limit the pace and effectiveness of empirical research. Complex problems are tested with many small, simple designs holding many variables constant, or manipulating very few levels of the independent variable(s). The ubiquitous 2X2 experimental design reflects the typical balance between retaining some degree of complexity against the realities of time, cost, and statistical power. However, such constraints make it difficult to identify interactive effects among multiple variables, non-linear patterns across the range of an independent variable, and other less constrained outcomes. Practical demands researchers focus on finding effects rather than clarifying the conditions under which an effect does or does not occur. Bigger studies can identify such conditions, but it is not clear in advance whether the added resource expense will bring commensurate payoff in the findings. Most research, therefore, is conducted on a small scale with simple designs, minimizing the cost of failure, but also crimping the returns on success.

For practicality's sake, some researchers use a narrow set of well-understood methods. A significant portion of the research literature on persuasion, for example, evaluates undergraduates' reactions to instituting comprehensive exams. With repeated use of this paradigm, researchers are prioritizing cumulative knowledge with a single topic and method, fully aware that such a choice may limit the generalizability of the discovery. This provides a limited, but more certain, understanding of a particular aspect of persuasion. For each new study, the researcher has to make a difficult decision of sticking with an established paradigm versus risking time, effort, and unanticipated confounds with a novel one.

Research topics are often selected based on the availability of a means to investigate them. It is usually easier to collect data with undergraduates than other samples, easier to run fewer participants than more, easier to run single-shot studies than longitudinal designs, easier to use a few methods than many, and easier to use an existing measure that is not perfectly appropriate than to develop a new measure that might be. It will surprise no one that studies using several such components of convenience dominate research and journal space.

Even when the behavior is not one that is likely to occur in sample A, we test sample A. For example, studies of hiring and jury decision making often use undergraduates who, more likely than not, have never sat on a jury or hired someone. Is this because researchers have determined that undergraduates are the best sample population to answer these questions? No. Researchers must make a calculated risk assessment: Are there study design factors that will interact with characteristics of the sample? Does the theoretical question demand a sample of people who hire or sit on juries? Is the expense of recruiting professionals worth it? Many times, doing such a study with undergraduates is a reasonable decision, and at times it may be an ideal one, but not always. If hiring professionals and actual jurists were readily available, they would surely be the sample of choice.

For every study, researchers face practical decisions about approximating the ideal design given the available resources. This article is not a call to rise above pragmatics and conduct only ideal studies. So far as we know, only thought experiments and philosophers can achieve that. We argue that the social and behavioral sciences cannot gain by masking, denying, or ignoring the existence of practical limitations on the impact of the work and the pace of discovery. The goal of this article is to describe technical and methodological innovations that face practical challenges directly with the goal of enhancing research progress in the social and behavioral sciences.

Technological and methodological innovation as a means for reducing practical constraints on research

Sometimes a new theory or idea spurs a burst in scientific knowledge. More often, it is a new way to measure, or a method, that offers a new way to look at the world. The history of science is marked by big innovations (telescope, electron microscope, fMRI) and small ones (petri dish, aperture size, tagging darts) that produced new learning, even unanticipated or unwelcome learning. The Internet has similar potential to fundamentally transform the research process and reduce a particular set of the pragmatic barriers that confront the mind sciences.

An interdisciplinary effort to exploit the Internet to enhance social and behavioral research

The focus of this paper is a multi-laboratory, interdisciplinary collaboration called Project Implicit, geared toward pursuing technological and methodological innovation to advance behavioral science. Eight project principles for developing Internet-based infrastructure emphasize reducing pragmatic barriers and increasing flexibility in striving for the ideal study:

1. *Accessibility*: Increase access to participants, materials, designs, and methods
2. *Automation*: Automate the repetitive components of the research process, and focus researchers' time and attention to the generative aspects of science
3. *Information integration*: Facilitate combining of data sources, fostering a cumulative science
4. *Collaboration*: Remove geography as a meaningful barrier to collaboration; simplify sharing of ideas, materials, methods and data
5. *Usability*: Capitalize on researchers' existing skills and minimize need for additional training
6. *Extensibility*: Consider future development and growth; emphasize general frameworks to support a wide variety of paradigms
7. *Scalability*: Infrastructure should support growth by orders of magnitude without redesign

8. *Sustainability*: Development and operations model should emphasize self-sustainability rather than depend on perpetual subsidization

The remainder of this article describes the first phase of an interdisciplinary effort to leverage the technological possibilities of the Internet to advance social and behavioral research. We first illustrate the potency of the Internet for advancing a research enterprise using implicit social cognition as an example. We then describe the Virtual Laboratory, the infrastructure that makes this research enterprise possible. We close with a vision for the future of research practices bolstered by Internet-based innovation.

A case example of leveraging technological and methodological innovation for empirical research in implicit social cognition

Project Implicit (http://projectimplicit.net/) is a multi-university collaborative enterprise, pursuing technological, methodological, and theoretical innovation in the investigation of implicit cognition. Project Implicit is a network of laboratories, technicians, and research scientists at Harvard University, the University of Washington, and the University of Virginia.

Project Implicit's origins are found in the launch of a web site in 1998 at Yale University to provide the equivalent of a science museum exhibit demonstrating the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). The IAT measures association strengths between concepts (e.g., young and old) and attributes (e.g., pleasant and unpleasant). The IAT has the conjoint qualities of being a productive research tool, and providing a provocative, palpable experience of response conflict for the test taker. For example, most respondents are able to sort exemplars of the concept and attribute categories faster when *young faces* and *pleasant* words are categorized with one behavioral response (a key press) and *old faces* and *unpleasant* words are categorized with an alternative response, compared to when *young faces* are categorized with *unpleasant* words and *old faces* with *pleasant* words. The differential response conflict between these two sorting conditions is taken to reflect association strengths between the concepts and attributes. Most respondents have an overt experience of finding it easier to sort the items in one condition compared to the other. This experience can be especially surprising when it suggests mental associations that run counter to one's explicit values or beliefs (see Nosek, Greenwald, & Banaji, in press for a review).

The launch of a public web site afforded an opportunity to engage the public in the on-going investigation of implicit biases and promote critical thinking about basic research, methodology, and evidence. The most optimistic estimate of the principal investigators was that 50,000 people would experience an IAT during the website's existence. That value was selected as roughly the equivalent to the number of participants that would pass through a bricks-and-mortar laboratory during a productive investigator's academic lifetime. This most-optimistic estimate required revision after the 50,000th visit to the website occurred on its second day of operation.

The surprising success launched an investment to extend the capabilities of the website into a robust environment for research. In the 10 years since the launch, Project Implicit's infrastructure has grown into a general study management, development, and administration engine. Multiple entry-points enable public access to a variety of demonstration studies, registration access to a virtual research laboratory, and private access for studies that identify participant populations in advance. The websites have administered more than 5 million study sessions to visitors, and the rate of participation has grown over time. In 2004, there were 466,740 completed study sessions. In 2005, there were 1,129,687. As of mid-2006, the sites were averaging about 20,000 completed study sessions per week across more than 140 active studies. With a network of international collaborators, the main website is served in a variety of languages including Chinese, Dutch, French, German, Hebrew, Italian, Japanese, Korean, Polish, Romanian, and Spanish making the research available to a majority of the world's population that has access to the Internet.

The substantive research in implicit social cognition is an application of the general Virtual Laboratory infrastructure. This infrastructure opens up a variety of methodological and design possibilities that would not otherwise have been feasible because of practical constraints. As a consequence, the Virtual Laboratory complements and extends the research possibilities beyond traditional laboratory methods. Next are some examples illustrating how Internet methods have increased the efficiency and capability of research in implicit cognition.

Separating study sessions

It is often desirable to split a study into two sessions so that responses on one set of measures do not contaminate responses on a second set. However, because of the added challenges of coordinating return visits of participants, a common, imperfect fix is to collect all the data in a single session with a break in between that is intended to clear the mental palette. Implementing half of the study on-line can make it easy to do a single lab session, and still separate measurement occasions. For example, Mendes, Gray, Mendoza-Denton, Major and Epel (2006) investigated the moderating role of implicit racial bias to neuroendocrine reactivity during interviews with Black or White interviewers. For the design, it was important that the implicit measures be administered separately from the interview session. Recruited participants were directed by email to a website that administered the implicit measures, and later came to the laboratory for the interview segment. Participants completed the web measures on their own time rather than coordinating schedules with the experimenters, and there was less concern about contamination between measurements.

