Daniel C. Dennett

Information, Technology, and the Virtues of Ignorance

When I was about ten years old, I read *Robinson Crusoe* for the first time and could scarcely contain my delight with the ingenious and resourceful ways Crusoe transformed his island world, bending its inventory to his own purposes, surrounding himself with contrivances of his own design and manufacture for enhancing his powers, and providing for his safety, nourishment, and pleasure. I discovered right then that I was in love with technology, as no doubt many of you are. We should recognize—we technophiles—that our love affair with technology, like all good love affairs, is not entirely rational, try as we may to rationalize our devotions. Crusoe the technocrat—it is one of the great fantasy themes, right up there with sexual adventure, athletic triumph, being able to fly, being invisible. It rivals them all in its captivation, in the luxuriousness of its details in our minds' eyes. It is far more satisfying than magic, precisely because it is *not* magic; it is something *we create,* and hence are presumably responsible for; it is something we *understand,* presumably, and hence, presumably, control.

We live today in a wonderful world in which the fantasy of unlimited technological enhancement seems to be coming true. This is convenient for us technophiles, for we can point to the many blessings our loved one has provided for us all—even the ingrates. Which technophobe would choose to live the harrowing and desperate life of the medieval peasant or the Stone Age hunter? Yet, like Crusoe, we pay some price for whatever we gain, requiring some minor revision in our habits, some curtailing of our options, some petty irritations that crop up as side effects with the adoption of each new marvel. Some find it easy to push these slightly nagging debits into the background; they say "Yes, there are costs of course, but it's not worth our while trying to sum them—whatever they come to, they are a small price to pay for the obvious gains." Others find their uneasiness harder to quell; they wonder if, when the total costs of some of the more indirect effects of new technology are rendered, we may not find that we have moved imperceptibly into a world we do not know how to inhabit and cannot leave.

I propose to focus attention on certain troubling aspects of this relationship between technology and morality.[1] I wish to consider the possibility that information technology, which has been a great boon in the past, is today poised to *ruin our lives*—unless we are able to think up some fairly radical departures from the traditions that have so far sustained us.

We all want to lead good lives—in at least two senses. We want to lead lives that are interesting, exciting, fulfilling, and happy, and we want to lead lives that are morally good as well: we would like to be useful, and to make a difference—a difference in the right direction, whatever direction that is. There is no doubt that technology in the past has facilitated both these aspirations, freeing us from drudgery and misery, making it possible for many of us to improve the lives of others. But unless we can find solutions to certain problems, the curve will turn. We have reached a point where the advance of technology makes the *joint realization* of these two goals less likely—we may have to make an unpalatable choice between lives that are morally good, and lives that are interesting.

Since my message is one that many may find unpalatable, it may be useful for me to start with a very specific instance of the trend that concerns me, before drawing on it for wider implications. As technology imposes new sources of knowledge, it renders obsolete the standards that guided our individual actions in the past. Consider, for example, the case of the rural doctor. Today, there are doctors who have chosen, commendably, to forsake lucrative urban or suburban practices for more valuable and meaningful lives as doctors in small rural communities. Their virtues have often been sung;[2] there is little need for me to dwell on them. These doctors know their patients well; their personal, intricate, involved knowledge stands them in good stead when they come to diagnose, treat, and advise the members of their communities.

Such doctors, for better or worse, are an endangered species. Technology is on the verge of rendering their style of medical treatment obsolete, and—because of its obsolescence—morally indefensible. As expert systems for medical diagnosis become available, these doctors will have to decide whether to avail themselves of the new technology. Let us suppose, for the sake of argument, that the systems will work as well as their supporters claim; they really will provide swift, reliable, and accurate diagnoses of ailments across the wide spectrum of medical cases the average physician is likely to encounter.

If so, the doctors, in good conscience, will have no choice: they will have to avail themselves of the new expert systems. To choose not to equip themselves with the best available means of securing accurate diagnosis would be a gross dereliction of duty, just as if—for some romantic whim—they chose to deny themselves use of a telephone, or insisted on making their rounds on horseback, or refused to consult x-rays before operating. Quaintness is acceptable when matters of life and death are not at stake; but few would be enthusiastic about a doctor who insisted on relying on old-fashioned methods, particularly if it entailed a serious and avoidable risk of misdiagnosis or mistreatment.

