The Meaning of Anonymity in an Information Age

Helen Nissenbaum

University Center for Human Values, Princeton University, Princeton, New Jersey, USA

Should anonymity be protected in electronic interactions and communications? Would this be a good thing for community, responsibility, free expression, political participation, and personal fulfillment? If so, when and why? These key normative questions probe the value of anonymity in our computerized society and political order. In this brief discussion, I do not directly address these important questions but address questions that undergird them about the meaning of anonymity in a contemporary, computerized society: What is anonymity? And what are we seeking to protect when we propose to protect it? Although answers to these foundational questions will not immediately yield answers to the key normative questions just mentioned, they are essential to understanding what is at stake in the answers to these questions. For, after all is said and done, we would not want to discover that the thing we have fought so hard to protect was not worth protecting after all.

The natural meaning of anonymity, as may be reflected in ordinary usage or a dictionary definition, is of remaining nameless, that is to say, conducting oneself without revealing one's name. A poem or pamphlet is anonymous when unattributable to a named person; a donation is anonymous when the name of the donor is withheld; people strolling through a foreign city are anonymous because no one knows who they are. Extending this understanding into the electronic sphere, one might suppose that conducting one's affairs, communicating, or engaging in transactions anonymously in the electronic sphere is to do so without one's name being known. Specific cases that are regularly discussed include:

- Sending electronic mail to an individual, or bulletin board, without one's given name appearing in any part of the header.
- Participating in a "chat" group, electronic forum, or game without one's given name being known by other participants.
- Buying something with the digital equivalent of cash.
- Being able to visit any Web site without having to divulge one's identity.

The concern I wish to raise here is that in a computerized world concealing or withholding names is no longer adequate, because although it preserves a traditional understanding of anonymity, it fails to preserve what is at stake in protecting anonymity. Why?

Information technology has made it possible to track people in historically unprecedented ways. We are targets of surveillance at just about every turn of our lives. In transactions with retailers, mail-order companies, medical caregivers, day-care providers, and even beauty parlors, information about us is collected, stored, analyzed, and sometimes shared. Our presence on the planet, our notable features and momentous milestones, are dutifully recorded by agencies of federal, state, and local government, including birth, marriage, divorce, property ownership, drivers' licenses, vehicle registration, moving violations, passage through computerized toll roads and bridges, parenthood, and, finally, our demise. In the great store of information, we are identified through name, street address, e-mail address, phone number, credit-card numbers, social security number, passport number, level of education, and more; we are described by age, hair color, eye color, height, quality of vision, purchases, credit-card activity, travel, employment and rental history, real-estate transactions, change of address, ages and numbers of children, and magazine subscriptions. The dimensions are endless (Nissenbaum, 1997).

Received 22 January 1998; accepted 1 September 1998.

Address correspondence to Helen Nissenbaum, University Center for Human Values, 5 Ivy Lane, Princeton University, Princeton, NJ 08544-1013, USA. E-mail: helen@princeton.edu

The Information Society, 15:141–144, 1999

0197-2243/99 $12.00 + .00

From these bits of information, public identities may be formed that are not only elaborate, but permanently accessible in an active electronic form for those who may need or want them. Even when these identities are not complete, and may in fact be quite fragmentary, inferential tools and network capabilities enable linking, matching, mining, and all the other activities that for one purpose or another transform bits of a person into a more complete, recognizable, possibly identifiable (virtual) person. Critically important to the question of anonymity is that these techniques allow linking of pieces and fragments of information; from a variety of pieces of information, or fragments of information, that are not each uniquely identifying, we may infer or link to those that are. For example, in most states, we can identify the owner's name and home address from the number on a car license plate; from a phone number we may reach a person or household; from an electronic mail address, or from an electronic pseudonym, we may be able to pinpoint a person's geographic whereabouts and physical identity.

Even where fragments of information do not lead to information that is uniquely identifying, people may be identified with a high degree of probability when various properties are compounded to include a smaller and smaller set of individuals who satisfy them all. If an unnamed individual, who regularly contributes to America Online discussion groups for Corvette owners and stamp collectors, reveals that he shops at Safeway, was born on 4 May 1965, graduated from Stanford in 1992, lives in Palo Alto in a three-bedroom house appraised at $525,000, and is divorced with two children in local public schools, we easily may be able to identify him without knowing his name. Although in the past the most direct and effective way of "getting at" a person was through his or her name, the electronic medium now offers many points of entry, some of which may be even more effective than a name. [Latanya Sweeney has carefully demonstrated this phenomenon in, for example, Sweeney (1997). Also, note close parallels to two of Gary Marx's categories of identification, namely, identification through distinctive appearance or behavior patterns, and identification through social categories (Marx, 1999).] Marketers use these techniques to track suitable targets to their home addresses by mining databases containing a diverse range of transactional information about them.