Longitudinal study

Tracking development over days, months, or years is highly valued and resource intensive. These demands include the mundane, but critical coordination of many visits to the laboratory, and connecting data across multiple sessions. Sherman, Chassin, Presson, and Macy (2006) have been following the smoking attitudes and behavior of a sample originating in the Bloomington, Indiana region since 1980. With ambitious project goals including not only longitudinal study of the original sample but also their children, and wrestling with geographic dispersion, these researchers introduced a web-based component of the data collection for questionnaires and IATs measuring implicit smoking attitudes. This approach saved resources that could instead be dedicated to maintenance of the sample, and more targeted laboratory data collections for measures that cannot be administered at a distance.

In a mini-longitudinal study, Ranganath and Nosek (2006) invited volunteers who had completed an initial attitude induction session to return to the web site three days later for a follow-up assessment. In this case, the all-volunteer sample never visited the laboratory and tracking participants was automated by Project Implicit's registration system (described later).

Rapid data collection

Some studies capitalize on cultural events or naturally occurring phenomena that have a punctuated time frame for measurement. It can be difficult to recruit participants and collect data quickly in the laboratory. To compensate, researchers will sometimes simplify study designs so that many people can be run together. Automating data collection with web-based methods allows many participants to be run at once without oversimplifying the designs.

Lane and Banaji (2005) investigated the malleability of implicit national and local identities in response to social events. Boston residents were recruited online in three one-week time periods, right after the Red Sox won the baseball World Series in 2004 (N=202), immediately before the season began in 2005 (N=297), and just after playoff elimination in 2005 (N=197). Participants visited a website and completed implicit and explicit measures of national and local identity. Administering this study in the laboratory would have been very difficult because of the design requirement to measure participants in one-week windows, and the expense of inducing participants (residents of the Boston area) to travel to a Harvard lab for a 30 minute study.

Recruiting special samples

Some research questions require participation of samples with very specific characteristics that are either hard to find, or are not centralized geographically. A common strategy for doing research with these samples is to recruit and administer studies by mail. In these cases, researchers are forced to limit the measurement methods to those that can be self-administered with paper and pencil. The Internet provides more technical flexibility for implementing many varieties of methods to a dispersed sample.

Sharp, Monteith and Banaji (2005) investigated change in implicit cultural attitudes among parents involved with cross-cultural adoption. In collaboration with an adoption agency, participants were recruited for two web-based studies via email (Total $N = 231$). All together, these studies required approximately two hours of researcher effort for recruiting, data collection and data entry because those components were automated. Further, the studies employed relatively complex methods, such as response latency measures, that could not have been administered via a postal mail packet.

Automated preselection

One of the most effective uses of university participant pools involves administration of preselection measures for many studies at once and allows researchers to identify specific sub-samples from the larger pool. Without preselection, the challenges of identifying a specific sample would make many studies unfeasible. Project Implicit's research study pool extends this idea by implementing a web-based preselection mechanism. Volunteer participants register for the study pool by creating a login identity and answering a demographics questionnaire (https://implicit.harvard.edu/). Participants can be automatically and unobtrusively selected for studies each time they enter the site based on their earlier responses.

Teachman, Peris and Nosek (2006) investigated the effects of stigma of mental illness among mental-health practitioners. After less-than-successful efforts at direct recruiting of practitioners, the study was added to the Project Implicit study pool. Site visitors who reported a census-defined occupation category including mental-health practitioners (from a list of 100) were automatically routed to the study. Even though this was a very specific sub-population, the daily traffic to Project Implicit was large enough to make this an effective recruiting strategy. Further, because recruitment and data collection was automated, the study could remain active as long as needed to reach the study sample size goals. After 5.5 months in the study pool, 638 participants had been identified and completed the study.

Large samples

The most obvious benefit of web-based data collection is the potential to increase the size and power of study designs. In many laboratories, participants are a scarce resource and the primary limiting factor on research productivity. Determining how many lab participants are necessary for sufficient statistical power is a critical – though frequently ignored – design element (Cohen, 1992).

Even when participants are plentiful, there are still practical concerns over the effort required to administer the study to more and more participants. There is a strong relationship between the number of participants and amount of lab resources expended for data collection: 40 participants require almost twice as much effort as 20, 80 nearly twice as much as 40. Automation of data collection can alter this relationship dramatically: once a study has been implemented, often there only small differences in comparative effort for running 10, 100, or 1000 participants.

When samples are orders-of-magnitude larger than the typical laboratory samples, power can be nearly 100%, trivializing significance testing and enabling an exclusive focus on effect size. Greenwald, Nosek, and Banaji (2003), for example, took advantage of the very large data sets of various IATs accumulated at Project Implicit websites to compare alternative scoring algorithms against a variety of criterion measures. The high power resulted in reliable estimates that could distinguish candidate algorithms across multiple parameters.

Nosek and colleagues (Nosek et al., 2006) summarized more than 3 million study sessions of implicit and explicit attitude and stereotype data collected at Project Implicit websites for 17 different content domains from 2000 to 2006. These data provide a useful complement to dozens of laboratory investigations, and tested questions that could not be examined any other way.

Cross-cultural and cross-regional comparisons

An entire field of research has grown out of the discovery that theory and evidence developed in American or European cultures does not necessarily generalize to other cultural contexts (Doob, 1980, 1988; Fiske, Kitayama, Markus, & Nisbett, 1998). Cultural psychology emphasizes the importance of considering contextual factors in psychological theory. Appreciation of this influence introduces a non-trivial problem of how to deal with it in the everyday activity of research. Most sampling is geographically-bound to a single university or locality. And, coordinating data collection at multiple sites, especially in multiple languages, is resource intensive. As a consequence, attending to possible contextual or cultural factors in research findings requires a very deliberate decision to devote significant resources to making those comparisons.

The reach of the Internet dramatically reduces the challenges of conducting studies across geographic boundaries. Internet access, not geography per se, is the limiting factor. Within this constraint a study can be made available as widely or narrowly as desired. While selection effects related to Internet access must continue to be seriously considered (e.g., more than two-thirds of North Americans are on the Internet, but access in Africa is estimated at under 3%; Internet World Stats, 2006), Internet access continues to grow rapidly. Traditionally underrepresented groups are rapidly becoming Internet users including: 32% of U.S. adults aged 65 or older, 53% of adults in households with less than $30,000 in annual income, and 40% of adults with less than a high school education as of early 2006 (Pew Internet and American Life Project, 2006).

The international reach is evident in publicly available research. At Project Implicit web sites, for example, 15% of participants were citizens of a country other than the United States when the sites were English-only (Nosek et al., 2006). With the addition of a network of international collaborators who developed country and language-specific versions of the websites and studies (16 countries as of August, 2006), the percentage of international visitors is increasing.

Seizing upon an opportunity for cross-cultural comparisons with existing data, Nosek and Smyth (2006) found a robust relation between implicit gender-science stereotype and national gender differences in science achievement across 21 nations. Standardized science test scores for 8th-graders were available for each nation (the TIMSS; Gonzales, 2004) and each had at least 100 self-reported nationals in Project Implicit's gender-science task. Greater national differences in boys' and girls' science performance was associated with higher mean levels of implicit stereotype—science is male/liberal arts is female—among site visitors from these countries, $r = .59$ ($p = .005$). Further, the performance differences did not correlate with explicit stereotypes, $r = .12$ ($p = .60$). This and similar cross-national comparisons with age attitudes (Lindner, Nosek,

& Banaji, 2006) and U.S. congressional district comparisons with racial attitudes (Nosek, Thompson, & Banaji, 2006) illustrate unique applications of broadly-based data collections, and the potential to integrate such data with existing census, voting, or other national and international datasets.

Another example illustrates the challenges for generalizability from findings with homogeneous participant samples. Richeson and Nussbaum (2004) found that racial attitudes were more egalitarian following a multi-cultural prime compared to a color-blind prime (see also Wolsko, Park, Judd, & Wittenbrink, 2000). In an attempt to identify the causal factors underlying this result, Smyth and Nosek (2007) repeated the study on the Internet with a larger, more heterogeneous sample ($N > 1500$). With difficulty replicating the effect and assistance from Richeson in theorizing why, they discovered that the 'original' effect may be specific to contextual characteristics of its relatively homogeneous, i.e., "very white" Dartmouth environment. When whites in the Internet sample were grouped according to their self-reported substantive contact with blacks, findings similar to Richeson and Nussbaum's were observed among the low-contact participants, while no effect of prime condition obtained among the high-contact participants. Identifying this interaction might not have been possible without the greater diversity of experience represented in the Internet.