Doctors have always been obliged to keep up-to-date with their medicine, and typically have responded to this obligation with fairly serious efforts to stay abreast of medical journals, to take refresher courses. The generation of information has recently been getting out of hand; until now, rural doctors have been excused from knowing everything that their urban colleagues were held responsible for—there are limits to what people can be expected to carry around in their heads.

Now, however, a technology is promised that will render such limits obsolete. All you will have to do is install a modem and a cellular telephone in your four-wheel-drive van and there, at your fingertips, will be a credible approximation of the finest corps of specialist consultants, available twenty-four hours a day. You would have to take a curious moral stand indeed to resist becoming reliant on such a system. How dare you turn your back on such a fine new source of information, when lives—lives entrusted to you—depend on you making the best informed diagnoses of which you are capable?

The standards of excusable ignorance for even the most isolated of rural doctors will shift, and the doctors will be required to alter their practices to meet wholly new standards. All doctors will be expected to avail themselves of the new technology, just as all doctors now are required to maintain standards of antiseptic practice. We may suppose that expert systems will enable doctors to practice much better medicine, but in order to use these systems they will have to relinquish some practices they may well have prized in their earlier *modus operandi.*

At present, rural doctors can take a varied, even informal, approach to gathering facts about their patients. If old Sam looks OK, sounds just about the way he always sounds at this time of year, and does not complain about anything new, the doctor can leave well enough alone. Besides, if there really is anything new wrong with old Sam, it is too esoteric, or too obscure at this stage, for the rural doctor to be expected to diagnose it. After all, rural medicine is not the Massachusetts General Hospital. But expert systems will change all this. Doctors will be obliged to ask all their patients a battery of questions they never felt the need to ask before—for what use could they have made of the answers?

They will also be obliged to perform a variety of largely simple tests they never felt bound to perform before. They will do so because the feeding of expert systems with such data will supposedly have proven to bear valuable results—permitting a higher rate of early diagnosis of treatable cancer, for instance. Gathering information by these two methods—asking questions and performing simple tests—will be made as easy, straightforward, unequivocal as possible. Indeed, the procedure will be *as routine as possible,* for the more routine it is, the more uniform the feeding of the expert systems will be, and hence the less likelihood there will be of misinforming them.

In this way, the "art" of diagnosis, and the "art" of "taking a patient's history" will be reduced, as far as possible, to an exercise in which the art is displaced by the mere capacity to follow directions. I am not claiming that such systems would place a positive value on the deliberate suppression of imaginative, artful investigation and diagnosis, but just that such activities would be relegated to whatever room remained *after* the doctors had done their duty by asking all the *obligatory* questions and performing all the *obligatory* tests.

Since "progress" in medicine (and technology generally) proceeds by replacing art with obligatory practices whenever the principles governing the art can be well enough understood and justified to be codified, we can expect that, insofar as the technology of medical diagnosis succeeds, insofar as it becomes so demonstrably reliable that doctors will be obliged to use it, it will do so by diminishing the regular, daily contribution of the medical practitioners who use it. Once in a while, the artful doctor may find a moment in which to exercise his or her art, and even save a life by filling a gap in the technology, but such opportunities will become rarer as the technology improves.

A subspecies of doctor will thus become extinct, succeeded by a new species that will delegate more and more diagnostic responsibility to expert systems, not because of their indolence or stupidity, but simply because they will not be able to defend the claim that they can do as well or better without the systems.

Should we mourn the passing of this species of practitioner? If we adopt the doctors' own point of view, we can see why they might well regret this development: it will make their own lives less exciting, less indispensable; they will begin to sink into the role of mere go-betweens, living interfaces between patient and system, who consolidate their direct observations into machine-readable symptomatology, and execute the therapeutic directives of the system.

It may help us to conceive of their predicament if we imagine their secret yearnings: they will occasionally be tempted to "live dangerously," to "fly by the seat of their pants," to take risks with their patients' lives just to prove to themselves that they still have the "right stuff"—that they can make bare-handed diagnoses as well as the best of the old-time doctors, the swashbuckling specialists of the 1970s and '80s. The more adventurous (or self-indulgent) of them may seek out the few exotic environments where they can practice medicine free from the obligation to use the boring technology—much the way some like to "rough it" by going camping, or by sailing small boats across the ocean. Yet, thanks to communication satellites, even Robinson Crusoe's island will provide no asylum for the physician who seeks refuge from expert systems. Being a doctor simply won't be anywhere near as much fun in the future.

