This work originally appeared in:

Dennett, Daniel C. "Formulating Human Purposes: Meta-Engineering Computers for People." In Information Processing 83, edited by R. E. A. Mason, 253-258. Amsterdam: Elsevier (North-Holland), 1984.

This is Daniel C. Dennett's final draft before publication. It has been modified to reflect the pagination of the published version of the work.

Meta-Engineering Computers for People

Daniel C. DENNETT
Department of Philosophy, Tufts University
Medford, Massachusetts 02155, USA

Invited Paper

The effort to consider the unintended side effects of the spread of computer applications in society has not proceeded with the energy, organization, or breadth of vision the topic requires. Two exemplary problem areas are discussed: speech recognition systems and expert systems. Both are found to portend serious social problems unless quite vigorous action is undertaken soon. Proposals for action are briefly sketched.

Almost twenty years ago Norbert Wiener foresaw a problem that is widely—even incessantly--recognized today, but still not being addressed with sufficient energy and organization:

"As engineering technique becomes more and more able to achieve human purposes, it must become more and more accustomed to formulating human purposes." (Wiener, 1964)

But surely, you may be thinking, nothing today is more vigorously being pursued in hundreds of research and development labs than "user-friendliness." And what is user-friendliness but the degree to which a computer is designed to meet the user on his or her own terms, to accommodate the hardware and software to the purposes and concerns and styles of the human user? And isn't it true that thousands of computer experts are ingeniously and insightfully examining virtually every aspect of human life in search of computer applications that will enhance human potential? What more could one ask of engineers in the way of "formulating human purposes"?

Quite a lot. My rather delicate task on this occasion is to take advantage of your invitation to speak, and to try to persuade you that all these efforts, excellent as many of them are, are not enough. Since I am not a computer scientist but a philosopher, I will--and should--be treated with skepticism. Who am I to tell the experts how to do their jobs? All I can reply is that, having tried to inform myself about these important matters, I find myself in the classic philosopher's position: I have some questions to ask, in the hope that we will all learn something in the process of answering them.

If the so-called Computer Revolution proves to be like all earlier revolutions, both political and technological, its intended, planned, foreseen effects will be overwhelmed, in the end, by the unanticipated and unintended side effects of its campaigns. And as with earlier revolutions, any hope of seizing control of it is out of the question; the most we can realistically hope to do is to steer it a little bit, here and there.

The comfortable response to such observations is stoic resignation and passivity, but there is really no excuse for it. If we care to, and if we organize our efforts with the same energy and intelligence that are normally poured into the development of a new computer, or weapon system, or political campaign, we can in fact achieve dramatic and almost instantaneous results worldwide. The computer revolution is, after all, a social and economic phenomenon, not a geological or climatic trend, and hence it is exquisitely information-sensitive. The apparent momentum and direction of the computer revolution is not a physical feature that can change only gradually and only by an enormous input of energy; a few well chosen and well placed words and deeds could swiftly change its trajectory. But what steering projects might we want to undertake?

Probably no professionals are more keenly aware of the problems surrounding unintended side effects than computer scientists, but isn't it the case that the standard, received practice is to narrow one's gaze when one goes hunting for troublesome side effects? There are many ways of narrowing one's gaze: not considering anything but those side effects that

are already clearly measurable and clearly detrimental to the "performance" of the system, not considering any effects for which one could not be held liable in a court of law, not considering any effect that would be outside one's professional expertise to assess, not considering the social impact of one's projects at all.

When I purchase a home computer for my children, I am not particularly concerned that it will emit harmful radiation, or ruin their eyesight, or produce blisters on their fingertips. The manufacturers can be counted on to have protected themselves by producing a product that is very probably locally safe, that has no immediate, assignable harmful effects on users. Of course long term cumulative effects are another matter, and no doubt there is careful research on long term effects currently being conducted by the major manufacturers. I am less sanguine, however, about the direction of their research when it comes to what my wife calls the goldfish problem. A child can buy a live goldfish, in a plastic bag of water, for less than a dollar. By the time that goldfish has been surrounded by a suitable tank, a filter, an aerator and plants, there is a considerable investment. Computer manufacturers play this game with particular skill, of course. They also play their own version of the encyclopedia salesman, who for a few dollars a month can give your children an educational advantage worth its weight in gold. Do the manufacturers also have research groups working on the possibly harmful effects on their customers of such seductive traps?

