The
KATHRYN FRASER MACKAY
Memorial Lecture Series
presents

MORAL THINKING UNDER TIME PRESSURE

Daniel C. Dennett
Center for Cognitive Studies
Tufts University

September 25, 1986

ST. LAWRENCE
UNIVERSITY

MORAL THINKING UNDER TIME PRESSURE

The Kathryn Fraser Mackay Lecture for 1986

by

Daniel C. Dennett
Center for Cognitive Studies
Tufts University

The cover illustration by Jim Benvenuto was inspired by a 12th century Byzantine ivory carving entitled "Adam."

The Kathryn Fraser Mackay Memorial Lectures in Philosophy and Religion

The Kathryn Fraser Mackay Memorial Lecture Series was established to honor the memory of Kathryn Fraser Mackay and to foster learned consideration of philosophical and religious issues, which were of vital interest to her as a student at St. Lawrence University.

Kathryn Fraser Mackay died in an aircraft accident in 1979, little more than two years after her graduation from St. Lawrence University, Bachelor of Arts *cum laude* with Honors in Philosophy. She was a young woman with a great gift for friendship—companionable, affectionate and generous. Bright and full of courage, Kathy waged an intellectual and spiritual battle to resolve for herself fundamental questions of philosophy and religion.

I. Mill's Nautical Metaphor

A hundred and twenty-five years ago, John Stuart Mill felt called upon to respond to an annoying challenge to his *Utilitarianism*: ". . . defenders of utility often find themselves called upon to reply to such objections as this—that there is not time, previous to action, for calculating and weighing the effects of any line of conduct on the general happiness." His reaction was quite fierce:

> Men really ought to leave off talking a kind of nonsense on this subject, which they would neither talk nor listen to on other matters of practical concernment. 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, as well as on many of the far more difficult questions of wise and foolish. And this, as long as foresight is a human quality, it is to be presumed they will continue to do. (*Utilitarianism*, 1861; p. 31.)

This haughty retort has found favor with many—perhaps most—ethical theorists, but in fact it papers over a crack that has been gradually widening under an onslaught of critical attention. The naive objector was under the curious misapprehension that a system of ethical thinking *was supposed to work*, and noted that Mill's system was highly impractical—at best. This is no objection, Mill insists; utilitarianism is supposed to be practical, but not *that* practical. Its true role is as a background justifier of the foreground habits of thought of real moral reasoners. This background role has proven, however, to be ill-defined and unstable. Just how practical is a system of ethical thinking *supposed* to be? Tacit differences of opinion about this have added to the inconclusiveness of the subsequent debate.

I want to try for a fresh perspective on these issues by considering the actual constraints and demands of real-time moral thinking, and the *possible* contributions of ethical theories, given those constraints and demands, to everybody's moral problem: "What should I do now?" The traditional simplifying strategy of ethical theorists has been to address the question of what the "perfectly rational" moral agent, with unlimited deliberation time and an unlimited ability to exploit the (possibly limited) information available, would decide was the right thing to do, *all things considered*. This

idealizing tradition has come under a barrage of criticism in recent years, and, valuable as I have found much of this discussion, I do not think it has gone far enough in its introduction of messy realities into the standard idealizations. I think philosophers have still not taken seriously enough the complaint that the ethical systems they describe, even with their new accommodations, *don't really work.*

Perhaps the grounds for my suspicion can best be brought out by reflecting on what is implicit in Mill's use of a metaphor drawn from the technology of his own day. The *Nautical Almanac* is a book of tables, calculated and published annually, from which one can easily and swiftly derive the exact position in the skies of the Sun, the Moon, the planets and the major stars for *each second* of the forthcoming year. The precision and certainty of this annual generator of expectations was, and still is, an inspiring instance of the powers of human foresight, properly disciplined by a scientific system (and directed upon a sufficiently orderly topic). Armed with the fruits of such a system of thought, the rational sailor can indeed venture forth confident of his ability to make properly informed real-time decisions about navigation. The practical methods devised by the astronomers actually work.

