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 philosopher is John Stuart Mill’s, who borrowed a technological metaphor from 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. . . . (Mill, 1863, p. 31)

This is a fine idea today as it was in Mill’s time, but what the comparison conceals is 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’re still debugging it. 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, most of 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 fairly easy for philosophers to overlook, is that the idea of actually producing a reliable or authoritative consequentialist almanac of any generality is sheer fantasy, now and forever. 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. Reliable, long-range forecasting of the weather months into the future is probably computationally intractable under any circumstances.\(^7\) If it proves not to be intractable, it will be only because microclimatic effects will be shown not to propagate chaotically after all. But we already know, from a thousand everyday experiences, that “microsocial” effects—for example, some unknown individual’s dislike of Tylenol—wildly interfere with the best-laid human plans and social trends.

\(^7\) Very short-range forecasting of local disturbances such as thunderstorms and tornadoes is proving extremely difficult, but is currently receiving considerable attention from NASA and the expert systems community, among others.
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 meltdown 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 hand-waving.

The suspicion that consequentialist theories are systematically infeasible in this way is nothing new. It has fueled support for the so-called Kantian or duty-based ethical alternative for over a century. As the Pirate King says to Frederick, the self-styled “slave of duty” in *Pirates of Penzance*, “Always follow the dictates of your conscience, me boy—and chance the consequences!” The trouble is, of course, that such duty-based theories, while not always leading to results as comical or pathetic as Frederick’s myopic posings and blunderings in *Pirates of Penzance*, have hardly coalesced into a stable and compelling system of recipes for real action. Kant’s own *categorical imperative*, which he quite consciously conceived as the one and only rule that needed to

---

8. The Kantian philosopher, Onora O'Neill (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, 1986).
be printed in the *Moral First Aid Manual*, appears today about as naive
and impractical a guide as Bentham’s hedonic calculus.

Still, it is a step in the right direction, and what *is* new is the opportu-
nity 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 per-
spective, 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.
Bibliography


———, 1988, *Dissociative Disorders: 1988*, Dissociative Disorders Program, Department of Psychiatry, Rush University, 1720 West Polk Street, Chicago, IL 60612.


—, 1990a, “The Interpretation of Texts, People, and Other Artifacts,” Philosophy and Phenomenological Research, 50, pp. 177–194


Bibliography

Content,” at a conference on Information, Semantics and Epistemology, SOFIA, Tepoztlan, Mexico, August 7, 1988.


394 Bibliography


Mill, J. S., 1863, Utilitarianism, Indianapolis, IN: Hackett.


Selfridge, O., unpublished, *Tracking and Trailing*.


