Ethics for Design-Based Research on Online Social Networks

R. Benjamin Shapiro, Tufts University, Medford, MA, ben@cs.tufts.edu
Pilar N. Ossorio, University of Wisconsin-Madison, pnossorio@wisc.edu

Abstract: Design-Based Research (DBR) allows learning scientists to investigate new processes, contexts, and technologies for learning. Social Networking Sites (SNSs) offer researchers rich new opportunities to create educational interventions that are deeply connected to learners' lives and relationships. We discuss legal and ethical challenges, and possible solutions to them, that face educational researchers as they begin to do DBR on SNSs. Addressing these issues will be crucial to design researchers wishing to use SNSs as sites for learning, and also offers an opportunity for the CSCL community to shape SNS research far beyond our field.

Introduction

Online social networking sites (SNSs) are powerful research tools because of their wide reach and deep connectivity to users' lives. In 2011, approximately 42% of U.S. adults belonged to a SNS (Hampton, Goulet, Ranie, & Purcell, 2011). In 2010, 73% of online U.S. teens used SNSs (Lenhart, Purcell, Smith, & Zickuhr, 2010). Facebook had over 1 billion users in October of 2012; one twelfth of the entire world's population now uses Facebook's mobile apps (Facebook, 2012), highlighting the enormous connectedness of social media to users' daily routines. SNS-based research offers learning scientists the opportunity both to understand human social activity, and to use SNSs for experimental interventions, such as increasing civic participation (Bond et al., 2012) or teaching about science (Shapiro, Squire, and ERIA, 2011). While observational studies offer researchers the ability to analyze activity – including learning – as it is already occurring, Design-Based Research (DBR) enables researchers to support and study new kinds of educational interactions (Brown, 1992; Collins, 1992). Though the number of studies conducted in SNS environments is growing, few DBR studies have been conducted to date. Instead the research has largely focused on publicly available data about extant activities.

Doing DBR on social networks offers researchers new opportunities to connect to learners' lives, and to understand how learning happens across levels of *space, time, and scale*. For example, a recent design experiment conducted by political scientists collaborating with Facebook staff reached over 60 million people and led to over 300,000 more of them voting in the 2010 United States elections (Bond et al., 2012). With SNSs we can build learning experiences that are deeply intertwined with learners' personal interests, and those of their friends and loved ones, such as by using information gleaned from users' posts on their Facebook feeds, or the content of web pages they've "liked." We can virally scale participation in learning environments by using the social sharing mechanisms that are central to SNSs, encouraging learners to invite their friends to learn along with them. We might use slow-moving interaction designs similar to social games like *FarmVille* to immerse learners in distributed embedded phenomena (Moher, 2006) that stretch over long periods of time and that differ depending upon users' physical locations. All of these possibilities highlight new opportunities for researchers to support and to understand learning at both individual and collective scales, over differing time scales, and to inform and study these experiences using new kinds of data.

Standard methodologies, tools, and norms for doing DBR on SNSs have not yet emerged, and we currently lack ethical frameworks for working with the unprecedentedly private data that SNS DBR makes available to researchers. Observational studies are already pushing boundaries; DBR will push further, such as by offering access to more information, exposure of information to, and about, peers. As learning scientists begin to develop DBR programs for SNS, we must also begin to develop a legal and ethical groundwork for doing our work in an appropriate manner. This groundwork can ultimately inform not just study design but the design of CSCL tools as well. Fields beyond the learning sciences are beginning to consider these issues as well (Introne, et al., 2013), and educational researchers have an opportunity to shape both the policy and the research practice landscape of SNS DBR.

The authors of this paper are, respectively, a learning scientist and a legal scholar specializing in bioethics who is a member of her university's IRB and the U.S. Secretary of Health and Human Services Advisory Committee on Human Research Protections (the analysis in this paper is our own and is not an opinion of any committee or governmental agency). We have begun working together to lay the necessary legal and ethical groundwork for DBR on SNSs. This paper illustrates the possibilities for DBR on SNSs using an example drawn from our own work, then discusses some of the legal and ethical considerations that educational researchers, as well as Federal regulators and university IRBs, must wrestle with for work like it to precede. Our *legal* analysis is grounded in study of US regulations including the Code of Federal Regulations, the Federal Policy for the Protection of Human Subjects (the "Common Rule"), as well as case law, though the *ethical* issues we raise are globally applicable.