Once we acknowledge the simple fact that they are, after all, commercial enterprises engaged in making a profit, we are bound to recognize that they cannot be relied upon to do all the examination of side effects we might want done. Powerful economic motivation will of course bias decision-making in industry, not because industry is evil, but just because it quite correctly sees that not all things worth doing are its particular responsibility to do. Insofar as the "environmental impact" of a new sort of product is elusive, not directly assignable to one's own brand, and sufficiently hard to predict so that no court could find one negligent for failing to make the prediction, one can argue that it would not be reasonable or acceptable stewardship of the corporation's assets to devote large sums of money to studying it.

I don't take myself to be telling you anything you don't already know. I'm just reminding you that the Platonic Ideal of how engineers and computer scientists sit down with their clients, humanity, and work out well-designed solutions to well-formulated problems, is an ideal seldom even approximated in the real world. Norbert Wiener was fond of citing the classic ghost story, "The Monkey's Paw", which repeats the ancient theme of Aladdin and his lamp. Like many other apparently fortunate characters in many other stories, the protagonists have the magic power to make their wishes come true, but they are held to the precise letter of their declared wishes, and because they fail to see the implications of their explicit choices, they come to regret most bitterly the effects their magic creates for them. But of course the genies in these stories can always defend themselves: you got just exactly what you contracted for, they can say. Caveat emptor. These stories would all be much more boring if the genies stopped their clients, and helped them to consider, cautiously and carefully, the implications of their desires. In fact, one wonders if the stories would ever get going, for there is no end to the reflection one can engage in when one's fundamental purposes and projects are called into question. It is much more fun, almost irresistibly more fun, to take the customer at his word and get on with the project, wherever it leads, comfortable in the illusion that one is not responsible for what happens, since "the customer is always right".

It is this comfortable insulating barrier we erect between our professional projects and the larger context in which they fit that creates the narrow gaze that concerns me. It's not much of a problem for a philosopher, usually, for our projects are notoriously ineffectual and arcane, usually touching only a few of our fellow human beings. Bu tit can be a problem even for us. Tradition has it that Nietzsche, hitting upon his truly depressing vision of the "eternal recurrence"--the idea, which he thought he had proved, that the history of the universe in its minutest details must repeat itself endlessly--gave very serious thought to the question of whether he should suppress this discovery, for the good of mankind. Perhaps this was ludicrous; perhaps this was a comical overestimation of the likely impact on society of a curious "result", of merely theoretical interest, and only of interest to like-minded thinkers, but perhaps not. In the Twentieth Century we have seen enough instances of "pure" ideas having impure results to learn we should not jump to conclusions.

Last year at a conference where I presented a paper on my materialist (and AI-inspired) theory of conscious experience, I was completely taken aback when an entirely serious and highly intelligent member of the audience asked me, point blank, if I didn't think I should suppress my own theoretical researches in this area, since their results seemed (to him) to be both immensely plausible and extraordinarily depressing and pessimistic. At first I couldn't imagine that anyone who understood my view could be anything but as elated and curious as I was to work out the details. What was depressing about it? What possible ill effects could the promulgation of a theory of consciousness ever have? I am still inclined to scoff at the idea, but I had to admit that it had simply never occurred to me that the question of the social impact of that work was worth a moment's thought. After all, one risks making a fool of oneself and seeming terribly self-important if one announces to one's colleagues "I have decided to take some time to examine the social and ethical consequences of my intellectual projects." So like most of us, I had insulated my professional projects from my political and ethical beliefs. My excuse was that I am a worldly enough philosopher to have a rather low opinion of my likely influence in the "real world". Certainly anyone who is professionally involved with computers is worldly enough to have a rather higher estimation of their potential influence.