Complex study designs

All laboratory experiments require practical decisions about which variables to hold constant and which to manipulate and the number of levels of the manipulation. The tradeoffs in these decisions usually pit statistical power against possible loss of generalizability. For example, the order in which two measures are administered may have an unintended influence on their relationship. Manipulating the order could reveal the influence, but at the expense of statistical power. Likewise, finding a relationship between variables in one domain may not apply in another domain. The resource costs of testing multiple domains must be weighed against the immediate and long-term threats to validity and generalizability.

A case in point is the observation that correlations between implicit and explicit attitudes vary from weak to strong depending on content domain. Nosek (2005) investigated the relationship between implicit and explicit attitudes in a study in Project Implicit's study pool. Participants were randomly assigned to complete implicit and explicit attitude measures for one of 57 different content domains. The study also contained a wide variety of manipulations to identify the influence of theory-irrelevant procedural factors on implicit-explicit relations. Finally, the study examined four potential moderators of implicit-explicit correspondence and found that all four accounted for variability in implicit-explicit relations even when considered simultaneously. Conducting this study via Project Implicit allowed for highly powered tests of a very complex experimental design ($N > 11{,}000$), and more confidence that the effects were not due to peculiar features of a single or small set of content domains.

Using traditional laboratory methods, moderators are usually examined one at a time because it is difficult to justify the resource expenditure to manipulate multiple moderators at once. This increases the difficulty of determining whether moderators are redundant, even meta-analytically, because they rarely appear in the same study. Automation of study administration eliminates some challenges to embracing complexity, and improves the facility for strong inference.

Validity of Internet Research

Younger, richer, better educated, non-rural people continue to be overrepresented among Internet users, but demographic characteristics of users are progressing rapidly in the direction of overall population parameters, and more than 70% of U.S. men and women are now accessing the Internet (Pew Internet and American Life Project, 2006). As always, caution must be used in generalizing results from any non-representative sample, and the Internet does not obviate this necessity. This is particularly true when interpreting the means and variances of the variables that may covary with the decision to engage with an Internet study. While Internet studies can yield large samples with more demographic diversity than other convenience samples (e.g., undergraduates) they may also be more highly self-selected with respect to key psychological characteristics (Best, Krueger, Hubbard, & Smith, 2001; Brenner, 2002).

Internet volunteers tend to be more interested in and involved with the given experimental topics than those recruited by more traditional means and are more likely to hold stronger views (Cooper, Scherer, & Mathy, 2001; Michalak & Szabo, 1998; O'Neil, Penrod, & Bornstein, 2003; Walsh, Kiesler, Sproull & Hesse, 1992). Furthermore, the characteristics of a Web sample can shift as a function of site presentation characteristics and the source from which users found their way to the site (Brenner, 2002; Schillewaert, Langerak, & Duhamel, 1998; Srivastava, John, Gosling, & Potter, 2003).

Despite these self-selection pressures, research with Internet samples that has focused on psychological processes and variable interrelations has yielded results consistent with those derived from either more representative or more typical (i.e., undergraduate) samples, whether obtained via computer or paper-and-pencil (Best, et al., 2001; Birnbaum, 2000; Gosling, Vazire, Srivastava, & John, 2004; McGraw, Tew, & Williams, 2000; Nosek, Banaji, & Greenwald, 2002a; Smith & Leigh, 1997; Smyth, Nosek, & McArdle, 2006). McGraw and colleagues (2000) replicated well-known laboratory effects in five Internet experiments, even though some involved between-subjects tests which carried greater risk of noise from artifacts of different computer hardware and experimental conditions. In a comparison of an Internet sample with a probabilistic telephone sample, Best and colleagues (2001) found congruent political attitudes. Smyth and colleagues (2006) found comparable factor structures for implicit and explicit gender-science stereotypes between a self-selected *Project Implicit* sample and a traditionally recruited undergraduate sample, demonstrating that measurement instruments were functioning in the same way for both samples. Furthermore, they found that the functional relation between identification with the science domain (i.e., majoring in the field) and implicit stereotyping varied for men and women of both samples in accord with the predictions of balanced identity theory (Greenwald et al., 2002; Nosek, Banaji, & Greenwald, 2002b). For women, regardless of sample, majoring in science predicted weaker science=male stereotypes, while for men majoring in science predicted stronger stereotypes.

Summary

Research via the Internet has advanced understanding of implicit social cognition through increased access to samples, flexible research design, automated data collection, and a global reach. These qualities provide a source of data that enables broad confirmation and extension of findings from other sources, and allows investigations that are not practical in a physical laboratory. An important advantage of Internet-based research is the increased opportunity to examine subject populations that are more diverse and harder to reach than typical samples. In addition, the Internet can provide access to samples that would not otherwise be available because of small numbers in a given locale, or brief windows of opportunity for data collection. Though work with Internet samples retains some of the shortcomings of other self-selected convenience samples, solid evidence has accumulated for the validity of the medium for experimental research on psychological processes and construct interrelations. Internet-based technical and methodological innovations have potential application to research domains beyond implicit social cognition. The next section describes the Virtual Laboratory, the infrastructure that made possible the just-described research.

Virtual Laboratory: A framework for research development, administration, and management

Using the Internet to advance research and collaboration

Alongside the commercial fervor generated by the Internet in the 1990's, some behavioral scientists anticipated a transformation in data collection and research forecasting that the Internet could become "the main venue for subject pools in coming years" (Birnbaum, 2000). While use of the Internet for research grows, the pace is far behind commercial applications. For example, a review of articles published in the Journal of Personality and Social Psychology for ones that used the Internet as a vehicle for data collection finds just 1 of 153 in 2001 (2 others used email), 2 of 173 in 2003, and 2 of 133 in 2005 (1 other used email). In Psychological Science there were 0 of 97 in 2001, 1 of 110 in 2003, and 0 of 171 in 2005. And, in Journal of Experimental

Psychology: General there were 0 of 42 in 2001, 0 of 30 in 2003, and 1 of 30 in 2005. This is 7 of 939 articles in total (0.7%), not an impressive number. If the Internet is transforming behavioral research, it is not happening through top journals.

The slow growth of excellent research conducted on the Internet likely reflects the methodological challenges and uncertainties facing the medium. As much as the previous section illustrated the mitigation of practical research challenges with web-based methods and the comparative validity of their results, for any particular research laboratory, the initial practical challenges for moving into web-based methods may appear greater than the long-term benefits.

As a team of technologists and behavioral scientists, our goal is to develop a virtual laboratory that will be a framework for developing, managing, and administering social and behavioral research. Individual researchers should not need to become experts in web technologies. Transformation of the research process will occur when researchers can engage in effective Internet-based research using the training and expertise that they already possess. Technological innovation will spur that transformation.

Overview

Project Implicit, our research and development network, is intended to become a nexus point for collaborative research in the social and behavioral sciences. A key component is the Virtual Laboratory (VL). The VL (a) is implemented on a secure, stable, extensible environment for study administration, data collection, and data storage; (b) provides a virtual workbench that researchers use to develop, manage, and administer studies through a browser in the lab or on the web; (c) improves access to samples, methods, and measures; (d) automates significant parts of the research process; (e) facilitates data integration; and, (f) fosters interdisciplinary and international collaboration. These tools were designed initially for advancing our substantive research work in implicit cognition. Our current development focus is on generalizing these tools for other research applications. This section describes the VL, and the next section charts a vision for Internet-based technical and methodological innovation in the service of scientific research.

Hardware infrastructure

Participants completed more than a million sessions last year on Project Implicit. The site's extreme popularity requires hardware with the capacity to securely and reliably administer studies to tens of thousands of subjects simultaneously. Spreading the connections across many servers prevents the system from becoming overwhelmed. The redundancy also insures that the failure of a single component will not impact participants.

Figure 1 illustrates Project Implicit's hardware architecture for study administration. A separate environment exists for study development and testing. A load balancer routes connections to an available application server that processes participant interaction during the study session. The application server communicates with a database cluster to store the resulting data as well as fetch relevant information from previous sessions.

The large volume of data and heavy usage requires a highly sophisticated database cluster consisting of several interconnected database server nodes. Fiberoptics link the nodes to a multi-terabyte fault-tolerant storage array. The entire database is regularly backed up to tapes stored off site to ensure recoverability in case of catastrophic failure. The backup system and structural design are scalable to accommodate continuing growth of the project, especially as more researchers develop and deploy studies.