The power of information technology to extract or infer identity from nonidentifying signs and information has been inventively applied by literary scholars to settling disputes and unraveling mysteries of authorship—say, to discover whether it was Shakespeare who wrote a given sonnet. These scholars infer authorship by comparing the stylistic and lexical features of anonymous text with the known style of authors whose texts have been analyzed along these same dimensions. In a recently publicized case, Donald Foster, a professor of dramatic literature at Vassar College, identified Joe Klein as the author of the controversial political novel *Primary Colors* (Pristin, 1997), published anonymously. Foster also helps law-enforcement officials identify extortionists and kidnappers by analyzing what they have written.

Why does this matter? For situations in which we judge anonymity acceptable, or even necessary, we do so because anonymity offers a safe way for people to act, transact, and participate without accountability, without others "getting at" them, tracking them down, or even punishing them. This includes a range of possibilities. Anonymity may encourage freedom of thought and expression by promising people a possibility to express opinions and develop arguments about positions that, for fear of reprisal or ridicule, they would not or dare not take otherwise. Anonymity may enable people to reach out for help, especially for socially stigmatized problems like domestic violence, HIV or other sexually transmitted infection, emotional problems, or suicidal thoughts. It offers the possibility of a protective cloak for children, enabling them to engage in Internet communication without fear of social predation or—perhaps less ominous but nevertheless unwanted—overtures from commercial marketers. Anonymity may also provide respite to adults from commercial and other solicitations. It supports socially valuable institutions like peer review, whistle-blowing, and voting.

In all these cases, the value of anonymity lies not in the capacity to be unnamed, but in the possibility of acting or participating while remaining out of reach, remaining unreachable. Being unreachable means that no one will come knocking on your door demanding explanations, apologies, answerability, punishment, or payment. Where society places high value on the types of expression and transaction that anonymity protects (alluded to in the previous paragraph), it must necessarily enable unreachability. In other words, this unreachability is precisely what is at stake in anonymity. If, in previous eras, namelessness—that is, choosing not to reveal one's name—was the best means of achieving unreachability, it makes sense that namelessness would be protected. However, remaining unnamed should be understood for what it is: not as the end in itself of anonymity, but rather, the traditional means by which unreachability has been achieved. It has been the most effective way to keep others at bay, avoid ridicule, prevent undeserved revenge, harm, and embarrassment, and so forth.

In the computerized world, with the systems of information that we currently have in place, namelessness by itself is no longer sufficient for protecting what is at stake in anonymity. If it is true, as I have suggested, that one can gain access to a person through bits, or constellations of bits, of information, then protecting anonymity today amounts to more than merely withholding a name. It means withholding the information or constellation of information it now takes to get at, or get to, a person. When

we think of protecting anonymity we must think about this broader range of possibilities; we must think not only of how a person can prevent his or her name from being divulged, but how a person can prevent all the crucial bits of information from being divulged, especially the bits of information that when divulged would enable access to him or her.

Deepening our understanding of the issue of anonymity in an information age, and reaching wise decisions about it, will, in other words, require not only resolving the key normative questions stated at the beginning (to achieve a balance among potentially conflicting interests). It also requires an appreciation of what it takes to be "unreachable" or "out of grasp" in a world where technologies of knowledge and information are increasingly efficacious at reaching, grasping, and identifying. This is a moving target.

To secure the possibility of being unreachable, we need both to promote understanding and also pursue advocacy. Understanding may be achieved partly through a priori reasoning (figuring things out) and partly through increased knowledge about networks of information. People may figure out, either on their own or through the insights of others, how various pieces of information may link to their identities and whereabouts and therethrough defy the efficacy of traditional anonymity. Thus, a person may suddenly become aware that bar codes link to her identity when she pays for purchases with a credit card, figures out that electronic mail sent pseudonymously (under a fictitious name, frequently devised specifically for electronic communications) or anonymously may nevertheless yield identifying information about her via her computer's IP address, or realizes that she becomes more easily identifiable through an electronic mail address that includes information about her geographic location (for example, by identifying her place of work).