Do the utilitarians have a similar product to offer to the general public? Mill seems at first to be saying so. Today we are inured to the inflated claims made on behalf of dozens of high-tech systems—of cost-benefit analysis, computer based expert systems, etc.—and from today's perspective we might suppose Mill to be engaging in an inspired bit of advertising: suggesting that utilitarianism can provide the moral agent with a foolproof Decision-Making Aid. ("We have done the difficult calculations for you! All you need do is just fill in the blanks in the simple formulae provided.")

Jeremy Bentham, the founder of utilitarianism, certainly aspired to just such a "felicific calculus," complete with mnemonic jingles, like the systems of practical celestial navigation that every sea captain memorized:

> *Intense, long, certain, speedy, fruitful, pure—*
> Such marks in *pleasures* and in *pains* endure.
> Such pleasures seek if *private* be thy end:
> If it be *public*, wide let them *extend.*
> —Jeremy Bentham, *Introduction to the Principles of Morals and Legislation*, Oxford, 1789, ch. IV.

This myth of practicality has been part of the rhetoric of utilitariansim from the beginning, but in Mill we see already the beginning of the retreat

up the ivory tower to ideality, to what is calculable "in principle" but not in practice.

Mill's idea, for instance, was that the best of the homilies and rules of thumb of everyday morality—the formulae people *actually considered* in the hectic course of their deliberations—had received (or would receive in principle) official endorsement from the full, laborious, systematic utilitarian method. The faith placed in these formulae by the average rational agent, based as it was on many lifetimes of experience accumulated in cultural memory, could be justified ("in principle") by being formally derived from the theory. But no such derivation has ever been achieved.

Not only have utilitarians never made an actual practice of determining their specific moral choices by calculating the expected utilities of (all) the alternatives (there not being time, as our original objector noted), but they have never achieved stable "off-line" *derivations* of partial results—"landmarks and direction posts," as Mill puts it—to be exploited on the fly by those who must cope with "matters of practical concernment."

What, then, of the utilitarians' chief rivals, the various sorts of Kantians? Their rhetoric has likewise paid tribute to practicality—largely via their indictments of the impracticality of the utilitarians.[1]

What, though, do the Kantians put in the place of the unworkable consequentialist calculations? Kantian decision-making typically reveals that a rather different idealization—a departure from reality in another direction—is doing all the work. For instance, unless some *deus ex machina* is handy to whisper in one's ear, it is far from clear just how one is to figure out how to limit the scope of the "maxims" of one's contemplated

---

1. The charge of practical imponderability is brought against utilitariansim with particular vigor and clarity by the Kantian Onoroa O'Neill in "The Perplexities of Famine Relief," in *Matters of Life and Death*, ed Tom Regan, (New York: Random House, 1980) and in her *Faces of Hunger*, (Boston: Allen and Unwin, 1986). An independent critic is Bernard Williams, who claims that utilitariansim makes

> enormous demands on supposed empirical information, about peoples' preferences, and that information is not only largely unavailable, but shrouded in conceptual difficulty; but that is seen in the light of a technical or practical difficulty, and utilitariansim appeals to a frame of mind in which technical difficulty, even insuperable technical difficulty, is preferable to moral unclarity, no doubt because it is less alarming. (That frame of mind is in fact deeply foolish . . .) (*Utilitarianism For and Against*, p. 137)

actions before putting them to the litmus test of the Categorical Imperative. There seems to be an inexhaustible supply of candidate maxims.

Certainly the quaint Benthamite hope of a fill-in-the-blanks decision procedure for ethical problems is as foreign to the spirit of modern Kantians as it is to sophisticated utilitarians. All philosophers can agree, it seems, that real moral thinking takes insight and imagination, and is not to be achieved by any mindless application of formulae. As Mill himself puts it, still in high dudgeon, "There is no difficulty in proving any ethical standard whatever to work ill if we suppose universal idiocy conjoined with it." So after all no ethical theory or system is designed to "tell you what to do."