Software framework

The Virtual Laboratory has multiple software components for research management, development, and administration. The user interface is a web-based, point-and-click virtual workbench that includes a private research environment with access to libraries of materials and methods, and tools for creating, maintaining, and archiving studies. Version control and a file-sharing system facilitate record-keeping, reusability, and collaboration. Another component is a collection of tools for tasks that comprise a study: questionnaires, instruction pages, scrambled sentence tasks, the Implicit Association Test, etc. Tasks are organized into studies and managed by study administration software. This software operates the experiment session, monitors participant behavior, and securely transfers data to the system databases. The components are described next and organized in terms of their function: study management, study development, and study administration.

Study Management

In many laboratories, study management is haphazard. A study is conceived, materials are prepared, IRB consent is obtained, data is collected, analyses are performed, papers are written, and these components are idiosyncratically stored in some combination of hard-drives, file cabinets and boxes. Perhaps because of these typical management mechanics, article method sections may not provide a comprehensive account of study design. Correspondence with a researcher often reveals nuances (or more) that did not appear in print and could be important influences on the observed effects. And, with time, many of the details of a study design are lost or forgotten. Failures to replicate across laboratories could either indicate that the first study was a false positive or that the original study conditions were inadequately reproduced. Further, secondary data analysis is much less frequent than it ought to be, in part, because of practical barriers to gathering, describing, and distributing old datasets. Meta-analysis is a popular secondary analysis approach, and is a simplistic version of what is possible if raw data were easily available. The VL is designed to establish a common, simple, and effective study management framework with a collection of tools: a virtual workbench, version control, and a buddy system for collaboration. These tools will help reduce inefficiencies in replication, accurate dissemination of research paradigms, and data sharing.

Virtual workbench. The VL is housed on servers at University of Virginia and is available via the Internet. When a researcher becomes a member of the VL (http://projectimplicit.net/) he or she receives a login identity and password to access a private virtual workbench. The workbench is the center for study management and development.

The workbench is a point-and-click interface for creating, editing, archiving, sharing, copying, deleting, and executing studies. Working with studies in the workbench is conceptually similar to managing financial accounts at banking sites such as http://yodlee.com/. Bank patrons can monitor, perform operations, and open or close financial accounts, and researchers can perform similar actions on studies in the workbench.

Studies exist in one of four phases that correspond to the familiar cycle of a research project: designing, submitted, running, archived. Each phase has corresponding *actions* that are performed on studies. In the design phase, studies are *created, edited*, *duplicated*, and *deleted*. Description of the editing process appears later in the Research Development Environment (RDE) section. Once designed and tested, the study is submitted and the VL administrator puts the study *under review,* tests that it conforms to system and web requirements, and either *accepts* or *rejects* it.

Rejected studies move back to the design phase for *editing* and *resubmission*. Accepted studies move to the *running* phase where the administrator can manage the study's availability for data collection with *play* or *pause* actions. Studies are moved to the archive phase when data collection is completed or if the researcher requests that the study be *withdrawn*. Researchers can *duplicate* or *revise* archived studies by bringing them back to the design phase.

The virtual workbench provides a straightforward study management framework that maintains active studies and archives studies for later review, sharing, or use. It also streamlines and automates interactivity between researchers and technical administrators for testing and posting studies.

Version control. Many research paradigms evolve across a sequence of studies that tweak various study parameters to perfect the experimental design. Productive lines of research often involve a series of studies that systematically manipulate factors in a single paradigm. The VL's version control system simplifies the tracking of variations to a study paradigm. When a study is moved to the running phase it receives a version number. Later, if those study materials are revised for a second data collection, the version number is incremented. All versions of the study are archived allowing researchers to examine and compare the changes to study procedures across a collection of related studies. Version control can also guarantee that a study design is identical to a previous version except for the changes made deliberately in the revision.

Buddy system. Science is a collaborative enterprise. Even on solo projects, it is common to ask colleagues to pretest and provide comments on a study design. And, when finished or published, colleagues often want to replicate or extend studies that have interesting implications, or want to obtain the original data for reanalysis. With the VL buddy system, researchers can give *execute*, *read*, or *edit* privileges to other researchers. Execute privileges allow a colleague to engage a study as participants would see it. Read privileges allow others to see and copy the study materials. Edit privileges give others the capability to edit the study design. A study log facilitates communication about design changes and features among collaborators. The buddy system simplifies collaboration, especially when researchers are geographically dispersed.

Access privileges can be granted for studies that are being designed, actively collecting data, or were completed long ago. For example, if someone requests materials or data from a study published five years earlier, the researcher can open the study in the archive and grant read and execute privileges. The requestor could then self-administer the study, and integrate or adapt the study materials into his or her own studies. This would eliminate the long searches through files and boxes of old materials, incomplete sharing of information, and failures of organization or memory in reproducing the particulars of the study design.

Finally, with a mouse click researchers can share, pretest, and get comments on study designs from people who do not have accounts in the VL by posting the study to a temporary webpage. Research assistants or other colleagues can visit the temporary page and execute the study before it is unleashed.

Study Development

Software facilitating study development has already made substantial improvements to research efficiency and capability. Packages such as Inquisit (http://millisecond.com/) and MediaLab (http://empirisoft.com/) provide flexible environments for creating studies. This software increases the reusability, sharability, and consistency of study materials. In addition to its own study development tools, the VL provides a general management platform enabling integration of study materials created with other popular development packages with web-delivery components.

Research Development Environment (RDE). The VL's study editor is called the RDE. Researchers access the RDE from the virtual workbench by selecting a study that they wish to edit. The RDE is a browser-based editor tuned especially for studies that are administered in a web browser. A point-and-click user interface enables researchers to define a collection of tasks – instructions, questionnaires, video clips, response latency measures, study manipulations – that comprise a study, and the experimental design for administering those tasks. Researchers can define any combination of random assignment rules, random selection of a subset of tasks (e.g., present a random 3 of 5 tasks), fixed task orders, and branching rules in which responses on an earlier task influence which task comes next or the contents of the next task. On a questionnaire about smoking, for example, participants self-identifying as non-smokers can be directed to one task, while smokers are sent to another.

Point-and-click editors facilitate creation of common tasks such as questionnaires or instructions. Editors can also be introduced for specific tasks that have a predictable format. For example, an editor for the Implicit Association Test (IAT) allows users to define the categories, response blocks, trials, response configurations, error handling, and instructions, simplifying the process of implementing an IAT while still maintaining flexibility in its design. Finally, tasks that are created by third-parties without an RDE editor can be uploaded and integrated into the study design. Any task that follows a simple set of technical rules defined by the VL can be implemented in the framework introducing the possibility of a general environment for study management and administration.

Libraries. The RDE contains a library of study materials and tasks facilitating reuse, replication, and redesign. The library has two modes: “My Library” and “System Library.” My Library contains study materials – questionnaires, instruction pages, stimulus materials, etc – created by the researcher for previous studies. The System Library contains the materials created by other researchers made available to the community of users. For example, instead of using an editor to create one’s own Rosenberg Self-Esteem scale, researchers need only open the System Library and import it. Some studies can be created in minutes using preexisting study materials combined uniquely into a new design. Reusability guarantees that the administration of a given task is identical across studies.

Templates. Sometimes researchers would like to define a general study template with some fixed features that are common across many implementations, and some flexible features that will change for each implementation. A common use would be cross-cultural research in which a single study design is implemented in multiple languages. The VL's templating system facilitates this process. Templates define the fixed and editable features of a study. Each implementation of the templates on a webpage presents the editable parts with the original information next to an open field for entering revisions.

For example, a researcher might create a study in English and make the viewable content (instructions, questionnaire items) editable in the template. A bilingual French-English collaborator would see each item in English and enter the French translation in the open field next to it. Then, both versions could be implemented with assurance that everything in the study design is identical except for the translated text. This way, individual translators need only focus on the translation; the study implementation is automatic. Any update or change to the design template can be applied to all versions of the study instantly without re-engaging the translator (unless new translations are needed). The international sites at Project Implicit (http://implicit.harvard.edu/) provide examples of sites and studies created with the template system.

Study Administration

Our vision for an effective Internet research environment includes (a) ethical administration of research protocols including consent and debriefing, (b) sufficient security for the privacy of participants and their data, (c) study design tools adequate to maximize the impact of the independent variables, (d) minimizing external elements that disrupt the experimental procedures and introduce error into research designs, and (e) flexibility to develop instruments beyond the standard survey (see Nosek, Banaji, and Greenwald, 2002c for a review). Here we describe the software with which we administer studies created in the VL, and three applications of the software that illustrate its flexibility.