Beyond what we can figure out, there is a great deal we can learn empirically about the linkages that exist that may potentially undermine the possibility of anonymity (and pseudonymity.) In general, these linkages establish a correspondence between the sign under which people attempt to act and transact anonymously (or pseudonymously) and information about them that either itself makes people reachable, or links to other signs and information that ultimately link to information that makes them reachable. These revelations of identity may occur by various means. One is by linking the sign under which an anonymous person is acting into a network of information that ultimately leads to the person him- or herself. As discussed earlier, those whose business it is to watch, record, match, infer, and identify may manage to converge on individuals only with some degree of certainty, or they may manage to do so by linking ultimately to that one crucial piece of information—the work address, the IP address, the street address, the motor vehicle registration—that places the unnamed person within their reach.

Another way of defying anonymity, not yet discussed, is by breaking systems of "opaque" identifiers. What I mean by an opaque identifier is a sign linking reliably to a person—chosen, assigned, or arising naturally—that, on the face of it, carries no information about the person. That is, the opaque identifier holds no clue, by itself, as to the real identity of the person or how to reach that person. The chosen screen names (or pseudonyms) of Internet service subscribers may serve in this way as opaque identifiers. The Social Security number is an instance of an assigned identifier, and biometrics, such as fingerprints, retinal images, and DNA profiles, are instances of naturally occurring ones. As well as serving important societal needs, such as law and order, secure entry, and financial transaction, these systems of identification offer the means of dealing reliably but anonymously with individuals. For example, a professor wishing to announce course grades anonymously may list grades alongside Social Security numbers. People may interact with a stable cohort knowing only screen names and not real identities, and so forth.

Problems arise when the key to a system of opaque identifiers is compromised, as, for example, critics say has occurred with Social Security numbers. They charge that the mapping between these numbers and information that allows people to be reached has seeped slowly but surely into the public domain. The Social Security number has become a sure-fire way not only to "get at" a person but to extract an enormous array of other information that has been keyed to it. In other cases, a key to the mapping can be less inadvertantly and more directly betrayed, such as occurred in a controversial case involving Timothy McVeigh, a member of the U.S. Navy. Navy personnel, investigating his alleged homosexuality, managed to elicit McVeigh's real identity from America Online on submitting his screen name, "boysrch" (*McVeigh v. Cohen*, 1998). It is of great importance that people at least have an accurate grasp of the existing level of integrity for each of these systems of opaque identification.

My purpose here is not to suggest that anonymity in an information age is impossible. I am mainly arguing that achieving it is a more demanding business than merely allowing people to withhold their names. Although I do not mean to imply that contemporary networks of information, and the compromise of opaque identifiers, are the result of insidious conspiracy and subterfuge, I recognize, at the same time, that all interests are not equally served by promoting a sufficient public understanding. It is this level of understanding that would make people more cautious, more guarded, more mindful of the information they divulge to others in various transactions,

and, as a result, more capable of protecting the possibility of anonymity. The understanding may also lead them to realize that anonymity and pseudonymity are not all-or-nothing qualities but can be achieved in degrees and through layers of cloaking. But public understanding is not, in my opinion, enough. Knowing where landmines are buried can help people avoid them, but clearing the landmines is a more robust and lasting solution.

Beyond the effort it would take to educate toward a more comprehensive understanding, we will need to pursue lines of advocacy. If, as a society, we agree that what is importantly at stake in anonymity is the capacity to be unreachable in certain situations, then we must secure the means to achieve this. This will include a dramatic reversal of current trends in surveillance, as well as a relentless monitoring of the integrity of systems of opaque identifiers. Without at least these measures, even if we nominally secure a right to anonymity through norms and regulations, we will not have secured what is at stake in anonymity in a computerized world.

REFERENCES

McVeigh v. Cohen Decision. 1998. U.S. District Court, District of Columbia.

Marx, Gary. 1999. What's in a name? Some reflections on the sociology of anonymity. *The Information Society* 15(2):1–15.

Nissenbaum, Helen. 1997. Toward an approach to privacy in public: Challenges of information technology. *Ethics & Behavior* 7(3):207–219.

Pristin, T. 1997. From sonnets to ransom notes. *The New York Times*, 19 November:B1.

Sweeney, Latanya. 1997. Weaving Technology and Policy Together to Maintain Confidentiality. *Journal of Law, Medicine and Ethics* 25:98–110.