An Illustrative Scenario

Consider the following vignette describing an online science education game:

> Katrina started playing Anatomy Pro-Am (APA) after receiving an invitation from her Facebook Friend Riley. Now, they play APA together daily. This afternoon, they work on the case of Mr. Badger, whose primary care physician suspects he may have liver cancer. They look for abnormalities in his CT scans. At first, they work independently. Katrina examines the images and decides everything looks OK. She clicks Looks Healthy. When she does, Riley's work appears. Katrina is surprised to see that Riley labeled some spots as cancerous.
>
> Their friend Marcus comes online and they invite him to give a third opinion. He agrees with Riley, but Katrina thinks those spots might be fat in the liver. To settle the argument, they search the APA Almanac and Google Images for reference images of fatty and cancerous livers. As it turns out Mr. Badger's liver looks more fatty than cancerous. Both Riley and Marcus change their diagnoses, and the team saves its work. The players then receive feedback comparing their judgments to those of experts.
>
> Finally, the team's work is sent to McLuhan Hospital, where Mr. Badger is a patient. Their work, along with that of hundreds of others, will be used to generate a representation that will guide McLuhan's radiologist in fine-tuning her analysis of Mr. Badger's case.

APA is a real project (see Figure 1) in which we wrestle with ethically creating and studying an online learning environment that is deeply integrated with social networking. We hypothesize that game players will learn scientific content, increase their feelings of self-efficacy, and increase their interest in scientific careers. In a pilot study, we found that middle-school girls who played a precursor to APA became significantly more likely to believe they could become physicians, as well as better understand the goals, tools, and challenges of cancer medicine (Shapiro, 2011). Moreover, we hypothesize that APA, and projects like it, can demonstrate new ways for the public to participate in science and medicine. In the above vignette, it is through providing crowd-sourced support for medical diagnosis that players both learn science and participate in medicine.

Figure 1. Anatomy Pro-Am.

Third-party Facebook applications (e.g., games) utilize programmatic interfaces to Facebook data, known as APIs, to obtain information about SNS users and their "Friends," and often to act on a user's behalf (for instance, by posting information to her profile). Thus far, commercial software developers have been the primary users of these APIs, but they also enable researchers to create rich, interactive experiences for study. However, unlike commercial software developers, academic researchers are subject to regulations governing human subjects research, as well as ethical guidelines about how to conduct research. These legal and ethical frameworks shape the kinds of studies we can conduct, as well as the manner in which we conduct them. As yet, however, we lack standards specific to SNS SBR. For example, in the above vignette, is it acceptable for Marcus (a teenager) to participate in the educational intervention without his mother's consent? The standard educational research approach is to seek parental permission for studies of minors' learning. But as we describe below, this may significantly impair SNS DBR, offer little actual protection to minors, as well as impose a burden on researchers and participants that exceeds the legal requirements that govern SNS DBR. Or,

considering the issue of privacy, would it be appropriate for us to use Facebook APIs to crawl through and save a complete record of all of Katrina's interactions across Facebook so that we can build a stronger model of relationships between player profiles and their learning through game play? Both of these possibilities are ones in which commercial developers routinely engage, but both are outside the bounds of what most researchers would find appropriate or routine in traditional (non-SNS) DBR.

Here, we address three ethical, regulatory, and technical challenges raised by SNS-embedded DBR: (i) whether adolescents who participate in such research through commercial portals, such as Facebook, should be categorized as children for regulatory purposes; (ii) the extent to which researchers may collect data about SNS participants and their Facebook "friends"; (iii) how CSCL researchers might construct their technical systems so as to maximally protect participants while minimally hindering legitimate research.