Production Environment. The PIBE (Project Implicit Back End) administers studies securely through a web browser. After studies have been designed in the RDE and moved into production through the virtual workbench, they are administered by the PIBE (pronounced *pie-bee*). PIBE executes the defined experimental design including random assignment, selection of tasks, and definition of branching rules that depend on participant responses. PIBE also provides a secure data connection between the Project Implicit servers and the participant machine, and for transfer of participant data to and from the databases. Participant behavior during a study session is tracked and retained for possible use in data analysis, including identification of participant misbehavior. Study tilt mechanisms are in place to prevent certain behaviors or disqualify participants who violate study rules. For example, participants are prevented from going 'back' in the web browser, or otherwise self-navigating through an experiment. Time spent on each task can be limited and recorded in the database. Researcher-enabled (or disabled) study continuation and withdrawal mechanisms allow participants to leave a study and continue where they left off at a later time, or withdraw from the study completely and be directed to debriefing materials.

Project Implicit – Demonstration sites. One implementation of this software is the Project Implicit demonstration websites (https://implicit.harvard.edu/) including the international sites. These sites prioritize educational goals about basic research in implicit cognition. Visitors to the site can review information about research in implicit cognition, self-select and complete an IAT from a list of topical domains (e.g., political, racial, age attitudes), receive feedback on their IAT performance, and review follow-up information and frequently asked questions about the tool and research area. Study behavior is carefully managed within the study session, and participants are encouraged to select whichever topical domains are of interest to them. This form of implementation has the advantage of being topically engaging and educational, and the disadvantage of exerting little control on selection processes. Nonetheless, the popularity of the sites has lead to very large datasets that have been useful for a number of research applications (e.g., Greenwald, et al., 2003; Nosek, et al., 2002a; Nosek, Greenwald, & Banaji, 2005; Nosek et al., 2006; Nosek & Smyth, in press; Smyth, Nosek, & McArdle, 2006).

Project Implicit – Research site. Accompanying the demonstration websites is a research site that prioritizes basic research goals by instituting stricter controls on participant behavior. Volunteers register at the research site using an email address as a unique identifier and login for future participation. Part of the registration process includes a demographics questionnaire assessing age, gender, ethnicity, education, country of residence, country of citizenship, fluency, political orientation, religion, religiosity, and occupation. Participants who self-identify as under 18 are informed that they must obtain parent/guardian approval prior to participation. When that user's login is entered again, a parental consent page appears and requires entry of a parent's email address as an approval signature.

Once registered, participants are eligible to participate in studies in the Project Implicit study pool. Each time participants visit the site, they are randomly assigned to one of the studies in the pool. The studies cover a variety of topics with a common theme of investigating implicit cognition. Researchers can define rules for participant selection. For example, most studies allow participants to be assigned to the study only once. Researchers can also define selection rules based on the responses provided during registration. This mechanism automates the selection of specific, sometimes very specific, samples from the participant pool. For example, Teachman and colleagues (2006) selected only those participants who reported occupational categories related to mental-health practitioners. The selection rules can also be weighted by category, e.g., for a study comparing Black and White participants, the researcher can increase the likelihood that Black participants are selected to compensate for base-rate differences in the pool. Selection weights can also be assigned between studies such that an urgent study can receive a higher percentage of pool participants than a lower priority study. These selection rules are applied unobtrusively so that participants are assigned to studies without knowing that some aspect(s) of their identity was relevant to study assignment. Finally, tracking participants over time facilitates longitudinal research, monitoring attrition, and making comparisons across studies.

Private studies. The demonstration and research sites are public sites for volunteer participants who find Project Implicit by any variety of means. These are the high-profile uses of the Virtual Laboratory. The more common and general applications of the tools are private studies for targeted samples. Many studies have private entry points to which a specific sample is invited. Those samples might be from a researcher’s usual university participant pool or a community sample recruited specifically for the study. For example, Sherman and colleagues (2006) used a private entry for participants in their 20-year longitudinal study of smoking attitudes and behavior. Likewise, Smith and Nosek (2005) asked participants who had visited the laboratory for a first session to complete the second session via a private webpage rather than by returning to the laboratory.

Private sites can also be used to match study implementations across multiple laboratories. A single study can be administered in multiple Internet-connected laboratories allowing distributed research teams to collaborate easily on data collection. The study design and collected data is stored and accessed through a single environment.

Addressing methodological challenges of web-based data collection

The accumulating experience with Internet research indicates that effective, reliable, and valid virtual laboratories are achievable with a new generation of technological and methodological innovation. While surveys are relatively easy to create and administer via the web, more sophisticated experimental protocols require a new generation of tools and resources than those available to most psychologists. Our Virtual Laboratory seeks to bridge some of the technical challenges to conducting effective and valid Internet research.

A continuing challenge is identifying the unique methodological strengths and challenges of Internet-based research compared to other research environments (reviews: Nosek et al., 2002c; Skitka & Sargis, 2005). For example, the experimenter is not present for most web-based research. While this is a boon for avoiding experimenter effects and minimizing coercion, it is a threat to comprehension of instructions and following of experimental protocols. Insuring that instructions are clear is not as critical in the traditional laboratory because the experimenter is present to answer questions, clarify confusion, and monitor progress through the protocol. Web protocols are more prone to damage resulting from inattention to clarity of instructional materials.

Another early concern for web-based research was the possibility that participants would be less truthful than in the traditional laboratory. Research comparing Internet and laboratory methods finds them to be comparable (Gosling, et al., 2004; McGraw et al., 2000; Smyth et al., 2006). The evidence suggests that participants may be somewhat less honest on the web than in the lab for some dimensions (e.g., mis-reporting gender or ethnicity, presumably because the lab experimenter would know), and somewhat more honest on the web than in the lab for other dimensions (e.g., reporting sensitive behaviors such as racial biases or illegal drug use, presumably because the lab experimenter would know; Evans, Garcia, Baron & Garcia, 2003).

Expansion of effective psychological research on the Internet will account for the constraints of the medium and generate alternative an eResearch methodology that works around or within those constraints (Nosek, et al., 2002c). For example, it is true that Internet samples are more diverse than the typical samples in the traditional laboratory. It is not true, however, that data collected via the Internet is necessarily generalizable to a definable population (e.g., the United States). So, like all other forms of data collection, factors involved in participant recruitment will necessarily affect the characteristics and generalizability. This is not a fatal flaw. Rather, it suggests that results should be interpreted just as carefully as one might with other samples – i.e. being vigilant to avoid claims beyond those warranted by the properties of the sample.

The fact that the Internet enables data collection away from the traditional laboratory has strengths and weaknesses, depending on the methodological factor of interest. A strength is that people are likely to participate while in a familiar and comfortable environment rather than having to visit an unfamiliar and unusual laboratory highlighting the emphasis on being in a study and being observed. A weakness is that the experimental environment is not controlled, thus allowing heterogeneity in participant circumstances. For generalizability, this may be a benefit, but for power and experimental impact it is not. Having multiple methods available increases flexibility to select a method of data collection that is strongest on features most critical for a given study.

Internet research is a complement, not a replacement, for the traditional laboratory. Capitalizing on the strengths of the Internet expands the domain of methods and testable questions. It would be unfortunate if the increased ease of doing web-based research diminished interest in conducting high-quality laboratory studies. The reverse should be true: Having viable alternative means for doing computer-based experimental and correlational research could free up laboratory resources for studies that must be lab-based because of special equipment needs or requirements for controlling or manipulating the social environment.