This can put a computer scientist in a distinctly painful predicament. Consider those who work on speech recognition programs, for instance. No area in AI has been more generously funded in the last decade than speech recognition and natural language processing systems, and it is easy to see why. A computer-based device that can turn spoken input into written input-or use it to control other processes-could in principle render obsolete the worldwide armies of stenographers, telephone operators, travel agents-and perhaps more specialized personnel as well.

The international research effort in speech recognition systems involves some of the brightest people in the field of computers, and now that the Japanese Fifth Generation Computer project has announced speech-recognition as a major goal, we can expect even more research energy will be devoted to this project in the immediate future. If it were to succeed, the enormous impact on the world's labor force, armies, and heaven knows what else would be something to ponder indeed, but I am not going to explore that topic on this occasion, because it is no doubt familiar to you and moreover a distant and rather unlikely prospect in any case. Instead, I want to draw your attention to a more pressing reason for concern--first pointed out, I believe, by Joseph Weizenbaum: crude speech recognition systems already exist, and while the level of improvement required to make a practical robot stenographer is not even on the horizon, current research directions could lead in the not too distant future to other applications with ominous consequences.

What would these applications be for? Surveillance. I have no idea how much wiretapping and hidden recording is currently being done around the world, but certainly it is an appalling amount. At present the effective limiting factor on the amount of electronic eavesdropping done is simply the tremendously time-consuming task of monitoring all the recordings collected. For every 5000 hours of recorded conversation one gathers, it takes roughly 5000 person-hours of listening to extract the usable (and abusable) knowledge from the huge store of raw data. This ratio can be pared down somewhat with the new computer- based speech compression technology, which can speed up recorded speech to several times its normal rate without serious loss of comprehensibility, but still-happily--the personnel cost of increased surveillance rises linearly with the amount of surveillance attempted. Hence it is extremely costly to expand such surveillance-especially when one considers that the personnel involved have to be alert, intelligent, and trusted. It must be very hard to find people who are "good" enough to do this work but who can also tolerate the boredom and human unpleasantness of searching for the seditious needles in the haystacks of quotidian trivia that make up the bulk of human conversation. Thank goodness.

It would also be a preposterously expensive project, given the current state of the art in artificial intelligence, to attempt to build an artificial speech-comprehending monitor to replace those human agents. Thank goodness again. But suppose you could put those miles of tape through a crude filter that would more or less reliably screen out 90% or 99% of the irrelevancies--and at speeds faster than human speech recognition, and without deterioration of performance due to boredom. You could increase surveillance, and surveillance effect-

tiveness, by orders of magnitude with no important increase in personnel. Such a crude filter, tunable to different keywords and topic areas, with a variable "relevance" threshold, is not far beyond the current state of the art.

So far as I know, drawing from public sources of apparent reliability, computer-based keyword monitoring of nonvoice electronic transmissions is already the normal state of affairs in advanced Western governments, and voice traffic is heavily monitored by teams of human agents whose task is rendered as efficient as possible by using whatever relatively cheap tricks are apt to work well enough: automatically deleting blank periods, computer-based speech compression, selection by called and calling number, and so forth. And it does seem clear that in England at least, the Joint Speech Research Unit of GCHQ in Cheltenham conducts both published and classified research on computer-based speech recognition systems. So there is little doubt that a well-financed niche exists for such technology to settle into just as soon as it becomes technically feasible. It cannot, then, be viewed as an unlikely development, given the current trends. It would moreover be a very bad development for computer scientists to have on their consciences.

Now I do not mean to suggest that computers in intelligence gathering or in military and police applications are always a bad thing. There are some projects in these areas of which I (cautiously, tentatively) approve. But the particular prospect of vastly increased surveillance, threatening the currently maintained security and anonymity of voice telecommunications (and other private conversations) is chilling indeed. In the security-versus-surveillance "arms race", the deployment of such content-sensitive searching filters would be a major destabilizing escalation. If the difference between character-recognition and speech-recognition by computer does not yet strike you as a difference with major social import, reflect on the difference it would make to your life to have live-in servants in your home (1) who were deaf, (2) who didn't know your language of daily conversation, or (3) who were fluent in your language. Would the benefits of (3) clearly outweigh the costs? How drastically, in any event, would you have to reorder your regular behavior to bring the unwanted consequences of (3) down to "acceptable" levels? Would any other envisaged "invasion of privacy" be more severe than this?