Consent and the Adolescent Player

University researchers studying players of an SNS-based educational intervention, such as a game like APA, are engaged in human subjects research and require either Institutional Review Board (IRB) approval or exemption, and players' informed consent/assent (U.S. Department of Health and Human Services, 1991). A surprising question is whether parental permission is required for researchers to study adolescent players (adults give consent for their own research participation; parents give permission, rather than consent, for their children's research participation) (U.S. Department of Health and Human Services, 1983).

Design researchers have good reason to want their experimental SNS applications to be accessible to as many learners as possible. SNSs offer excellent viral recruitment potential, and most Facebook applications enable users to recruit their Friends to participate. These mechanisms are attractive to researchers because they both enable greater scale of participation and because their allow research on learning interventions to be ecologically valid. Research on how social sharing mechanisms can drive participation can increase ecological validity if those mechanisms mirror those used by commercial applications. From this perspective, it is desirable that adolescents participating in our DBR can invite their Friends to participate too, and that those Friends can accept those invitations and participate immediately. Other commentators on SNS research have considered adolescents to be children, and have stated that parental permission will usually be necessary for such research (Bull et al., 2011; Moreno, Fost, & Christakis, 2008). This would considerably damp the participation growth curves of our applications compared to commercial applications.

What The Federal Regulations Say

The Common Rule defines children as: "Persons who have not attained the legal age for consent to treatments or procedures involved in the research, under the applicable law of the jurisdiction in which the research will be conducted." This definition coordinates research regulation with other laws, such as laws permitting adolescents to consent on their own to treatments for sexually transmitted diseases. The regulatory definition of a child permits reasonable research on the effectiveness of those interventions, such as STD treatments, to which adolescents can consent. The regulations are clear that so long as an adolescent can consent to an intervention, she can participate in research about that intervention without parental permission.

The federal Children's Online Privacy Protection Act (COPPA) defines children as *persons under the age of thirteen* (15 U.S.C § 6501 *et seq*, 1998). COPPA prohibits commercial operators of online services (such as Facebook) from engaging in unfair or deceptive practices with respect to the collection, use, or disclosure of personal information from and about children (i.e., those 12 and younger) on the Internet. Congress deliberated extensively before determining that 13 - 17 year olds ought to be treated the same as adults in the online context.

Facebook and other SNSs have responded to COPPA by prohibiting access to their services by people under thirteen. They are not legally required to ban users 12 and younger, but do so in order to avoid the costs of complying with COPPA's requirements that information collection and sharing practices be disclosed to, and parental permission obtained from, the parents of young users. Although some parents help their younger children to evade these age limits, SNSs take measures to restrict underage access, including deleting accounts identified as belonging to underage users (boyd & Hargittai, 2011).

In light of COPPA and the Common Rule's provisions, we believe IRBs should allow adolescents who enter an experimental game through a commercial website that limits access to people thirteen or older to consent on their own to research on participation and learning in that intervention. The existing commercial context of the research, in which 13 year olds have, according to COPPA, obtained legal age to consent to participation, including to data collection, triggers the common rule provisions for treating these adolescents as adults for the purposes of IRB review. In the non-research context, adolescents can legally use SNSs to play games and to provide identifiable, private information about themselves to commercial application developers, and so they should be able to consent in the research context. Parental permission should not be required. This approach is not a waiver of the requirement for parental permission. Rather, we argue that people thirteen and older should not be categorized as children for IRB review of SNS DBR. Researchers will nonetheless need to ensure that they take all possible steps to articulate to adolescent participants how their data will be used and the risks of participation.

Other education and social media researchers with whom we have spoken about doing SNS research report that their universities' IRBs have required them to obtain parental permission for all minors. This requirement contradicts U.S. Federal regulations governing research. While the regulations do permit local IRBs to impose additional protections for participants, we suggest that researchers might use the argument presented here to negotiate with their IRBs for greater latitude to conduct ecologically valid DBR.

What About Parents?