Vision: Using technological innovation to transform the imaginative and practical potential of social and behavioral research

Our methodological and technological innovation efforts are motivated by the following vision for the future of the scientific research process from the perspective of an individual scientist:

[Cape Cod, 2017] On the first day of her vacation Dr. Stenfanie Milgraine realizes that a laboratory study that she is running could be advanced significantly by adding an experimental manipulation reported in a recent article by Dr. B.S. Finner. Milgraine logs into her Virtual Laboratory (VL) and calls up Finner's experiment, accessible on the VL system since the article's publication. She runs through Finner's experiment to clearly understand the study manipulation and watches the study video clip to fully comprehend the situational context of the study execution. Convinced that the manipulation is useful, she selects the task implementing the manipulation in Finner's design and copies it into her library. She also copies the raw data from the Finner study for comparing it with her sample data. Milgraine inserts the Finner manipulation in her study and runs through it to confirm that it works as expected. Milgraine replaces the active study with this updated version so that it will appear on the laboratory machines when her lab assistants collect data that week. Milgraine has a hunch that Finner will be interested in the experiment, so she sets Finner's permission to view her study on the VL. Finner, who happens to be at his virtual workbench, is intrigued by this novel application, and messages Milgraine with an offer to run some participants in his lab. Finner also suggests a study variation that would compare Asian and Asian American science majors. He suggests recruiting through the International Participant Pool (IPP) – a consortium of universities that share part of their participant pools for studies that can be executed via the Internet. Milgraine agrees to both ideas and promotes Finner to *collaborator* status on her study. Finner clones the study to his queue of laboratory studies, copies the IRB materials for submission, and assigns two research assistants as collaborators. Then, Finner edits the study design for the Asian sample variation, and submits it to the IPP administrator noting the sample selection criteria – 100 Asian science majors, and 100 Asian American science majors. After clearing IRB and IPP approval processes, the study is included in the IPP, and Finner and Milgraine check-in occasionally on the pace of data collection. The dynamic data monitoring system detects a couple of bugs in the first day and the study is paused briefly while corrections are made. Within a few days, data collection is completed; the investigators retrieve the data, and plan the next step of Milgraine-Finner collaboration.

This imagined scenario illustrates multiple features of a research process that are possible now by developing existing web technologies. The remainder of this section describes a 10-year vision of technological and methodological innovation in the service of social and behavioral research.

A common infrastructure for conducting and managing research

The major innovation articulated in the scenario is a shared study management framework. A generalization of the Virtual Laboratory described in the previous section would be widely available, adaptable to a variety of research methodologies, and integrate study development, management, and administration.

Removing geography as a pragmatic constraint for collaboration and data collection. As a common work space for social and behavioral researchers, a virtual laboratory can remove geography as a practical barrier for data collection and collaboration. Researchers can work together on common projects through a single interface. The same study can be easily implemented in multiple laboratories, and participant samples that are geographically distributed can participate as long as they have Internet access. With increasing global availability of the Internet, even to traditionally underrepresented groups such as the elderly or the poor, the viability of the Internet to access hard-to-reach populations will continue to be one of its most attractive features.

A common framework can cross international borders as well. With tools such as the templating system, cross-cultural research is relatively simple to implement and manage via the Internet. Updating study materials can be relatively instantaneous with a single study design template that is operationalized in multiple languages. Such tools would make the conduct of cross-cultural research almost as easy as research in a single cultural context.

Real-time data integration. The Internet allows integration of data from a variety of sources. Many large-scale databases are available via the Internet such as census data (http://www.census.gov/; U.S. Census Bureau, 2006), weather (http://www.noaa.gov; National Oceanic and Atmospheric Administration, 2006), and social science databases (e.g., http://www.icpsr.umich.edu/; Inter-University Consortium for Political and Social Research). This introduces the possibility of real-time data integration for use during data analysis, or even in the study design itself.

For example, a researcher might be interested in the effects of weather on mood. At the beginning of the study, the Virtual Laboratory uses an IP locator database to automatically identify the participant's location, then accesses NOAA's National Weather Service to find the current weather in that location, and then add that information to the study. If it is raining, the participant gets one set of materials, if it is sunny, a different set is presented. Likewise, a participant's zipcode could be matched with on-line Census databases and variables relevant to that study could be immediately incorporated into the dataset for later use. Possibilities like this will expand the imaginative and practical potential for investigating human experience.

Sharing designs, materials and data. The libraries described in the previous section illustrate the advantages for widespread availability of study materials. With a common framework, all components of research – designs, materials, instructions, and data – can be shared among researchers. This will remove substantial degrees of uncertainty for the replication of research designs, and add enormous flexibility for secondary data analysis.

We envision a system in which identifying a particular task or measure instantly generates a list of studies that used it; identifying a study generates a list of measures that it used; and identifying a dataset produces a list of measures that generated it. Such a system would make it easier to conduct meta- and mega-analyses and, to the extent that the common framework is widely used, increase confidence that the search identified all of the relevant studies. In fact, for some types of comparisons, meta-analytic results could be automated so that selecting a measure identifies all the studies in which it was applied and provides an aggregate data summary. A common framework can be a great leap forward toward eliminating the file-drawer problem (Rosenthal, 1979) by creating a single, shared file-drawer.

An International Participant Pool. The university participant pool is a model for expanding research opportunities and efficient use of resources. Sharing the responsibility of recruiting, maintaining, and administering a participant pool among laboratories (usually organized at the department level) is much more efficient than leaving each laboratory to recruit its own samples. Further, many participant pools include a preselection mechanism in which laboratories that need to recruit a sample with specific characteristics, can identify that sample from the larger pool. This single innovation has been a boon to social and behavioral research. Even so, the department participant pool is a simple starting point, compared to more encompassing, complex possibilities.

With the establishment of a common research framework, we can conceive an International Participant Pool (IPP) that could include participants from universities, colleges, junior colleges, high schools, and other institutions from all over the world. An IPP would have numerous benefits for advancing science.

First, an IPP would expand the pool of active participants and researchers. In the U.S. alone there are approximately 4,168 universities, colleges, and junior colleges (Infoplease, 2006) and only about 500 of them (12%) have MS or PhD programs in Psychology (American Psychological Association, 2004), where most research is conducted. The practical barriers facing academics at smaller institutions – e.g., teaching demands, availability of resources, number of researchers who would use a pool – make it impractical to develop and maintain a pool making it more difficult for researchers at those universities to conduct research, and for educators at those university to expose their students to the scientific process through research participation. An IPP would present an opportunity for educators at teaching-focused institutions to include a participation requirement as part of the curriculum, as it often is at research universities.

Second, an IPP would increase comparability and generalizability of research. There is huge variability in the demographics of student bodies across universities. Some institutions attract exceptional students, some attract highly diverse student bodies, some are especially attractive for particular political or religious ideologies, and most are highly regionalized. These selection factors introduce many sampling idiosyncracies for studies conducted at a single university. When a study conducted at one university fails to replicate at another, it is difficult to tell if it is a subtle change in the study procedures or something relating to differences in the participant populations. It can be very difficult to identify such influences (recall the earlier discussion of Richeson & Nussbaum, 2004 and Smyth & Nosek, 2007). With an IPP, samples would be pooled into a single population maximizing the consistency of participant sampling across universities.

Third, an IPP would increase access to specific subpopulations. Researchers are usually constrained to the samples that are available in their immediate region. With an IPP, researchers at Michigan could easily conduct a study comparing Northern and Southern participants, researchers at predominately liberal colleges could obtain conservative samples, researchers could maximize or minimize diversity on a variety of characteristics as demanded by the research question, and cross-cultural research would be at least an order-of-magnitude easier to conduct. Further, preselection on a grand scale could facilitate recruiting very specific samples.

Finally, a well-functioning IPP could be a nexus for samples of many varieties: ‘captive’ student samples who participate as part of an educational requirement, public volunteer samples, paid samples, and probability samples. National and international organizational bodies could administer the IPP to ensure fairness in the maintenance and distribution of its resources. We are developing a pilot proposal to investigate the feasibility of an IPP. With appropriate support from national funding agencies and professional organizations we believe that an IPP is achievable within 10 years.

Developing a sustainable common framework. In 2005, Project Implicit incorporated as a non-profit entity. One of its goals is to advance basic research through technological and methodological innovation by making the tools described here available to others. As of 2006, Project Implicit supports individual research laboratory studies on a contract-by-contract basis. When more of the study design and implementation process is automated, the development team will be able to support more studies for lower costs. With full automation, the tools should be broadly accessible and flexible for many research applications in the social and behavioral sciences. With each step forward in development, the accessibility, usability, and affordability of the framework will increase.

The success of this model will require that the tools be designed well enough to attract a sizable user base, and be managed efficiently so that use of the framework is affordable and extensible. The transition from research project to a national or international framework should involve professional organizations and oversight committees that define standards and principles for the management of shared resources.

A model for "big science" in the social and behavioral sciences

The currency of the academic researcher is publications. Careers are made or broken on their number and impact. At the same time, the social and behavioral sciences are still deeply grounded in the humanities/philosophy model of "every academic is an island." Collaboration is valued, but the mechanisms of reward and advancement emphasize achievements that are easily identified as *individual* contributions. Researchers who collaborate heavily, especially early in their career, risk acquiring the label "lacks independence." The structure of reward in psychology creates the normative research process – small-scale studies, lots of publications, and few authors on each publication. This model is efficient for some substantive questions, but not others.