Are there in fact any benefits to be derived from the development of versatile speech-comprehending computers that could outweigh the opportunities for abuse? Why on earth do the Japanese think this is a good idea? It is often suggested that workers who must use both hands and concentrate their visual attention on tricky jobs could greatly benefit from being able to direct robotic assistants by voice command. Suppose they could; are there not many other ways the same effect could be achieved? Wouldn't it be almost trivial to design a whistle the workers could keep in their mouths and blow signals on? The fact that voice commands would be easier or more versatile is by itself no argument. No doubt many people could perform their daily tasks more safely and efficiently if they had human slaves at their disposal, but that hardly would justify slavery.

Probably the defenders of speech-recognition system development can think up some applications where nothing short of versatile speech recognition would be the appropriate interface between person and computer. That would not settle the issue, for then we should want to know whether these applications are after all worth it. When the stakes are this high, the social utility of introducing the system anywhere must be great indeed to overcome the anticipated risks.

And why, one might well pause to ask, should anyone be so eager to dream up justifications for this particular technology? There's a very good reason—almost as good as any I know: sheer intellectual, scientific curiosity and enthusiasm. My own research has occasionally brought me into happy collaboration with researchers in natural language processing, and as I share their fascination, I must admit that I also share their reaction when it is pointed out to them that their deliciously interesting "pure" research might have a social effect so dire as to warrant abandoning the research: irritation, dismay, and extreme reluctance even to consider turning to other lines of research. One turns almost irresistibly to wishful thinking. Perhaps there are safeguards, one thinks; or perhaps we can do the theoretical research without building the hardware--or enabling others to build it. Besides, one is tempted to add, this is an arms race of sorts, and if the other side gets this capability first…

Well let us see how plausible these themes of wishful thinking are. Can anyone actually imagine the sorts of safeguards that could prevent the misuse of speech comprehension systems if they existed? The problems are somewhat different, but when one reflects on the present escalation of computer crime versus computer security, one is led, I think, to agree with the observation of Colin Cherry, who suggested that information is inherently leaky; that the cards are stacked against any security system, however sophisticated.

What then of the prospect of doing the fascinating theoretical, scientific work but taking steps to prevent that work from being harnessed by engineers into usable systems? While one may on occasion be able to erect such a barrier between, say, physics and engineering, computer science is already so much an engineering enterprise that there is often scarcely any room at all between the basic research and the application. Can anyone imagine a way of isolating this research from applications?

Finally, what are we to make of the familiar idea that our own research is a defensive measure, against the evil technology of the other side? Here a version of Colin Cherry's observation suggests, all too plausibly, that there simply is no way to hold back the tide of scientific or technological discovery, and that while individuals can choose not to risk soiling their own hands with a particular bit of regrettable technological advance, there is no practical action they can take beyond that to bottle up a "secret" forever. Even if this is so, however, there may be a great positive value simply in delaying the onset of some such effect, until we can better think through ways of accommodating it. There might be a real project beyond moral handwashing to consider here.

These considerations are not cheering, but if they are realistic, we had better acknowledge them as a first step towards seeing what might be done. One fact seems clear: the logical motivation driving these technological developments--the motivation at the point of research and development—is very powerfully in favor of pushing ahead, and hence one cannot reasonably expect the practitioners themselves to engage in a whole-hearted and properly cautious assessment of the likely impact of their efforts, no matter how strongly the wider public might feel about the dangers--unless there are some dramatic changes in the way people think about these issues.