Of course, parents may control their adolescents' online activities. Parents who monitor their adolescents' Internet activities can observe the consent process and read the information provided, and parents are always free to prevent their adolescents' participation in IRB-approved research. IRBs may also impose additional protections for vulnerable populations (45 U.S.C § 46.111b, 2005).

Furthermore, we believe that researchers should make use of the online environment to deliver innovative, truly informative consent processes for *anybody* participating in online research. This is especially important given the general public's substantial ignorance about data collection over the Internet. For instance, while 44% of American parents "are extremely or very concerned that their children might have information about them used for targeted advertising," only 9% of parents whose children use SNSs believe their children's data have been used in this manner (boyd & Hargittai, 2011). In reality, 100% of SNS participants have their data mined for targeted advertising. In light of these data, the traditional approach of IRBs, to assume that parents will be able to weigh the risks of participation in research for their children, seems grossly insufficient to protect youth. Furthermore, many youths interact with social media from across a range of locations and devices, making parental mediation of activity extremely difficult (Yardi & Bruckman, 2011).

Given these and other findings, it is important that online researchers take advantage of their medium to create high quality consent processes that explain how data will be analyzed and clearly inform participants of possible risks. Because participants may only understand the full risks of participation after the fact, researchers should create easily accessible mechanisms for post-hoc withdrawal from participation in research. Allowing participants to delete data about themselves after they have provided it creates additional opportunities for participant education, such as through discussion between adolescents and teachers, parents, and peers, to shape informed consent. Many commercial application developers provide less than straightforward tools for accessing and controlling the information that they maintain; academic researchers could use the more stringent ethical criteria of our field to create and demonstrate higher standards for participant protection.

Scaling from Individuals to Networks

The APA vignette also raises regulatory and ethical questions about collecting information on members of participants' social networks (i.e., their Facebook "Friends"). An important research question for environments like APA is whether they can help overcome race and gender disparities in science. Understanding the demographics of player networks will help answer this question. For instance, to learn how biases shape APA participation, one might examine the demographic characteristics of Facebook Friends whom our main player Katrina invites to work on particularly hard problems vs. easy problems vs. the demographics of her network generally. The average Facebook user has 190 Friends (Backstrom, 2011); it is not plausible that all would consent to data collection. Requiring consent from non-players in order to characterize the overall characteristics of Katrina's network would introduce statistical bias and diminish such a study's rigor.

Facebook APIs offer researchers access to a great deal of identifiable information about game players and their Friends, much of which is necessary for providing game user interfaces. Should design researchers who use Facebook APIs be required to not store identifiable information on non-playing Friends? Should researchers be prohibited to access this identifiable Friend information at all, should they be permitted to access the information only to save it in non-identifiable form, or should they have unrestricted access to it under the guise of Facebook being a public space? In considering this question, we also must figure out how researchers should weigh the difference between Facebook's legal status as a public space (see below) and many users' expectation that it is semi-private.

Perhaps the closest already-understood analogy involves collection of family history information from research participants. Although researchers have engaged in this practice for decades, it became controversial in 2001: regulators temporarily halted all human subjects research at Virginia Commonwealth University after a research participant's father objected to the collection of sensitive family history information (Bolkin, 2001). Commentators and regulators emphasized that researchers who collect identifiable private information about relatives are engaged in human subjects research on the relatives and must obtain their consent, unless an IRB waives this requirement (U.S. Department of Health and Human Services, 1991). However, if the information is not identifiable, or is identifiable but not private, then the researcher is not engaged in regulated human subjects research on relatives, and their consent is not required. Such unregulated research has been called "human non-subjects research" (Brothers & Clayton, 2010).

Under the regulations, information is private if a person reasonably expects it will not be observed or recorded, or if it is provided for a specific purpose and the provider reasonably expects it will not be made

public (U.S. Department of Health and Human Services, 1991). Social conventions regarding information privacy on SNSs are still developing. Facebook warns users that some information is "always publicly available," including one's name, profile picture, network, username and Unique Identifier (UID), and any additional information one chooses to make public (Facebook, 2012). To date, state and federal courts that have addressed the question have held that individuals do not have a reasonable expectation of privacy in information posted to a SNS (Newell, 2011); however, the issue is new enough that courts have not yet considered many aspects of online privacy.