This extinction of social roles is a familiar process in history. Artists, calligraphers, potters, and tailors used to be more indispensable to their communities than they are now. Although there still is a role for such artists, it is a luxury role; some people are willing to pay extra for that special, personal, artistic touch—but the realm in which the hand-made is superior to the machine-made has shrunk to an almost purely ceremonial, even mystical remnant of its former status.

Fortunately for potters, there are still enough people who prize hand-made pottery so that it is possible to sustain a career as a potter, but the social position of the potter has been ineluctably damaged; potters are simply no longer indispensable as they once were. While being a potter is still a good life compared with most others—it has more than its share of satisfactions and delights—it is not as fulfilling a life as it used to be, since any reflective potter must recognize that he or she survives by gratifying the desires of a rarified subset of the population. Doctors will not even be that lucky, for who in his right mind would acquire a taste for funky, hand-made medical care—just like Grandma used to get?

No doubt the rich and foolish would recognize a certain *cachet* in keeping a personal—and personable—physician in their entourage. Compare the doctor of the future with the apartment doorman. This descendant of the *concierge,* who had a relatively challenging and varied life work, has an almost purely ceremonial function today. You can telephone for a taxi with greater ease than your obliging doorman can lure one to the door, and the security he provides is typically almost redundant, given the twenty-four-hour surveillance and alarm system. But it looks nice to have a doorman. He adds a personal touch—of sorts. It is posh to live somewhere that is so well-heeled that it can afford to pay a grown human being to stand around in a uniform smiling all day. The doorman's life is not pleasant to contemplate; it is a travesty of human service, however well reimbursed.

Every doctor must begin to worry that he or she is heading towards becoming a health-care doorman. Can it be that, in a future generation, all that will be left of today's doctor will be minimal "computer literacy" and a bedside manner?

The advocates of expert systems in medicine may wish to intervene here, pointing out that, far from diminishing the life of the physician, expert systems will enhance it. The physician will have *more* time to deal personally with patients, and can care effectively for greater numbers, because the drudgery and galling uncertainty of poring through textbooks and journals for snatches of half-remembered wisdom will be eliminated. Indeed, and today's apartment doorman can "deal personally" with ten times as many inhabitants as the old-fashioned concierge, since all the drudgery has been removed from his life as well. The doorman has certainly been relieved of such menial labor, but also of responsibility, variety of challenge, and autonomy. Like the Cheshire cat, all that is left is the smile. As the responsibility for diagnosis and treatment shifts imperceptibly away from the physician—the "field operative"—and lodges in the expert system (or system of expert systems), doctors will suffer a similar, if less drastic, diminution of role.

I am not for one minute maintaining that today's rural doctors are heroes, and that their sad fate is the result of evil, rapacious technocrats seducing them from their noble lives. Greed and evil intentions do not enter this equation—though they are not in short supply. It is precisely because doctors want to practice the best medicine they can that they will find it incumbent on them to make these choices; for they will see that they will actually be able to save lives more reliably and efficiently by availing themselves of the technology. The interesting and risky life they had been leading will no longer be morally defensible. Wanting to be responsible and to do good, they will have to settle for a less exciting service role. We may suppose equally pure and altruistic motives on the part of those who design, develop, and promote the technology. They do not *intend* to spoil career opportunities; it is simply one of the foreseeable side effects of their effort to do a better job of saving lives through technology. What I am referring to is not a cheap melodrama with a convenient villain at which I can shake my finger, but more in the nature of a tragedy.

In a tragedy, the hero's ultimate fate must be seen to be inevitable; that is one reason why I hesitate to call this a tragedy. If I thought that this unhappy *dénouement* were strictly inevitable, I would perhaps have decided to keep the grim news to myself. How, then, might some alternative future await the physicians?

First, the technology of expert systems may turn out not to work all that well. We may discover that expert systems are so limited and unreliable, taken by themselves, that doctors will still have to be very self-reliant, very knowledgeable individually, very artful in the use they make of technology. Perhaps they will not even be obliged to use it, so untrustworthy will it prove to be. (In several conversations with advocates of such technology I have been amused to be assured, most solemnly, that I have vastly overestimated the actual powers of expert systems. These spokespeople for expert systems have failed to see the irony in their protestations: "Don't worry!" they say. "These expert systems aren't going to be *reliable!*—they won't be *foolproof!* Why, in the hands of an unskilled practitioner they would be positively dangerous!" I am strongly inclined to agree, but to suggest that to them would be to risk being dismissed as a technology-hating humanist.)