I now want to turn to the development of expert systems, an area where the forecasts are not so gloomy, both because the threats are less ominous and irrevocable, and because there is apparently so much more room for meaningful action. At first glance the most benign application of expert systems to date would appear to be in medicine, where tireless specialist programs will soon be able to "consult" with hundreds of far-flung general practitioners every day. While the prospects for socially valuable applications in this area seem immense, there are some nagging worries well worth careful exploration. For instance, the introduction of expert systems in medicine will probably accentuate, rather than ameliorate, several well known difficulties in medical education. The education of doctors is a highly pragmatic—not deeply intellectual – educational project. There is a considerable pressure on doctors to learn habits of thought that are efficient, swift, and--to put it bluntly--as superficial as safety will permit. So doctors are, as a group, prone to taking advantage of any short cuts that they can convince themselves are authoritative. They are apt to be suckers, in short, for expert systems.

Will their own diagnostic skills, their deepest understanding of medicine, be jeopardized by the wide adoption of expert systems? There is reason to fear it. The trend is already discernable in engineering, which relies heavily on computer-aided design techniques, and is a profession intellectually quite similar to medicine. Some years ago a major builder of jet engines ran an informal experiment: its top design team (then already equipped with the finest computer-based design aids of the day) was given a delicate problem in turbine compressor design that had been elegantly solved by traditional methods some years earlier. On several occasions this design problem was presented to the engineers and solved, but none of the new solutions were as good as the original solution. Why? Apparently the engineers were no longer thinking deeply about the problem; they were letting their computers churn out candidate solutions by a crude trial-and-error process. When their systems had worked long enough in their opinion, they took the best fruits and decided these were as good as they could get. Now there is surely a major place for such prosthetic extension of engineers'--or doctors'--talents. Not

every design or diagnostic problem requires the best efforts of human genius. This is particularly true, no doubt, in medicine. If the shortage of highly trained doctors in underdeveloped parts of the world can be met in part by networks of expert systems interacting with paramedics and nurses and other moderately trained personnel, this will no doubt bring much better medical care to vast numbers of people. But if we don't take steps to differentiate areas of application, and if we ignore or deny the fact that those who rely on such systems are paying a heavy cognitive price for their dependence, something like an atrophy of human understanding (and concomitantly, human control) can occur.

We might also ponder in advance the justice in the charge that will be leveled against those who saturate the Third World with expert systems: intellectual imperialism, designed to create an abject dependence on a technology that must be purchased from, and serviced by, the technocratic nations, who thereby increase the gap between those peoples who have some measure of control over their destinies and those who are in many important regards at the mercy of others.

Expert systems and other "knowledge engineering" applications are currently a commercially hot trend, with venture capital spawning new firms with almost the same enthusiasm that greeted gene-splicing a few years ago. The comparison with the recent experience of the molecular biologists, whose field is, like nuclear physics and computer science, a "scare technology", suggests a timely project. In 1975 at Asilomar in California, the world's leading molecular biologists gathered for an intensive workshop conference out of which emerged principles for an international canon of procedures and safeguards for recombinant DNA research.
There was considerable initial skepticism among those leaders about the value of that conference at the time, but in retrospect most, I think, would agree that it was an extremely important step to have taken--and taken in the nick of time. Not because biological catastrophe was just around the corner, but because no one could in good conscience say at the time that it wasn't. Having taken some extraordinary, time-consuming steps to put their house in order, they could then go back to work--not just "business as usual" but a significantly improved version of business as usual—no longer hiding behind the pretext that they were just doing their narrow, technical jobs, giving the customers what they were asking for.

A comparable project in expert systems and speech recognition systems (to speak only of the examples I have discussed) would require, as Asilomar showed, the cooperation and active participation--however begrudging—of the leading workers in the field. No other participants (certainly not philosophers and politicians and citizen action groups) could have the expertise (and the natural authority that springs from that expertise) to produce a result that would be anything other than just another policy study to gather dust on library shelves.

If IFIP wants to take seriously its commitment to "Social and Economic Implications", as the formation of this new section suggests, it should recognize that bringing philosophers and social scientists and political activists to speak to it, while perhaps a good first step, is no more than that.