The emerging legal consensus – if not that of the social media research community – is that much information on SNSs is not private, so for now IRBs will often be justified in treating the collection of identifiable information about Facebook Friends as human non-subjects research. Even in the absence of clear regulation, however, investigators have ethical duties to minimize risks to people whose data they use. Even data that may be innocuous in its raw form may be embarrassing when crystalized into accountable facts (e.g., that someone exhibits racial bias in his/her daily choices; we describe a hypothetical SNS-based analysis of this below). A primary way to reduce risk to research participants is limiting their identifiability as research participants, while simultaneously mitigating the consequences to them of such identifiability. A standard practice for doing so is to anonymize collected data as early in the research process as possible.

Reidentification Risk

Though anonymization can be a way to protect the identities of participants and their associates, the challenge of successfully anonymizing data, while still permitting useful research, is surprisingly difficult (Ohm, 2010). Even seemingly sanitized datasets can be de-anonymized, such as by using network structure to re-identify individuals (Narayanan & Schmatikov, 2009), or statistical inference combined with non-random identifier generation algorithms to predict social security numbers (Acquisti & Gross, 2009).

Perhaps the most notable (or notorious) case of re-identifiability of an *academic* SNS dataset thus far is the Tastes, Ties, and Time (T3) project (Lewis et al., 2008). Researchers at Harvard and University of California, Los Angeles downloaded the Facebook profiles of freshmen at a "diverse private college in the Northeast U.S." for four years, and combined those data with housing records obtained from that university in order to study relationships between online participation and physical space. This longitudinal dataset offered deep insights into the social behavior of four cohorts of undergraduates across different levels of *space, time, and scale*. The research was IRB-approved, and the data released with assurances that "all identifying information was deleted or encoded" (ibid.).

However, nearly immediately after the data were released, other researchers quickly identified the data as belonging to Harvard College students (Zimmer, 2010); some students could be easily identified based upon the sparseness of the space of individual characteristics present in the dataset (students from state X, majoring in Y, interested in Z). For example, only one student was present in the dataset for a given year with the home state of Mississippi. It would be easy to combine this data with publicly available information to identify many of the individuals present in the data set. Thus even though researchers made good faith efforts to protect participants (by removing their names from the dataset), the information that was released was enough to reveal participants' identities and reunite it with the large amount of personal data that researchers harvested.

New Identifiability Challenges for Design Researchers

Design-based research raises new issues that projects like T3 have not yet contended with. T3, like almost all other SNS-based research to date, is exclusively observational in nature. Researchers collected data on already unfolding online activities. But design-based research on SNSs goes a step further, to providing the context for new activities that are situated in the existing context of sites like Facebook. This raises new challenges because it pits two important aims – both of urgent importance to design researchers – against each other: the needs of being an experience provider and facilitator, and the de-identification needs of ethical research. Most Facebook games (including our own) are long-running socially connected experiences. They are microworlds where users can log in repeatedly over time and have a stable profile, connected to their Facebook profiles and Friend networks, that follows them from session to session. Designers facilitate the social elements of these experiences for users by disclosing information to their peers (Friends) about their participation. For example, in the above vignette, Riley, Katrina, and Marcus are all informed of their friends' participation in the APA game. In order to do any of this, application developers must maintain a considerable amount of identifying information about participants. If Katrina's game play history were totally disassociated with her identity, then she would be required to start anew each time she played, which would significantly alter the nature of the educational project. Similarly, if users could not see which of their friends play the game, including perhaps their competencies at different skills the game demands, then opportunities for collaborative learning (such as through creating teams of complementarily skilled players) would be substantially diminished. Insofar as peer-driven discovery is the major means by which Facebook applications grow their user bases, this would limit the potential of an educational project to succeed, as well as the generalizability of the research. Ultimately the need for collecting and (sometimes) disclosing identifiable information is irreconcilable with the protective heuristic of anonymizing data as early in the research process as possible.