We have found one escape route: this particular technology will not work after all, and hence will not be obligatory, and hence will not spread to destroy this enviable and admirable variety of human life. There are several other ways out. If one thought that the technology *might* work, and thought that preserving the way of life of today's physician was of prime importance, one could take steps to avert this future: either by the Luddite tactic of destroying expert systems as they appeared; or by attempting to prohibit or prevent the development and improvement of the technology in the first place. But Luddism has never worked well in the past. It tends to postpone crises and aggravate situations, and is in any event not likely to inspire those who would have to support the policy today.

Alternatively, it may turn out that I have overestimated the physicians' commitment to practicing the best medicine they can. According to several observers, many doctors have given the new expert systems a lukewarm reception largely because they are more interested in "talking shop" with consultants, and in spreading liability, than in obtaining diagnostic assistance. If such resistance is widespread, it may prevent the public from perceiving the value of expert systems, and thereby keep the obligation to use them at bay.

Finally, of course, one could decide that saving the role of the *mid-twentieth-century physician* was, in the end, no more defensible than saving the role of the linotype operator in the production of newspapers. These roles must pass, perhaps, and as long as we ease the plight of the current holders of the positions, and prevent the recruitment of a new generation, little harm will be done to specific individuals. People in the future will just have other, no doubt better, occupations.

While that sentiment has a certain plausibility when the displaced workers are miners, linotype operators, or secretaries, it is far from clear what exalted work will remain for displaced physicians. If a social role as obviously valuable and impressive as that of the physician is in jeopardy, what future awaits the rest of us?[3]

Let us review the situation: if expert systems in medicine live up to their promise, then the tradition and the current trajectory of development suggest that they will probably ruin one of the most exciting and fulfilling careers in modern life. Without destroying it, they will diminish it enormously; people who want to live a good life—not just do good in life—will think twice before entering this part of the service sector. Perhaps the role of physician is not worth preserving. Alternatively, perhaps expert systems will not prove all that powerful, so that physicians will not be obliged to cede their responsibility to them. Or, in the hope that expert systems will fail to establish themselves, we might even take steps, violent or legislative, to forestall their deployment.

I see two further possibilities. The first and most probable outcome is that we shall be faced with the worst of both worlds: expert systems will not work anywhere near well enough for physicians to be *obliged* to rely on them, but the physicians will come to depend on them anyway, succumbing to the pressure of over-optimistic public opinion, their lack of self-confidence, and even laziness, greed, and fear of malpractice suits. A second somewhat utopian possibility is certainly worth striving for: perhaps we can design computer systems to support only the wily and self-reliant physician. We should look for design principles that would lead to the creation of systems that preserve or (better yet) enhance the contribution of the individual physician, while not sacrificing diagnostic power. I do not think that creating such systems is impossible, but it will not be easy; it will require rethinking the basic design task.

Compare expert systems to musical instruments: today's expert systems are similar to autoharps, designed so that anyone can learn to play them, and with an easily reached plateau of skill. We should aim instead to develop systems more like violins and pianos—instruments that indefinitely extend and challenge the powers of the individual.

I have some inklings about how this might be accomplished. They stem from ideas I have been developing at Tufts' Curricular Software Studio with my colleague, George Smith.[4] We are creating several different kinds of "concept pianos" for the exploration of complex phenomena—such as population genetics and the computer's own internal architecture. If our ideas survive their current testing, we shall subsequently present them as steps towards a new design philosophy for expert systems, but in the meanwhile there is still plenty of philosophical work to be done on these issues, to which I shall devote my remaining observations.

Why should doctors find themselves riding this obligation-train to tedium? To understand this particular phenomenon, we must step back and take a more general view of the relations between information technology and our ethical lives as decision-making agents.

