The Cutting Edge

Helen Nissenbaum
University Center for Human Values
Princeton University

Values in the Design of Computer Systems

There are many points of entry to the study of social, ethical, and policy dimensions of information technology. In some of my work, where I have been concerned with implications or consequences of computerization for society and moral values, I have examined such issues as accountability and property rights. Within this same paradigm, I am currently examining moral foundations of a right to privacy that would account for the violation we feel in response to contemporary practices of public surveillance. Recently, however, I have also been intrigued with another, somewhat different approach, one that is concerned with the implications of societal values for information technology, and in particular with the idea of values embedded, or embodied, in the design of computer systems. In this column, I briefly describe past, current and planned work within this alternative framework.

My interest in studying values embedded in computer systems, inspired by work of scholars in the area of science, technology and society like Wiebe Bijker and Langdon Winner, grew directly out of a joint project with Batya Friedman. This project examined the problem of bias in computer system design, taking off from work by the philosopher James Moor, who argued that software could embody biases, empirical work by Chuck Huff & Joel Cooper, which pointed to the potential of gender bias in educational software, as well as results of our own investigation of a number of systems. Convinced that bias was a particularly important issue in a society (like ours) which places high value on justice and fairness, we devised a definition of a biased system—namely, a system that systematically and unfairly discriminated against some in favor of others. We also developed a theoretical framework which identified three categories of bias—namely, preexisting bias (reflecting biases preexisting in society), technical bias (arising from technical constraints), and emergent bias (arising as a result of contextual shift). Finally, on the basis of our framework, we suggest ways of remedying bias in computer systems.

During the course of this project, we became convinced that bias was a particular instance of the more general phenomenon of values embedded in computer system design. This realization, in turn, has suggested three lines of further research. One maintains the level of particularity of the bias study extending its scope to other values. A second seeks to develop a methodology for computer scientists and engineers interested in values. A third, examines emerging themes and questions. Ongoing work that Friedman and I are conducting together as well as independently has concentrated on the first of these two lines, but I find the general, philosophical questions, increasingly compelling and requiring attention. In the remaining paragraphs, I briefly discuss all three.

Other Values.

Following the methodology we developed for bias, Friedman and I are conducting a similar study of user-autonomy, which would explicate how system design may either enhance or inhibit the autonomy of users. We chose to study autonomy because, like bias, its embodiment in technology may significantly affect individuals and society. Furthermore, a tradition of intellectual criticism sees human autonomy as particularly vulnerable to technological advance. In the case of computer technology, at the same time that its champions tout its capacity to control human affairs, its detractors, worry that this same capacity bodes ill for humanity. As computer control increases, human control and autonomy would diminish.

As we did with bias, we seek to link specific technicaal features of design with autonomy. Work in progress indicates, not unexpectedly, that control is an important mediator of user-autonomy but it also suggests that this relationship is more indirect and complex than one might have expected.

Methodological Implications.

Beyond theoretical analysis, we are interested in developing methodological tools for systems designers. This explicitly normative dimension to our work urges computer scientists and engineers to think beyond the normal set of technical criteria of excellence such as power, efficiency, and correctness, and toward an augmented list that would include consideration of moral values such as bias and user-autonomy. We see its capacity to provide practical guidelines to designers as a measure of the success of our work. Putting this to the test in a project recently started, we are collaborating with computer scientists at Princeton to strengthen systems

of web security through technical improvements that, at the same time, are sensitive to values.

As a philosopher, my central role in the practical side of this work has been to interpret concepts of value. The great challenge of this role is to work with deep, rich, and nuanced conceptions of values such as justice, responsibility, and autonomy, that have developed over decades, even centuries of philosophical and political thought. Out of these conceptions, in order to be able to map values to characteristics of computer systems, I must construct concepts that are operational within a practical setting, create precision where none naturally exists. While certain subtleties are necessarily compromised, it is not (I hope) at the expense of the essence of the original conceptions.

Philosophical Themes and Questions.

This approach opens the way to a set of questions that might not normally arise in the study of consequences or implications of information technology. Where the latter creates an impression of technology as a given, the former highlights the malleability of design. It projects the possibility of values injected into a system during construction, a possibility that leads naturally to conceiving of ways to intervene in the design process to promote better (morally speaking) technology. As described earlier Friedman and I, and others, have taken this route. (See work by Philip Brey at Technical University Twente, The Netherlands. Also, see work by Deborah Johnson examining the meaning of values embedded in design.)

This consequence will only be cause for rejoicing after we have thought carefully through a number of related issues, including the various possible sources of values in technology, as well as their legitimacy. One important source, public policy, is already widely studied. The work described here focuses on the phase of engineering design and development, implicating designers and builders of systems — scientists and engineers — whether their interaction with values is conscious and explicit, or unconscious and inadvertent. If it is true that values may be incorporated even at the fine-grain level of design and development, then nothing short of a deep appreciation of the technical character of a system will suffice to articulate a full picture of embedded values.

But are computer scientists, systems engineers, and designers, ready and able to shoulder this responsibility? And if yes, what about legitimacy? Is it legitimate for scientists and engineers to be affecting values in this apparently undemocratic way? Who is to say that the values embedded in a system are the "right" values? These questions fill out a research program driven by the idea of values embedded in computer systems. I pose them to myself and hope others, too, will pursue them.

Selected References

Brey, P. The Politics of Computer Systems and the Ethics of Design, Computer Ethics Philosophical Enquiry, *Proceedings of CEPE97*, Erasmus University Rotterdam 11-13 June 1997. An ACM/SIGCAS Conference, ed. M.J. van den Hoven

Huff, C. and Cooper, J. Sex Bias in Educational Software: The Effect of Designers' Steriotypes on the Software They Design. *Journal of Applied Social Psychology* 17 (1987), 519-532

Johnson, D., Is the Global Information Infrastructure a Democratic Technology?. *Computers and Society*, Vol 27, no 3, September 1997. 23-26

Nissenbaum, H. and Walker, D. Will Computers Dehumanize Education? A Grounded Approach to Values at Risk, *Technology in Society* (forthcoming)

Nissenbaum, H., Toward an approach to privacy in public: the challenges of information technology, *Ethics and Behavior*, 7(3), 207-219

Friedman, B. and Nissenbaum, H., Bias in Computer Systems, *Transactions on Information Systems* July 1996.

Nissenbaum, H., Accountability in a Computerized Society. *Science and Engineering Ethics* (1996) 2. 25-42

Nissenbaum, H., Should I Copy My Neighbor's Software? *Computers, Ethics, and Social Responsibility*. D. Johnson and H. Nissenbaum (eds) Prentice Hall. 1995

—Please send any communications to: Helen Nissenbaum, University Center for Human Values, Marx Hall, Princeton University, Princeton, NJ 08544. Email: helen@princeton.edu