Technical Guidelines

The thicket of issues raised above is primarily ethical and legal in nature. We cannot find salvation from the described tensions among important values in technical solutions alone. However, it may be possible to architect systems for design-based research on SNSs in ways that reduce risk to participants and their SNS Friends.

In a typical Facebook game scenario, the user plays a game in his or her web browser (and increasingly on mobile devices). The game appears to be a part of Facebook, but resides on a different server (Figure 1 illustrates how a game hosted at a university appears seamlessly integrated with the rest of the Facebook interface). When the user first navigates to the application from a link on Facebook, Facebook's servers, the user's browser, and the application servers exchange information about the user. Applications can request, among other things, access to the user's profile, the ability to send messages on behalf of the user, to see Friends' information, and the ability to do all of this and more when the user is not even logged in. If the user has not previously agreed to the access that the application requests, then Facebook asks the user for permission. If the user agrees, access keys are sent from Facebook, via the user's browser, to the application developer. These keys are used in subsequent Facebook requests to authenticate access to users' profiles. The user may subsequently revoke these tokens using through Facebook's.

Once an application has these keys, it can use them to interact with the user's profile via the Facebook APIs in two distinct ways, each with strikingly different implications for users' privacy: The application's servers may use these APIs to directly obtain information from or post information to Facebook. Or, the application may execute JavaScript in the user's browser that interacts directly with Facebook via these APIs to obtain the information, and then parlay that information back to the university-based application servers. The former approach enables application developers to shift as much of the computational burden away from the user's browser as possible, as well as to maximize the amount of information that application developers have access to. The latter approach limits the demands on application developers' servers, may have worse performance characteristics from the user's perspective, but also has some interesting potential uses for increasing users' and users' Friends' privacy.

Consider the APA vignette above and suppose we wish to understand how racial bias creeps in to Katrina's decisions about whom to invite to collaborate with her. To study this, we would want information about the demographics of Katrina's network overall, such as the race of each of her Friends, as well as about the people Katrina specifically chooses to invite to work with her. We might be interested in comparing her choices in game play with her general choices about whom to interact with, and so also grab race data about the people Katrina chats with or is tagged in photos with. In the end, our analysis does not require these raw data, only aggregate statistics about Katrina's network and her collaborators (e.g., that 10% of her Friends are African American, that she is just as likely to chat with African Americans as other Friends, but far less likely to ask an African American to help her on a difficult game challenge).

We could obtain these aggregate data using either the server- or the browser-based approaches. In the server-based approach, our university-based servers would use Facebook APIs to crawl through Katrina's social network, retrieving and accumulating information about each Friend, her messaging history, etc., eventually distilling this information down to the necessary aggregates. In the browser-based approach, our code, running in Katrina's browser, would do much the same work. The difference is that in the latter scenario the raw data need never exist on our university servers. Only the aggregated information would be sent to us. This is strongly preferable from the standpoint of protecting SNS members' privacy, as it allows researchers to ask questions about participants who are situated within their networks without requiring researchers to ever see raw information about Friends that could be considered private.

This approach also has the benefit of tying data collection to explicit actions that users take to participate in research. A server-based approach allows researchers to harvest data about users and the Friends at will, with no active involvement by users. Researchers could periodically harvest information from all users' profiles and networks without users knowing about this ongoing activity. In contrast, if data collection code runs in the user's browser, it will only be active when the user has actively chosen to use our research-driven application. Unlike traditional research, when the event of participating in data collection is explicit and signaled to subjects by their unusual interaction with researchers, online research (such as by Facebook APIs) permits ongoing data collection about users once they have agreed to participate. A server-driven data-collection approach permits data collection months or even years after a user has consented, when he or she may not even remember doing so or even be aware that it is continuing. The browser-based approach requires explicit action by the user to re-enter the research space, and so permits the user to make an explicit, conscious choice about continuing to provide data to researchers. We believe that this is ethically preferable. Of course, this approach is only suitable to some research questions and methods, but exploring software architectures like this is a first step toward building systems that enable design-based research while maximizing participant protections. We hope other researchers will explore the space of possible technical systems designs that permit SNS-based design research while maximizing participant protections.