This is not meant to be a shocking indictment, but just a reminder of something quite obvious: no remotely compelling system of ethics has ever been made *computationally tractable* for real world moral problems. So even though there has been no dearth of utilitarian (and Kantian, and contractarian, etc.) *arguments* in favor of particular policies, institutions, practices and acts, these have all been heavily hedged with *ceteris paribus* clauses and plausibility claims designed to overcome the combinatorial explosion of calculation that threatens if one actually attempts—as theory says one must—to *consider all things*.

It will help us appreciate this obvious fact about ethical theories if we compare them, not to the productions of the Astronomer Royal, as Mill did, but to a more contemporary technique of expectation-generation: computer-aided weather forecasting.

The current North American data-gathering grid divides the atmosphere into cells approximately thirty miles on a side and ten thousand feet in height. This yields in the neighborhood of 100,000 cells, each characterized by less than a dozen intensities: temperature, barometric pressure, wind direction and velocity, etc. How these intensities change as a function of the intensities in the neighboring cells is fairly well understood, but computing these changes in temporal increments small enough to keep some significance in the answers challenges today's largest supercomputers. Obviously, a weather prediction must be both accurate and timely; achieving accuracy at the cost of taking 36 hours to calculate a 24-hour prediction is no solution.

It is not clear yet whether reliable long-range weather forecasting is possible, since the weather may prove to be too chaotic to permit *any* feasible computation. The behavior of the weather is strikingly unlike the behavior of the heavenly bodies. Suppose though, for the sake of illustration, that there were a *proven* forecasting algorithm—one that could successfully "predict" tomorrow's weather if allowed to engage in a *month* of number crunching on a bank of supercomputers. This would be scientifically very interesting, but not very useful. We can imagine taking the tour of the weather bureau and being shown the gleaming giants at their work. "How do you actually *use* the algorithm in figuring out the forecasts you are obliged to issue every day?" we ask. "Oh, we don't use the algorithm at all. We sort of eyeball the maps and the local conditions and then apply our favorite maxims. Jones is partial to 'red sky at night, sailor's delight' while I am more into aching joints and looking for the groundhog's shadow. We vote, in the end, and our track record is pretty good."

That is the way it is with ethics too—only with ethics, things are worse. At least with meteorology, there is an uncontroversial and widely accepted ideal background theory—however infeasible it might be in practical calculations. Now we can see that there are actually three ways in which Mill's metaphor is misleading. First, as just mentioned, no ethical theory enjoys the near-universal acceptance of astronomy or meteorology, in spite of vigorous campaigns by the partisans. Second, there are no feasible algorithms or decision procedures for ethics, as there are for celestial navigation. Third, the informal rules of thumb people actually use have never been actually derived from a background theory, but only guessed at, in an impressionistic derivation rather like that of our imagined meteorologists.

I am not saying that ethical theories are not valuable at all, but just that they are not valuable—and not valued, in fact—as systems for actually calculating courses of action. They are sometimes valued as calculation aids in restricted contexts, e.g. in a cost-benefit analysis that considers only narrow—typically financial—costs and benefits for a specified individual or organization over a relatively short span, but no one trusts their moral decisions to them.

If there is a *Moral Almanac* actually in use, then, it is less like the *Nautical Almanac* than it is like *The Old Farmer's Almanac*—an unsystematic collection of wise sayings, informal precepts, traditional policies, snatches of taboo, and the like, a *vademecum* vaguely approved of by the experts—who, after all, rely on it themselves—but lacking credentials.