Our ancestors were, relative to us, epistemically impoverished: there were few means of finding out much about non-local, non-immediate effects and problems, so they could plan and act with a clear conscience on the basis of a more limited, manageable stock of local knowledge. They were thus *capable* of living lives of virtue—of a virtue that *depended on* unavoidable ignorance. Modern technology has robbed us of the sorts of virtue that depend on such ignorance, for ignorance is all too avoidable today. Information technology has multiplied our *opportunities to know,* and our traditional ethical doctrines overwhelm us by turning these opportunities into newfound *obligations to know.*

We have always had "principles of excusable ignorance." According to tradition, we are responsible for knowing whatever is "common knowledge," plus whatever is the received wisdom of those who occupy our specialized social role—such as the role of physician—plus whatever is obviously and directly relevant to our particular circumstances of the moment. We are all responsible for knowing the standardly understood relationships between smoke and fire, rainstorms and slippery roads, voting and democracy. Plumbers—but only plumbers—have been responsible for knowing the particular effects, opportunities, and hazards of the plumbing trade, and everyone is responsible for knowing whether anyone is standing behind one's car before backing out of a parking place.

The rough-hewn boundaries of these classes of knowledge were fixed by default by the limitations of human capacity. One could not be expected to carry around vast quantities of information in one's head, nor to calculate, in the time available, any of the longer-range effects of action. The example of the physician showed in some detail how technology interacts with the obligation to know in a specialized field, but its effects on "common knowledge" are even more severe and imponderable.

"Common knowledge" is no longer the relatively stable, inertial mass it once was. We *can* acquire knowledge with little effort on almost any topic; when knowledge is "at your fingertips," how can you not be responsible for acquiring it? The obligation to know—a burden of guilt that weighs heavily on every academic, but that in milder forms is ubiquitous today—creates the situation where, if we read everything we "ought" to read, we would have time to do nothing else. Thanks to science and mass communication, we *all* now know that, in addition to worrying about whether someone is standing behind our car when we back up, we also have to wonder about the effects of our personal auto-driving (and auto-buying) activities on air pollution, acid rain, the local and global economy, and so forth.[5]

The well-known glut of information has inspired a host of responses from those who must cope with it, or wish to exploit it. Since everyone knows that no one can possibly keep abreast of all this information, meta-techniques, meta-strategies, meta-meta-structures, meta-meta-meta-tactics have arisen. The "common knowledge" we are now held responsible for is not the whole of what is almost instantaneously *available* to almost everyone, but rather a small, shifting core of what might be called "temporarily famous" common knowledge. (Recall Andy Warhol's prediction of the future time when each person will be famous for ten minutes.) Getting items of information into the spotlight of temporary fame has become a major enterprise. Whether your problem is the eradication of Third World hunger, the deposition of an evil dictator, stopping the Star Wars lunacy, or selling cornflakes, your solution must begin with "advertising"—attracting the fleeting attention of the well-intentioned, and *imposing* that item of information on them.

So much information is available that mere accessibility is no better than invisibility. Most books that are published are not read, and even being read does not guarantee their influence. This depends on higher-order effects: a book must not only be reviewed, but (thanks to the reviews) be included on an influential list of books to be read, for example. If it achieves sufficient visibility in the higher-order structures, it need not even be read to have vast influence. This profusion of information filters, duplicators, and amplifiers is the product of helter-skelter competition, and there is little reason to suppose such a process is even approximately optimizing. On the contrary, there is probably scant direct relationship between the value of items of information and their capacity to exploit the publicity environment and reproduce themselves across the society.

Richard Dawkins' excellent book, *The Selfish Gene,* introduces the idea of what he calls *memes*—a "new kind of replicator" living in "the soup of human culture." Memes are, to a first approximation, ideas in particular forms—the sort of thing one might be able to patent or copyright:

> Examples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperm or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain via a process which, in the broad sense, can be called imitation.[6]

The analogy between memes and genes runs deep, as Dawkins shows. Recasting my argument in his terms, my claim is that, thanks to *some* technological memes, we have entered a population explosion of memes—parasites that are overwhelming their hosts. Unless we can find some new ideas, some antibodies for these new antigens, hard times are in store for us. The new memes we need are *conceptual* innovations, not just new technology.

It is technology that has created this embarrassment of riches. Consider, in this respect, our obligations to those in misery on other planets. It is quite possible that there is life elsewhere in the universe, and if there is life there is almost certainly misery as well. Fortunately for us, however, the most shocking and gruesome calamities on other planets—plagues, dictatorships, nuclear holocausts—are nothing to us, because even if we knew about them (which, fortunately for our peace of mind, we do not) there would be absolutely nothing we could do about them. Perhaps we should be wary about proceeding with the project Carl Sagan champions of trying to communicate with the civilizations on other planets; after all, we might succeed and find that their first message to us was a heart-rending plea for help—

together with detailed information on just how we could be of assistance![7]

Not long ago, measured by the astronomical or even the biological time scale, the Western Hemisphere was as remote from the Eastern as any planet is from us now. Even well into the nineteenth century, few people had the knowledge and power to have any clear obligations to anyone or anything beyond their local communities. The average person could not reasonably expect to have much effect on the lives of those in distant lands, and hence was absolved from worrying about them. Such questions just did not arise—any more than the question of what to do about starvation in other solar systems arises for us.