Conclusions

We have described several of the challenging questions facing design-based educational researchers hoping to utilize social networks like Facebook to create the learning environments of the future. A major promise of social media is peer-driven discovery, and we are excited about the prospects of creating learning experiences so compelling that learners voluntarily share them with their friends. Traditional IRB processes, particularly the requirement of parental permission that educational researchers usually obtain, would seem to preclude the possibility of viral growth in participation (and learning). But our analysis of the legal regulations governing commercial and academic work in this area supports the surprising conclusion that adolescents thirteen and older should be treated as adults for the purposes of research consent.

We examined the question of whether information about consenting users' peers (i.e. their Facebook Friends) should be accessible to researchers. Though the legal landscape in this area is changing quickly, there is currently a basis to permit researchers to use information about participants' broad networks without the explicit consent of others in the networks. Nonetheless, there is a need for researchers to develop ethical standards that are more stringent than the law alone requires, and to devise such standards in ways that both permit promising research and protect online users' privacy.

IRBs have been criticized for their inconsistent assessment of social, psychological, and economic harms and benefits in other research contexts (Department of Health and Human Services, 2011), and they will undoubtedly need guidance for assessing SNS research. Design-based research raises new ethical issues regarding risk minimization. IRBs must be educated regarding the array of technical options for addressing ethical issues. Further, funding agencies such as NSF and NIH might issue calls for research on building open source participant-protective cyberinfrastructure for SNS DBR, as well as work with HHS to encourage IRBs to consider the use of such infrastructure as a means of participant protection when evaluating research designs.

OHRP recently circulated a proposal to revise the Common Rule (Department of Health and Human Services, 2011). The proposal mentions the Internet as an emerging experimental sphere; however, it contains no discussion of how research in this sphere ought to influence regulatory revisions, or how such research would be reviewed under the new proposals. In the revisions, OHRP takes the position that all informational risk in research can be minimized by adequate data security precautions. While we believe data security is important, we suggest that appropriate research design also affects the risk calculus of SNS research. As the T3 example above highlights, even anonymized observational data bears identification risks for participants. We do not yet know how to balance the technical needs of creating persistent online experiences with the ethical need to minimize participant identifiability.

Lack of ethical guidance can stymie academic SNS research, potentially rendering academia irrelevant to an important and growing domain of online activity, and educational researchers unable to utilize SNSs to create and study better learning environments. The private sector is charging ahead, creating de facto standards for data use. Privacy invasion, or the imposition of risk, is not made acceptable because somebody else is already doing it. However, the goals of most academic research surely have as much social value as the goal of selling more products to SNS users. Permitting marketers greater access than academic researchers to peoples' online information is a dubious ethical outcome. CSCL researchers who figure out how to productively navigate the complex array of issues raised in this paper will not only impact the kinds of technologies and contexts that we can create to support collaborative learning, but have an impact far beyond educational research, impacting fields like public health and political science, as well as shaping research policy for the 21st century.

References

Acquisti, A. & Gross, R. (2009). Predicting social security numbers from public data. PNAS 106(27). 10975–10980.

Backstrom, L. (2011) Anatomy of Facebook. Webpage of the Facebook Data Science team. https://www.facebook.com/note.php?note_id=10150388519243859 Accessed October 28, 2012.

Barab, S., et al. (2009) Transformational Play and Virtual Worlds: Worked Examples from The Quest Atlantis Project. IJLM: International Journal of Learning and Media, 1(2), published online.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012) A 61-million-person experiment in social influence and political mobilization. Nature, 489, 295-298.

Botkin, J. R. (2001) Protecting the privacy of family members in survey and pedigree research. JAMA 285, 207-211.

Boyd, D. & Hargittai, E. Facebook privacy settings: Who cares? First Monday, 15 (8), published online.

Brothers, K. B., & Clayton, E. W. (2010) Human Non-Subjects Research: Privacy and Compliance. American Journal of Bioethics 10(9), 15-17.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. The Journal of The Learning Sciences, 2(2), 141–178.