For the most part, philosophers have been content to ignore the practical problems of real-time decision-making, regarding the brute fact that we are all finite and forgetful, and have to rush to judgment, as a real but irrelevant element of friction in the machinery whose blueprint they are describing. It is as if there might be two disciplines—ethics proper, which undertakes the task of calculating the principles of what the ideal agent

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 "on line" decisions under time pressure.

In practice, philosophers acknowledge, we overlook important considerations that we really shouldn't overlook and bias our thinking in a hundred idiosyncratic and morally indefensible ways, but *in principle*, what we ought to do is what the ideal theory says we ought to do, and philosophers have concentrated on spelling out what that ideal theory is. No philosopher has tried to write *The Moral First Aid Manual*.

This could perhaps have been a fruitful division of labor, but the lack of stable progress on the ideal theory suggests that if we paid closer attention to the constraints on real applications we might get a better sense of the issues. Might it be that features of our actual practice have a tacit wisdom, a practical rationale that can shed light on the perennial debates, such as those between consequentialists and Kantians? Every ethical theory must at some point confront the tribunal of *our moral intuitions*, that shifty mob of uncredentialed but heartfelt judgments that condemn or condone—without saying why—the deliverances of theory. Every ethical theory honors some of these intuitions while seeking to overthrow the rest. The proper habitat of these intuitions is in the *Moral First Aid Manual*, where they are relied on daily. Might our allegiance to them sometimes be more a matter of "practical wisdom" than of "theoretical insight?" Perhaps trying to write the *Moral First Aid Manual* will prove to be a theoretically interesting task after all.

I should say at the outset that the job I envision, if done right, would involve *systematic* empirical studies and experiments by psychologists in addition to my informal and anecdotal explorations, and *formal* analyses of the task domains and the useful heuristics for them (of the sort produced by people in artificial intelligence), in addition to my intuitive guesswork. I am not ready to do this work, but—as philosophers are wont—I am ready to talk about why it would be interesting work for somebody to do.

II. Judging the Competition

Let us consider a moral problem which, while a few of its details are exotic, exemplifies a familiar structure. Your department has been chosen to administer a munificent bequest: a twelve-year fellowship to be awarded in open competition to the most promising graduate student in philosophy. You duly announce the award and its conditions in the *Journal of Philosophy*, and then to your dismay you receive, by the deadline, 250,000 legal entries, complete with lengthy dossiers, samples of written work and testimonials. A quick calculation convinces you that living up to your obligation to evaluate all the material of all the candidates by the deadline for announcing the award would not only prevent the department from performing its primary teaching mission, but also, given the costs of administration and hiring additional qualified evaluators, bankrupt the award fund itself, so that all the labor of evaluation would be wasted; no one would gain.

What to do? If only you had anticipated the demand, you could have imposed tighter eligibility conditions, but it is too late for that: every one of the 250,000 candidates has, we will suppose, a right to equal consideration, and in agreeing to administer the competition you have undertaken the obligation to select the best candidate.[2]

When I have put this problem to colleagues, I find that after a brief exploratory period, they tend to home in on one version or another of a mixed strategy, such as:

> choose a small number of *easily checked* and *not entirely unsymptomatic* criteria of excellence—such as Grade Point Average, number of philosophy courses completed, weight of the dossier (eliminating the too-light and the too-heavy)—and use this to make a first cut. Conduct a lottery with the remaining candidates, cutting the pool down randomly to some manageably small number of finalists—say 50 or 100—whose dossiers will be carefully screened by a committee, which will then vote on the winner.

There is no doubt that the policy I have described is very unlikely to find the best candidate. Odds are, in fact, that more than a few of the losers, if given a day in court, could convince a jury that they were *obviously* superior to the elected winner. But, you might want to retort, that's just tough; you did the best you could. It is quite possible, of course, that you would lose the lawsuit, but you might still feel, rightly, that you could have arrived at no better decisions at the time.