A few people of enhanced power and knowledge found, however, that they could not hide behind their powerlessness. Their attitude was captured in the slogan *noblesse oblige*—those of noble birth had special obligations.[8] While the slogan applied originally only to the titled few, the idea was subsequently extended to all those who had power, inherited or otherwise acquired. The price they paid for their "*noblesse,*" in the extended sense of not having to devote their waking hours to providing daily bread and shelter, was an enlarged social purpose. In the nineteenth century, every well-read person could ask whether he or she should become fully committed to ending slavery, for instance, and many decided they should. Their efforts succeeded in practically eradicating slavery.

Others took on different causes, with varying degrees of success. It is often noted how curious and sometimes eccentric the focus of such people's moral sensitivities can be, both then and now. Some anti-slavery crusaders were strikingly oblivious of the suffering of their own poor, the degradation of their own servants, the exploitation of the workers in their factories. Today single-minded zealots may uphold the cause of environmentalism or animal rights, apparently unmoved by the outrages committed against their own species by various dictators—at least they devote none of their energies to action on these fronts. Similarly, there are ardent nuclear disarmers who do not care to reform the sexism out of their own language and practices, or who routinely discard unopened all mail beseeching them to enlist in the cause of Amnesty International or Oxfam.

There can be little doubt how these exotic specimens of do-gooder come into existence. We are *all* of the *noblesse* these days. We all have the daunting luxury of the time, energy, knowledge, and power to undertake a broader purpose than merely staying alive and keeping our immediate kin in the same condition. Technology has created innumerable opportunities for us to know, and to act. We want to deal responsibly with this bounty, but *we do not know how*. When we turn to the question of which priority should engage our best efforts, we drown in the available information, unable to make truly principled decisions. Our responses exhibit a sort of Rorschach magnification of whatever minor personal proclivities emerge from the noise of competing and imponderable alternatives. The results, as we have seen, may often be eccentric, but they are arguably better than the course chosen by those who sit on their hands and ignore all appeals, on the grounds that they cannot calculate—have no time to calculate—which appeal is the worthiest.

One would think that any solution there might be to this practical dilemma would stem from philosophy, and more narrowly from ethics, but much as I would like to relate that the humanists either have the answer or at least have undertaken the research program that ought to yield the answer, I must report that almost no direct attention has yet been paid to this troubling moral problem by professional philosophers.

The reason for this is not hard to find. Ethics, like any theoretical enterprise in science, has always been conducted with the aid of idealizations. Reality, in all its messy particularity, is simply too complicated to theorize about taken straight. A favorite idealization in ethics has been the useful myth of the moral agent with unlimited knowledge and time for ethical decision-making. For instance, consequentialist theories, such as the various brands of utilitarianism, declare that what ought to be done is always whatever course of action will have the best expected consequences *all things considered*. Although consequentialists know perfectly well that no one can ever truly consider all things—even all relevant things—they still choose to couch their theories in terms of what the ideally reflective and conscientious decision-maker would have made of all the available facts. This presumably gives one a standard of conduct at which to aim, if never in fact to reach. *In practice,* we tend to overlook important considerations and bias our thinking in numerous idiosyncratic ways, but *in principle* what we should do is what this ideal calculator of consequences decides will most probably maximize utility (or whatever we call the good consequences).

The plain fact that we are all finite, forgetful, and have to rush to judgment is standardly recognized, not implausibly, as a real but irrelevant element of friction in the machinery whose blueprint we are describing. It is as if there might be two disciplines—ethics proper, which undertakes the task of calculating the principles of what one ought to do under all circumstances—and then the less interesting, "merely practical" discipline of *Moral First Aid,* or *What to Do Until the Doctor of Philosophy Arrives,* which tells, in rough and ready terms, how to make decisions under time pressure.