Bull, S. S., et al., (2011) The Cold Pressor Task: Is it an Ethically Acceptable Pain Research Method in Children? Journal of Pediatric Psychology, 36(10), 1071-1081.

Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), New directions in educational technology (pp. 15–22). New York: Springer-Verlag.

Department of Health and Human Services, Additional Protections for Children Involved as Subjects in Research. Federal Register 48, 9814 (March 8, 1983); codified at 45 Code of Federal Regulations 46, Subpart D.

Department of Health and Human Services, Federal Policy for the Protection of Human Subjects. Federal Register 56, 28003 (June 18, 1991); codified at 45 Code of Federal Regulations 46.

Department of Health and Human Services, Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators. Federal Register 76, 44512 (July 26, 2011).

Facebook, Information We Receive and How it is Used, http://www.facebook.com/about/privacy/your-info - everyoneinfo, visited January 5, 2012.

Facebook (2012). One billion – key metrics. http://newsroom.fb.com/imagelibrary/downloadmedia.ashx?MediaDetailsID=4227 Accessed online October 28, 2012.

Hampton, K. N. , Goulet, L. S., Rainie, L., & Purcell, K. Social networking sites and our lives. Pew Research Center's Internet and American Life Projects, Internet and American Life. Pew Research Center, Washington, DC, 2011.

Introne, J., Levy, K., Munson, S., Goggins, S., Wash, R., Aragon, C. (2013) Design, influence, and social technologies: techniques, impacts, and ethics. Workshop at the 2013 ACM Conference on Computer Supported Cooperative Work. http://distworkshop.wordpress.com/

Jianquiang, D. S., et al. (2011) Farmer's tale: a facebook game to promote volunteerism. CHI 2011, 581-584.

Lenhart, A., Purcell, K., Smith, A., & Zickuhr, K. (2010) Social Media & Mobile Internet Use Among Teens and Young Adults. Milleni-als: A Portrait of Generation Next Pew Research Center's In-ternet and American Life Projects, Internet and American Life (Pew Research Center Washington, D.C., 2010).

Lewis, K., et. Al. (2008) Tastes, ties, and time: A new social network dataset using Facebook.com. Social Networks, 30, 330-342.

Moher, T. (2006). Embedded phenomena: supporting science learning with classroom-size distributed simulations. CHI '06 Proceedings of the SIGCHI conference on human factors in computing systems.

Moreno, M. A., Fost, N. C., & Christakis, D. A. (2008) Research ethics in the MySpace era. Pediatrics, 121(1), 157-161.

Narayanan, A., & Shmatikov, V. (2009). De-anonymizing social networks. Paper presented at the 30th IEEE Symposium on Security and Privacy.

National Research Council, Learning Science Through Computer Games and Simulations. Committee on Science Learning: Computer Games, Simulations, and Education. M. A. Honey, M. L. Hliton, Eds. (The National Academies Press, Washington, DC, 2011).

Newell, B. C. (2011) Rethinking Reasonable Expectations of Privacy in Online Social Networks. Richmond Journal of Law and Technology 17, 12.

Ohm, P. (2010) Broken promises of privacy: responding to the surprising failure of anonymization. UCLA Law Review, 57, 1701-1777.

Shapiro, R. B. (2011). Anatomy Pro-Am. http://benshapi.ro/projects/anatomy-pro-am/

Shapiro, R. B., Squire, K., and the Educational Research Integration Area (2011). Games for Participatory Science. Educational Technology.

Steinkuehler, C., Duncan, S. (2008) Scientific Habits of Mind in Virtual Worlds. Journal of Science Education and Technology, 17(6), 530-543.

Yardi, S. & Bruckman, A. (2011). Social and technical challenges in parenting teens' social media use. In Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'11). Vancouver, Canada. May 7-12, 2011.

Zimmer, M. (2010) But the data is already public: on the ethics of research in Facebook. Ethics and Information Technology 12(4), 313-325.

Acknowledgments