Many questions in the social and behavioral sciences are of the *big science* variety – large, complex problems that would be most effectively answered with broad, collaborative research efforts. Such projects are rare in the behavioral sciences. Because of the system of reward and the availability of resources, behavioral disciplines approach big questions with many small studies each manipulating one or two variables, measuring one or two outcomes, assessing a small, specific sample, and holding constant a variety of factors that could be important qualifiers. For big questions, this approach is inefficient, especially for social and behavioral sciences where interactions are the norm, but can only be identified if both factors are measured or manipulated.

Science is at its best as a collaborative discipline. Technological innovation makes is possible to pursue big questions with big studies. A common framework makes it possible for 5, 20, or 50 research laboratories to collaborate on a study, or group of studies, that embraces the complexity of a problem, manipulates many variables, measures a variety of outcomes, includes heterogeneous samples from around a country or the globe, and is not limited by the doctrine of *ceteris paribus* – holding many variables constant. Done separately, individual laboratories could spend many years debating what procedural or methodological factor explains their divergent results. Done together, such differences would be identified quickly, at early stage, and may lead to an unanticipated discovery.

Many questions are bigger than what is practical for an individual researcher to answer, and many institutions – especially those without graduate programs – do not have the resources to support large-scale research. The Virtual Laboratory model emphasizes collaboration – the building of a community resource for the collective advancement of science. The individualistic model and unequal distribution of resources are two factors that contribute to an uncomfortable reality: Most research has little impact beyond lengthening vitas. One analysis of 783,339 articles in the Institute for Scientific Information database found that 368,100, or 47%, had never been cited, and 81% had been cited 10 times or less (Redner, 2004).

The social and behavioral sciences can better capitalize on its intellectual resources by making it easier and more desirable to collaborate. The Virtual Laboratory will provide a nexus point so that individual researchers, who on their own could only examine a small part of a large problem, work collectively to pursue stronger inference, more efficiently. Social and behavioral sciences can also mature by developing a sophisticated ecosystem that encourages productive collaboration without imposing career losses to individual contributors.

Keeping pace with technology

Technological innovation in the 21st century is rapid and accelerating. Within 10 years, the Internet turned from an interesting, untested environment into a fundamental part of many people's lives. It is easy for scientists to get comfortable with the technologies and methods that they know, and stop paying attention to the potential offered by new technologies. Promoting awareness that new technologies are accessible and offer profound opportunity to improve the practice of science is a fundamental goal of this article. The social and behavioral sciences would be well-served to encourage constant attention to technological innovation and adaptation to scientific uses. Even better, social and behavioral scientists should be participating in the invention of new technologies with their potential uses for science in mind.

As of mid-2006, video conferencing, for example, is on the verge of turning from cool novelty to a mainstream application. Instant messaging packages like AIM, Yahoo, and Skype offer free video conferencing, and computers such as Apple's iMac now arrive with a built-in video camera and microphone. For some web-based studies, experimenters may need to have face-to-face communication to administer consent or debriefing, to deliver instructions or measures, or to verify participant identity. With mainstream availability, video conferencing suddenly becomes an easy way for experimenters and participants to interact in real-time during a remotely-administered study.

With video conferencing, experimenters have another means to measure information about the participant and his or her social environment. A modern innovation in developmental psychology involves measuring infant looking time as a measure of expectancy violation (Baillargeon, 1995). This type of research is difficult to conceive conducting with the experimenter and child being in different places, and yet, the challenge of bringing parents and infants to the laboratory is an expensive and time-consuming affair. Infant research-at-a-distance is conceivable with video conferencing.

Experimenters could schedule an on-line session with a parent, administer the study via the Internet, and use video conferencing to record infant looking time. The practical benefits are obvious – scheduling is easier, costs are reduced, data collection is more efficient. Certainly there would be challenges to surmount before actually implementing such a study, but the possibilities for such innovation are tantalizing.

If done well, such applications of technology could revolutionize research with children and families, as might the other technologies described in this article revolutionize social and behavioral research generally. The conditions for implementing such methods on a broad scale are just a few years away, especially if we work on it now.

References

American Psychological Association. (2004). *Graduate Study in Psychology*. Washington, D.C.: American Psychological Association.

Baillargeon, R. (1995). A model of physical reasoning in infancy. In C. Rovee-Collier & L. P. Lipsett (Eds.), *Advances in Infancy Research* (Vol. 9, pp 305-371). Norwood, NJ: Erlbaum.

Best, S. J., B. Krueger, Hubbard, C., & Smith, A. (2001). “An assessment of the generalizability of Internet surveys.” *Social Science Computer Review, 19*, 131-145.

Birnbaum, M. H. (2000). *Psychological Experiments on the Internet*. San Diego, California: Academic Press.

Brenner, V. (2002). Generalizability issues in Internet-based survey research: Implications for the Internet addiction controversy. In *Online Social Sciences,* B. Batinic, U-D. Reips, M. Bosnjak (Eds.), pp. 93-113. Gottingen, Germany: Hogrefe & Huber.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Cooper, A., Scherer, C., & Mathy, R. M. (2001). Overcoming methodological concerns in the investigation of online sexual activities. *Cyberpsychology & Behavior, 4,* 437-447.

Doob, L. W. (1980). The inconclusiveness struggles of cross-cultural psychology. *Journal of Cross-Cultural Psychology, 11(1)*, 59-73.

Doob, L. W. (1988). Avoding an impression of ethno- and egocentrism. *The Journal of Psychology, 122(3)*, 259-262.

Evans, D. C., Garcia, D. J., Baron, R. S., & Garcia, D. M. (2003). In the privacy of their own homes: Using the Internet to assess racial bias. *Personality and Social Psychology Bulletin, 29*, 273-284.

Fiske, A. P., Kitayama, S., Markus, H. R., & Nisbett, R. E. (1998). The cultural matrix of social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 2, pp. 915-981). Boston, MA: McGraw-Hill.

Gonzales, P., Guzmán, J. C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., & Williams, T. (2004). *Highlights From the Trends in International Mathematics and Science Study (TIMSS) 2003* (NCES 2005–005). U.S. Department of Education, National Center for Education Statistics. Retrieved May 1, 2006, from http://nces.ed.gov/pubs2005/2005005.pdf

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59*, 93-104.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellot, D. S. (2002). A unified theory of implicit attitudes, beliefs, self-esteem and self-concept. *Psychological Review, 109(1)*, 3-25.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74(6)*, 1464-1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.

Infoplease. (2006). Number of U.S. Colleges and Universities and Degrees Awarded. Retrieved August 8, 2006, from the World Wide Web: http://www.infoplease.com/ipa/A0908742.html

Internet World Stats (2006). http://www.internetworldstats.com/stats.htm

Lane, K. A. & Banaji, M. R. (2005). Boston and national identity. Unpublished data: Harvard University.

Lindner, N., Nosek, B. A., & Banaji, M. R. (2006). Implicit attitudes toward older and younger adults: Effects of gender, culture, and age. Poster presented at the annual meeting of the Society for Personality and Social Psychology, Palm Springs, CA. January 27-28, 2006.

McGraw, K. O., Tew, M. D., & Williams, J. E. (2000). The integrity of web-delivered experiments: Can you trust the data? *Psychological Science, 11*, 502-506.

Mendes, W. B., Gray, H., Mendoza-Denton, R., Major, B., & Epel, E. (2006). Why egalitarianism might be good for your health: Physiological thriving in stressful intergroup situations. Unpublished manuscript: Harvard University.

Michalak, E. E., & Szabo, A. (1998). Guidelines for internet research: An update. *European Psychologist, 3*, 70-75.

National Oceanic & Atmospheric Administration. (2006). NOAA Past Weather. Retrieved August 8, 2006, from the World Wide Web: http://www.noaa.gov/pastweather.html

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134,* 565-584.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002a). Harvesting implicit group attitudes and beliefs from a demonstration website. *Group Dynamics, 6*, 101-115.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002b). Math = Male, Me = Female, therefore Math ≠ Me. *Journal of Personality and Social Psychology, 83(1)*, 44-59.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002c). eResearch: Ethics, security, design, and control in psychological research on the Internet. *Journal of Social Issues*, *58(1)*, 161-176.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (in press). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes*. Psychology Press.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31,* 166-180.

Nosek, B. A., & Smyth, F. L. (2006). *Implicit Cognition in STEM Education*. Grant proposal to the National Science Foundation's Research and Evaluation of Education in Science and Engineering (REESE) Program.