My suspicion is that traditional theories of ethics all either *depend on* or *founder on* the very elements of friction that are ignored by the standard idealization. Information technology, by removing the friction, helps expose the weakness of much that has passed for sound in ethics. For instance, a bench test that most ethical theories pass with ease is the problem: what should you do if you are walking along, minding your own business, and you hear a cry for help from a drowning man? But almost no one faces predicaments with that logical form anymore; instead we hear, every day, while desperately trying to mind our own business, a thousand cries for help, complete with volumes of information on how we might oblige.[9] On this ubiquitous problem, traditional ethical systems are essentially reduced to silence or transparent handwaving.

This is too large a claim to support here, but I can at least sketch the problem as I currently envision it. How could we write the *Moral First Aid Manual?* Or, might we replace the manual with something fancier—an Expert System for Moral Advice-Giving in Real Time?

The fantasy of just such an expert system often lurks in the shadows of ethical theory. "If what I ought to do is whatever has the highest expected utility, how on earth shall I calculate it in the time available?" This question has been familiar for over a hundred years, and the standard response from the moral philosophers is well expressed by John Stuart Mill, who borrowed a metaphor from the technology of his own day:

> Nobody argues that the art of navigation is not founded on astronomy because sailors cannot wait to calculate the Nautical Almanac. Being rational creatures, they go to sea with it ready calculated; and all rational

creatures go out upon the sea of life with their minds made up on the common questions of right and wrong . . . .[10]

This is as fine an idea today as it was in Mill's time, but the metaphor misleadingly invites us to ignore the fact that the future position of the heavenly bodies could *actually* be calculated in advance, using the technology of the day. Where is the Moral Almanac that would guide the moral chooser through the stormy seas of life? We are still debugging it.[11] Jeremy Bentham, Mill's contemporary, set out to create a "hedonic calculus," and while no one takes it seriously today, the descendants of this quaint museum piece are still being produced, elaborated, and, above all, advertised, not just by philosophers, but by "cost-benefit analysts," computer modelers, and other futurologists.

What should be evident to computer scientists, if still easily overlooked by philosophers, is that the idea of actually producing a reliable or authoritative consequentialist almanac of any generality is sheer fantasy, now and at any future time. Compare the demanding specifications for such a system with the now well-known limitations on far simpler forecasting and problem-solving tools. *Short*-range real-time weather forecasting, for instance, has reached useful levels of reliability by restricting itself severely to a handful of measures, coarse-grained data-grids, and relatively simple equations, and then exhausting the powers of the world's fastest super-computers. Detailed forecasting of the weather months into the future is probably computationally intractable under any circumstances.[12] If it proves not to be intractable, it will be only because micro-climatic effects will be shown not to propagate chaotically after all. Yet, we already know, from countless everyday experiences, that "micro-social" effects—some unknown individual's dislike of Tylenol, for example—can create major perturbations in the best-laid human plans and social trends.

Even supposing the prediction problem could somehow be tamed, the evaluation problem would remain. In chess-playing programs, the problem of when to terminate look-ahead and evaluate the resulting position has led to the framing of the *principle of quiescence:* Always look several moves beyond any flurry of exchanges and postpone final evaluation until a relatively quiescent board position obtains. This satisfactory, though not foolproof, strategy of chess design, is systematically inapplicable to the design of our moral advice-giver, because of what we might call the Three Mile Island Effect. It has now been several relatively quiescent years since the melt-down at Three Mile Island, but can we yet say, with confidence better than a coin flip, whether that was one of the good things that have happened or one of the bad? If our imagined system were to generate a future path of probability *p* with Three Mile Island as its terminus, should it assign a high or low utility to the event? The trouble is, of course, that in life there is no checkmate, no fixed point finitely in the future at which we get one definitive result or another, from which we might calculate, by retrograde analysis, the actual values of the alternatives that lie along the paths followed and not followed. So there is no way, and *could be* no way, to tune the parameters of any prototype expert system we designed—except by the invocation, as usual, of ideology and handwaving.

The suspicion that consequentialist theories are systematically infeasible is nothing new. It has fueled support for the so-called Kantian or duty-based ethical alternative for over a century.[13] As the Pirate King says to Frederick, the self-styled "slave of duty" in *The Pirates of Penzance,* "Always follow the dictates of your conscience, me boy—and chance the consequences!" The trouble is that such duty-based theories, while not always leading to results as comical or pathetic as Frederick's myopic posings and blunderings in *The Pirates of Penzance,* have hardly coalesced into a stable and compelling system of recipes for action. Kant's own *categorical imperative,* which he quite consciously conceived as the one and only rule that needed to be printed in the *Moral First Aid Manual,* appears today about as naive and impractical a guide as Bentham's hedonic calculus.