My example is meant to illustrate, enlarged and in slow motion, the ubiquitous features of real-time decision-making. First, there is the simple

2. I don't mean to beg any questions with this formulation in terms of rights and obligations. If it makes a difference to you, recast the setting of the problem in terms of the overall disutility of violating the conditions set forth in your announcement of the competition. My point is that you would find yourself in a bind, whatever your ethical persuasion.

physical impossibility of "considering all things" in the allotted time. Note that "all things" doesn't have to mean *everything* or even *everybody in the world*, but just "everything in 250,000 readily available dossiers." You have all the information you need "at your fingertips"; there need be no talk of conducting further investigations. Second, there is the ruthless and peremptory use of some distinctly second-rate cut rules. No one thinks *Grade Point Average* is a remotely foolproof indicator of promise, though it is probably somewhat superior to *weight of dossier*, and clearly superior to *number of letters in surname*. There is something of a trade-off between ease of application and reliability, and if no one can think of any easily applied criteria that one can have *some* faith in, it would be better to eliminate the first cut step and proceed straight to the lottery for all candidates. Third, the lottery illustrates a partial abdication of control, giving up on a part of the task and letting something else—nature or chance—take over for a while, while still assuming responsibility for the result. (That is the scary part.) Fourth, there is the phase where you try to salvage something presentable from the output of that wild process; having *over*-simplified your task, you count on a metaprocess of self-monitoring to correct or renormalize or improve your final product to some degree.[3] Fifth, there is the endless vulnerability to second-guessing and hindsight wisdom about what you should have done—but done is done.

Notice how the pattern repeats itself, rather like a fractal curve, as we trace down through the sub-decisions, the sub-sub-decisions, and so forth until the process becomes invisible. At the department meeting called to consider how to deal with this dilemma, (1) everyone is bursting with suggestions—more than can be sensibly discussed in the two hours allotted, so (2) the chairman becomes somewhat peremptory, deciding not to recognize several members who might well, of course, have some very good ideas, and then (3) after a brief free-for-all "discussion" in which—for all anyone can tell—timing, volume and timbre may count for more than content, (4) the chairman attempts to summarize by picking a few highlights that somehow strike him as the operative points, and the strengths and weaknesses of these are debated in a rather more orderly way, and then a vote is taken. After the meeting, (5) there are those who still think

3. See my "A Route to Intelligence: Oversimplify and Selfmonitor," (CCM-85-4, Center for Cognitive Studies, Tufts) forthcoming in J. Khalfa, ed., *Can Intelligence be Explained?*, Oxford University Press; and "Designing Intelligence," (CCM-86-4, Center for Cognitive Studies, Tufts), British Association for Advancement of Science, September 2, 1986.

that better cut rules could have been chosen, that the department could have afforded the time to evaluate 200 finalists (or only 20), etc., but done is done. They have learned the important lesson of how to live with the sub-optimal decision-making of their colleagues, so after a few minutes or hours of luxuriating in clever hindsight, they drop it.

"But *should I* drop it?" you ask yourself, just as you asked yourself the same question in the midst of the free-for-all when the chairman wouldn't call on you. (1) Your head was teeming at that moment with reasons why you should go along with your colleagues quietly, and all this was competing with your attempts to follow what others were saying, and so forth—more information at your fingertips than you could handle, so (2) you swiftly, arbitrarily and unthinkingly blocked off some of it—running the risk of ignoring the most important considerations, and then (3) you gave up trying to *control* your thoughts; you relinquished meta-control and let your thoughts lead wherever they might for a while. After a bit you somehow (4) resumed control, attempted some ordering and improving of the materials spewed up by the free-for-all, and made the decision to drop it, suffering (5) instant pangs of dubiety and toying with regret, but, because you are wise, you shrugged these off as well.