Nosek, B. A., Thompson, E., & Banaji, M. R. (2006). Intergroup attitudes and voting behavior: The 2000 and 2004 U.S. presidential elections. Paper presented at the

Annual Meeting for the Society for Personality and Social Psychology, Palm Springs, CA, January 28, 2006.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N., Smith, C. T., Olson, K., Chugh, D., Ranganath, K. A., Banaji, M. R., & Greenwald, A. G. (2006). Pervasiveness and variability in implicit bias. Unpublished manuscript: University of Virginia.

O'Neil, K. M., Penrod, S. D., & Bornstein, B. H. (2003). Web-based research: Methodological variables' effects on dropout and sample characteristics. *Behavior Research Methods, Instruments, & Computers, 35*, 217-226.

Pew Internet & American Life Project (February 15 – April 6, 2006). http://www.pewinternet.org/trends/User_Demo_4.26.06.htm

Ranganath, K. A., & Nosek, B. A. (2006). Implicit and explicit attitude formation. Unpublished manuscript: University of Virginia.

Redner, S. (2004). How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B – Condensed Matter and Complex Systems, 4(2)*, 131-134.

Richeson, J. A., & Nussbaum, R. J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology, 40*, 417–423.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86(3)*, 638-641.

Schillewaert, N., Langerak, F., & Duhamel (1998). Non-probability sampling for WWW surveys: A comparison of methods. Journal of the Market Research Society 40(4): 307-322.

Sharp, L. B., Monteith, M. J. & Banaji, M. R. (2005). Adopting new attitudes: Implicit attitudes of parents adopting Chinese children. Poster presented at the 6th annual meeting of the Society for Personality and social Psychology, New Orleans, LA.

Sherman, S. J., Chassin, L., Presson, C. C., & Macy, J. T. (2006). Implicit attitudes toward smoking among adolescents and their parents. Unpublished data: Indiana University.

Skitka, L. J., & Sargis, E. G. (2005). The Internet as psychological laboratory. *Annual Review of Psychology, 57*, 529-555.

Smith, M. A., & Leigh, B. (1997). Virtual subjects: Using the Internet as an alternative source of subjects and research environment. *Behavior Research Methods, Instruments, & Computers, 29*, 496-505.

Smith, C. T., & Nosek, B. A. (2005). Unpublished raw data: University of Virginia.

Smyth, F. L., & Nosek, B. A. (2007). The impact of multiculturalism versus color-blindness on racial bias revisited. Poster presented at the annual meeting of the Society for Personality and Social Psychology, Palm Springs, CA. January 2007.

Smyth, F. L., Nosek, B. A., & McArdle, J. J. (2006). *"Choosing" to stereotype: Balanced identity accounts for selection effects in an Internet study of gender-science stereotypes*. Unpublished manuscript: University of Virginia.

Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology, 84*, 1041-1053.

Teachman, B. A., Peris, T., & Nosek, B. A. (2006). Stigma of mental illness among mental-health practitioners. Unpublished data: University of Virginia.

U.S. Census Bureau. (2006). Data Access Tools from the Census Bureau. Retrieved August 8, 2006, from the World Wide Web: http://www.census.gov/main/www/access.html

Walsh, J. P., Kiesler, S., Sproull, L. S., & Hesse, B. W. (1992). Self-selected and randomly selected respondents in a computer network survey. *Public Opinion Quarterly, 56,* 241-244.

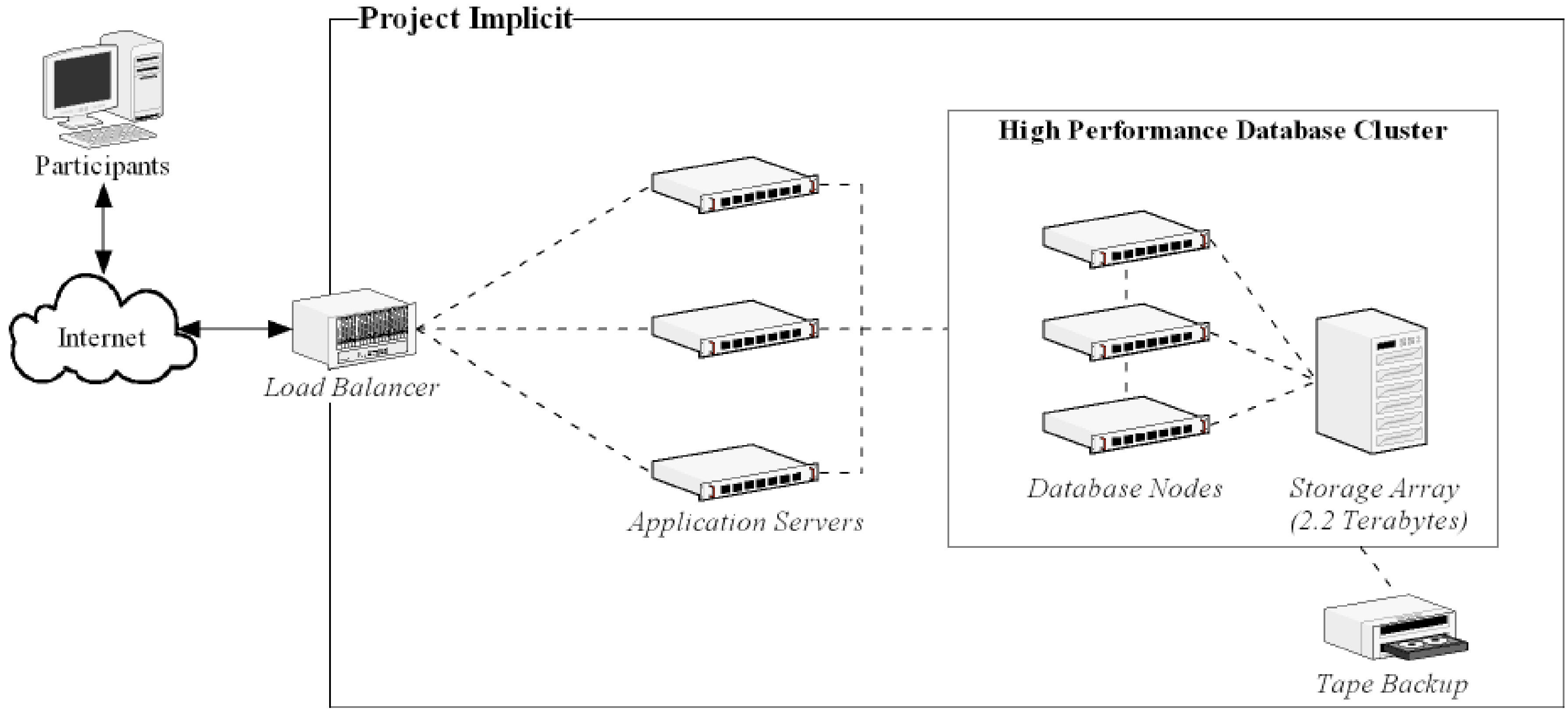
Wolsko, C., Park, B., Judd, C.M., & Wittenbrink, B. (2000). Framing interethnic ideology: Effects of multicultural and colorblind perspectives of judgments of groups and individuals. *Journal of Personality and Social Psychology, 78*, 635-654.

Authors' Note

This research was supported by the National Institute of Mental Health (MH-41328, MH-01533, MH-57672, and MH-68447) and the National Science Foundation (SBR-9422241 and SBR-9709924). The authors are grateful to the technical and administrative support for this research program: Roger Despres, Ford Fay, Richard Hackman, Elizabeth Hess, Mary Hunt, Larry Levine, Philip Long, Pepe Lopez, Peter Mankey, Patrick Ohiomoba, and Brian White. B. A. Nosek, A. G. Greenwald, and M. R. Banaji are officers of Project Implicit, Inc., a non-profit corporation described in this report. Nosek, Greenwald, and Banaji do not receive personal financial compensation from this organization.

Correspondence concerning this article should be addressed to Brian Nosek, Department of Psychology, University of Virginia, Box 400400, Charlottesville, VA 22911; e-mail: nosek@virginia.edu. Related information is available at http://briannosek.com/ and http://projectimplicit.net/.

Figure 1. Hardware Architecture for Project Implicit


Project Implicit
Participants
Internet
Load Balancer
Application Servers
High Performance Database Cluster
Database Nodes
Storage Array
(2.2 Terabytes)
Tape Backup