It is a step in the right direction however, and what *is* new is the opportunity to reconceive of these alternatives to consequentialism through the lens of artificial intelligence as responses to the inescapable demands of real-time heuristic decision-making. When viewed from this perspective, for instance, what would count as a justification or defense of an ethical principle shifts significantly. This opens up a promising research program in philosophy, in my opinion, and I think it will gain more than just jargon from its engineering perspective.

The first, general result is appealing: we can already see that, since *any* "system" for ethical decision-making must be bounded arbitrarily by limitations that are far from content-neutral, no technological black-box oracle can give you a principled, objective, reliable answer to your ethical problems, no matter what anyone advertises. When the choice is between "flying by the seat of your own pants" on the one hand and paying to fly by the seat of somebody else's pants on the other, you are entitled to keep both the responsibility and the excitement to yourself.

ENDNOTES

[1] These reflections have grown out of discussions in the Norbert Wiener Forum, a policy workshop at Tufts, funded by the CSK corporation of Japan, and under the co-directorship, currently, of Professors Tadatoshi Akiba and David Isles. Earlier versions of parts of this paper were presented at the joint meeting of the Norbert Wiener Forum with its counterpart forum at Tokai University in Japan, July, 1985, and in lectures at the MIT Laboratory for Computer Science, and the Yale Humanities Center this Spring.

[2] Most recently and convincingly by John McPhee, in "Heirs of General Practice," which first appeared in *The New Yorker*, and has since been reprinted in McPhee's collection, *Table of Contents* (New York: Farrar, Straus, Giroux, 1985).

[3] "Even physicians, formerly a culture's very symbol of power, are powerless as they increasingly become mere conduits between their patients and the major drug manufacturers." Joseph Weizenbaum, *Computer Power and Human Reason* (San Francisco: Freeman, 1976), p. 259.

[4] Daniel C. Dennett, "Notes on Prosthetic Imagination," *Boston Review* 7 (3) (June, 1982), pp. 3–7; George E. Smith, "The Dangers of CAD," *Mechanical Engineering* 108 (2) (Feb. 1986), pp. 58–64.

[5] "Now I used to think that I was cool
  Runnin' around on fossil fuel
Until I saw what I was doin'
  Was drivin' down the road to ruin"
    —James Taylor, "Damn This Traffic Jam"

[6] Richard Dawkins, *The Selfish Gene* (Oxford: Oxford University Press, 1976), p. 206.

[7] There are suggestive observations on the role of technology in expanding our moral universe in Peter Singer, *The Expanding Circle* (Oxford: Oxford University Press, 1981), and in Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), part I.

[8] Duc de Lévis, (1764–1830) *Maxims, Préceptes et Reflexions,* is the first use I have uncovered so far in my casual inquiries.

[9] John Stuart Mill, in *Utilitarianism,* 1863, thought he could defend his utilitarianism thus: ". . . the occasions on which any person (except one in a thousand) has it in his power to . . . be a public benefactor . . . are but exceptional; and on these occasions alone is he called on to consider public utility; in every other case, private utility, the interest or happiness of some few persons, is all he has to attend

to." I doubt that this was an entirely convincing claim in 1863; it is transparently unrealistic today.

[10] Ibid., p. 31.

[11] Mill's idea was that the everyday maxims of morality that people had "made up their minds" about could be shown to be reliable rules of thumb that followed, somehow, from the more laborious and authoritative calculations of utilitarianism.

[12] Very short-range forecasting of local disturbances such as thunderstorms and tornados is proving extremely difficult, but is currently receiving considerable attention from NASA and the expert systems community, among others.

[13] The Kantian philosopher, Onora O'Neill, in "The Perplexities of Famine Relief," in *Matters of Life and Death, ed. Tom Regan* (New York: Random House, 1980) offers a convincing analysis of the fundamental embarrassment of utilitarianism: two competent and well-informed utilitarians, Garrett Hardin and Peter Singer, addressing the same issue (what if anything to do about famine relief), holding the same ethical theory, and having access to the same empirical information, arrive at opposing counsels: one thinks the case is compelling for dramatic forms of aid; to the other it is equally "obvious" that all such aid should be withheld. See also O'Neill's *Faces of Hunger* (Boston, MA: Allen and Unwin, 1986).