And how, precisely, did you go about dismissing that evanescent and unarticulated micro-wonder ("should I have dropped it?")? Here the processes become invisible to the naked eye of introspection, but if we look at cognitive science models of "decision-making" and "problem-solving" *within* such swift, unconscious processes as perception and language comprehension, we see further tempting analogues of our phases in the various models of heuristic search and problem-solving.[4]

So far as I can see, then, time-pressured decision-making is like that *all the way down*. There may be dividing lines to be drawn between biological, psychological and cultural manifestations of this structure, but not only are the structures—and their powers and vulnerabilities—basically the same; the particular contents of "deliberation" are probably not locked into any one level in the overall process but can migrate. Under suitable provocation, for instance, one can dredge up some virtually subliminal consideration and elevate it for self-conscious formulation and appreciation—it becomes an "intuition"—and then express it so that others can

4. The suggestion of temporal ordering in the five phases is not essential, of course. The arbitrary pruning of randomly explored search trees, the triggering of decision by a partial evaluation of results, and the suppression of second-guessing need not follow the sequence in time I outline in the initial example.

consider it as well. Moving in the other direction, a reason for action perennially mentioned and debated in committee can eventually "go without saying"—at least out loud—but continue to shape the thinking, both of the group and the individuals, from some more subliminal base (or bases) of operations in the process.

III. Toward Designing the Manual

I think this is the *basic* structure of all real decision-making, moral, prudential, economic or whatever. To say that this structure is basic is not necessarily to say that it is best, but that conclusion is certainly invited—and inviting. We began, remember, by looking at the highest level instantiation of the structure, the candidate evaluation process, and we treated our task as a design problem. Suppose we decide that the system we designed is about as good as it could be, given the constraints. A group of roughly rational agents—us—decides that this is the right way to design the process, and we have reasons for choosing the features we did. The same reasons (good or bad) apply to any level of the structure we encounter.

Given this genealogy for one version of the design, we might muster the *chutzpah* to declare that this is optimal design. Optimality claims have a way of evaporating, however; it takes no *chutzpah* at all to make the modest admission that this was the best solution *we* could come up with, given our limitations.[5]

This same slipperiness can be observed in the debates in ethics about rationality, over such questions as whether it could be rational for an agent to choose to be irrational on some occasions (e.g., the recent discussions by Parfit, Gauthier, Pettit) and the more long-standing debates among "act" utilitarians and "rule" utilitarians over whether one can "justify" reliance on "sub-optimal" rules. The mistake that is sometimes made is supposing that there is a single (best or highest) perspective from which to assess ideal rationality. Does the ideally rational agent have the all-too-human problem of not being able to remember certain crucial considerations when they would be most telling, most effective in resolving a quandary? If we stipulate, as a theoretical simplification, that our imagined ideal agent is immune to such disorders, then we don't get to ask the question of what the ideal way might be to cope with them.

5. Compare that with the claim: "Mother Nature isn't perfect, but she does the best she can." Is that a Panglossian statement or not? See the discussion of this issue in my "Intentional Systems in Cognitive Ethology: the 'Panglossian Paradigm' Defended," *Behavioral and Brain Sciences*, 1983, pp. 343–90.

The *Moral First Aid Manual* should not be considered a grubby compromise with practicality, but itself just as pure an ideal vision as any other in ethics: the book the ideally rational agent would write as his own *vademecum*, written in the light of his perfect self-knowledge about his limitations. Any such exercise presupposes that certain features—the "limitations"—are fixed, and others are malleable; the latter are to be adjusted so as best to accommodate the former. But one can always change the perspective and ask about one of the presumably fixed features whether it is something one would want to tamper with in any event; perhaps it is for the best as it is. Addressing that question requires one to fix still further ulterior features as fixed, in order to assess the wisdom of the feature under review.

So without pausing further to ask whether it is good, or even inevitable, that we human beings rush to judgment in the hectic way we do, let us just suppose that this is the non-optional background against which we should examine systems of ethics. That is, as rational agents able to some degree to change our ways, we can consider the pros and cons of *adopting* or *converting to* a particular ethical system, but taking such a step would not mean *abandoning* the way of thinking I have just described in favor of some other way of thinking. Our conversion, no matter how heartfelt and sincere, would amount to the provisional incorporation of the ethical system in question into the bag of tricks we rely on when decision-making under time pressure. We may give it "pride of place" among our tricks, but that is the pinnacle of its authority. The familiar face of this feature is our recognition that we "wait to see how it comes out" before endorsing and acting upon the deliverance of an ethical argument. We never cede irrevocable control to the system. (We will reconsider the "wisdom" of this shortly.)

If *The Moral First Aid Manual* is to be optimally addressed to a time-pressured decision-maker, it may help us design it if we slow the process down once more and look at what makes for good decision-making at the departmental level. First, of course, you want to have good colleagues: people who can be relied upon to come up with the right sorts of considerations right away, without wasting precious time on irrelevancies. (This translates readily into the discussion among utilitarians, familiar since Mill's day, of the value of inculcating good *habits of thought*.) There is no point having more than one colleague if they are clones of each other, all wanting to raise the same consideration, so we may suppose them to be specialists, each somewhat narrow-minded and preoccupied with protecting a certain set of interests.

Now how shall we avert a cacophony of colleagues? We need some *conversation-stoppers*. We need some ploys that will arbitrarily terminate reflections and disquisitions by our colleagues and cut off debate independently of the specific content of current debate. Why not just a *magic word*? Magic words work fine as control-shifters in artificial intelligence programs, but we're talking about controlling intelligent colleagues here, and they are not apt to be susceptible to magic words. That is, good colleagues will be reflective and rational, and open-minded within the limits imposed by their specialist narrow-mindedness. (They could take their motto from the philosophical journal, *Nous: "Nihil philosophicum a nobis alienum putamus."*) They need to be hit with something that will appeal to their rationality, while discouraging further reflection.

It will not do at all for these people to be *endlessly* philosophizing, endlessly calling us back to first principles and demanding a justification for these apparently (and actually) quite arbitrary principles. What could possibly protect an arbitrary and somewhat second-rate conversation-stopper from such relentless scrutiny? A meta-policy that forbids discussion and reconsideration of the conversation-stoppers? But, our colleagues would want to ask, is *that* a wise policy? Can it be justified? It will not always yield the best results, surely and . . . and so forth. One cannot expect there to be a single stable solution to this design problem, but rather a variety of uncertain and temporary equilibria, with the conversation-stoppers tending to accrete pearly layers of supporting dogma which themselves cannot withstand scrutiny, but do actually serve on occasion, blessedly, to deflect and terminate consideration.

Here are some promising examples:

"But that would be to break a promise."
"But that would be to use someone merely as a means."
"But that would violate a person's *right*."

Bentham once rudely dismissed the "theory" of absolute rights as "nonsense upon stilts" and we might now reply that perhaps he was right; perhaps talk of rights *is* nonsense upon stilts, but *good* nonsense—and good only because it is on stilts, only because it happens to have the "political" power to keep rising up above the meta-reflections and reasserting itself as a compelling—that is, conversation-stopping—"first principle."

In short, "rule worship" of a certain kind is a good thing, at least for agents designed like us. It is good not because a certain rule, or set of rules, is probably the best, or always yields the right answer, but because

having rules works—somewhat—and not having rules doesn't work at all.

If we agree that ethical decision-making cannot be an *effective procedure* in the technical sense (an algorithm guaranteed to give us the solution we are seeking, the "right" solution) we still want it to be an effective procedure in the everyday sense: a procedure that actually terminates with a good chance of leaving us with a presentable, acceptable, better-than-nothing solution. The uneasy equilibrium of any such design can be understood if we return to the question of whether we would want any ethical system to hold more than "pride of place" among our bag of tricks. What if someone becomes so impressed with the yield of some tool, some new colleague in the department, that pride of place turns into a total monopolization of the discussion? The result would be fanaticism.

The risks of lapsing into fanaticism, of having one's hectic and informal heuristic operating system completely occupied and controverted by some new routine, are apparently considerable, if we look around at our fellows. Probably the chief "virtue" of fanaticisms of all stripes—religious, nationalistic, political and indeed philosophical—their holding power, their parasitical strength in competition with rival ways of thinking, is their capacity to alleviate the discomfort of uncertainty. When one is overwhelmed with knowledge—of candidates for a fellowship, or of wrongs to be righted—and gets desperate for a reason to move one way rather than another (recalling the plight of Buridan's ass), almost any Oracle that will reliably and swiftly give an answer will loom in attractiveness and plausibility—simply because it gives blessed relief from indecision.

If we think about the predicament of a moral agent as a constant, time-pressured competition of allegiance to rules or principles, perhaps we can make more sense of some of the phenomena traditional ethical theories wave their hands about. For one thing, we might begin to understand our current moral position—by that I mean yours and mine, at this very moment. Here we are, devoting an hour to my meta-meta-meta-reflection on values and valuation. Is this time well spent? Shouldn't we all be out raising money for Oxfam or picketing the Pentagon or writing letters to our Senators and Representatives about various matters? Did you consciously decide, on the basis of calculations, that the time was ripe for a little sabbatical from real world engagement, a period "off line" for maintenance and inventory control? Or was your process of decision—if that is not too highfalutin' a name for it—much more a matter of your not tampering with some current "default" principles that virtually ensure that you will ignore all but the most galvanizing potential interruptions to your rather narrow, personal lives?

The problems of justifying "personal projects" has been quite properly a focus of attention in recent ethics. Consider a traditional bench test that most systems of ethics can pass with aplomb: solving the problem of what you should do if you are walking along, minding your own business, and you hear a cry for help from a drowning man. That is the easy problem. The hard problem is, how can we justifiably get ourselves into that relatively happy predicament? Our prior problem, it seems, is that every day, while trying desperately to mind our own business, we hear a thousand cries for help, complete with volumes of information on how we might oblige. How on earth could anyone prioritize that cacophony? Not by any systematic process of considering all things, weighing expected utilities, and attempting to maximize. Nor by any systematic generation and testing of Kantian maxims—there are too many to consider. The actual process must be much more like our helter-skelter heuristics, with arbitrary and unexamined conversation-stoppers bearing most of the weight.

That arena of competition encourages escalations, of course. With our strictly limited capacity for attention, the problem faced by others who want us to consider their favorite consideration is essentially a problem of advertising—of attracting the attention of the well-intentioned. This is the same problem whether we view it in the wide-scale arena of politics, or in the close-up arena of personal deliberation.

For better or for worse, your attention got attracted to my considerations for more than my share of time. I am grateful for it, and hope it proves to have been time well spent.

About the Author

Daniel C. Dennett received his B.A. from Harvard University and Ph.D. from Oxford University. Professor Dennett taught from 1965 to 1971 at the University of California at Irvine. Since 1971 he has been in the philosophy department at Tufts University, where he has been the director of the Center for Cognitive Studies since 1985. He has held numerous visiting lectureships (e.g. Yale, Dartmouth, MIT) and is author of many important works in the field, including *Content and Consciousness* (1969), *Brainstorms: Philosophical Essays on Mind Psychology* (1978), *Elbow Room: The Varieties of Free Will Worth Wanting* (1984), and *The Intentional Stance* (1987). Professor Dennett's latest book is *Consciousness Explained*, a 1991 publication. We welcome Professor Dennett to the Mackay Lecture series.

One of 200 copies printed by Ryan Press, Ogdensburg, New York,
and bound by The Soleil Bookbindery, Rochester, New York.
Acid-free Mohawk paper was used for both text and cover.
The type, Times Roman, was set by
Partners Composition,
Utica, New York.