

ARTICLES

**Saving Lives, Saving from Death, Saving from Dying:
Reflections on ‘Over-Valuing’ Identifiable Victims**

Mark Kelman*

INTRODUCTION52

**I. DO SUBJECTS MEET THEIR GOALS INEFFICACIOUSLY OR DO CRITICS
MISCOMPREHEND THEIR GOALS?61**

 A. CONCEPTUALIZING DISTINCT HESITATIONS ABOUT CLAIMS OF
 COGNITIVE ERROR61

 B. DEATH AND DYING66

 1. THE POSSIBILITY OF COGNITIVE BIAS68

 2. THE POSSIBILITY OF UNACKNOWLEDGED GOALS: DEATH, DYING
 AND DREAD74

II. CONSEQUENTIALISM AND AGENT-RELATIVITY (A): CONDUCT V. CAUSE ...88

III. CONSEQUENTIALISM AND AGENT-RELATIVITY (B): PARTIALITY.....95

CONCLUSION98

* James C. Gaither Professor of Law and Vice Dean, Stanford Law School. While errors are mine, I remain grateful for the help I received from my research assistant, Elena Coyle, and for the feedback I received from participants at workshops at UCLA and Stanford Law Schools. Jared Curhan, Dick Craswell, Barbara Fried, and Seana Shiffrin were especially insightful and generous with their time and attention.

INTRODUCTION

The canonical case that psychologists, philosophers, and policy analysts reflect upon in considering how and why individuals and collective decision makers allocate resources that diminish the number of preventable death in seemingly irrational ways is the “Baby Jessica” case. The simple behavioral observation we make is that there was a generous, spontaneous outpouring of aid to save Baby Jessica, a young child trapped in a well.¹ At the same time, those people who sent checks and cash to save the trapped child seem to be willing to expend far fewer resources to *prevent* such accidents or other fatalities. The case raises at least two quite distinct descriptive and normative issues that are often conflated, though each is worth independent attention: First, how do and should we think about the ways we react to saving identifiable victims rather than “unknown persons”?² Second, how do and should we make decisions about

1. Actually, little, if any, of the approximately \$700,000 in unsolicited funds that were sent to the McClure family back in 1987 while Baby Jessica was trapped for fifty-eight hours after falling into a well in Midland, Texas, went to defraying the costs of the rescue—or to pumping additional resources into an otherwise-underfunded rescue effort. Instead, the funds were put into a trust both to pay to treat Jessica’s injuries and for her education and long-term support. See Susan Schindehette & Anne Maier, *The Joy of Life Hard Won*, PEOPLE, Nov. 7, 1988, at 146, 151, available at <http://www.people.com/people/archive/article/0,,20100404,00.html>. The trust, in fact, was never really used even for those purposes, even though Jessica did sustain some quite serious injuries: Jessica (at age twenty-one) reported that the trust, which will be distributed to her at twenty-five, will go to her own child, Simon. See Mike Celizic, ‘Baby Jessica’ 20 Years Later, TODAYshow.com, June 11, 2007, <http://today.msnbc.com/id/19165433/?GTI=10056>.

2. Over the course of the paper, it will become clear that the term “identifiable victim” is by no means readily defined; certain explanations of *why* people apparently care more (or, to put it more normatively, unduly) about identifiable victims depend on distinct accounts of what the attitude-affecting traits of such victims might be. For now, I believe it will be helpful in considering these issues to emphasize a few distinctions that have not been made in precisely this fashion in the literature. A victim is *strongly* “identifiable” if she is a known, named person, at least modestly vividly described in terms of her particular attributes—she has at least a thin bio. She is a strong *victim* if we know that there are discrete steps we can take now that will save her from a particular short-term threat. Victims are *weakly* identifiable if no one is identified as a particular person but we know that *some* person or people could already be labeled as in need of action that will prevent their deaths. However, the distinction between weakly identifiable persons and “mere” statistical lives seems extraordinarily psychologically flimsy. While it is reasonably clear that 1) releasing a toxin that will increase mortality rates of a large group ordinarily merely threatens “statistical lives,” and 2) releasing a toxin that will imminently kill a particular family, vividly pictured, threatens “identifiable lives,” there are many intermediate and blurry cases. Imagine the toxin will kill all members of whatever geographically confined community it happens to “land on,” but we do not yet know where that community will be. The victims are plainly not strongly identifiable, but it is not obvious whether they are weakly identifiable in attitude-relevant ways or not; the answer to that question depends on whether one believes that a significant part of what drives the

expending resources to *cure* or *rescue* those who are already known to be in mortal danger rather than to *prevent* people from dying or developing fatal maladies?³

effect is that people care more about a large proportion of a group, however defined, dying than they care about a smaller proportion of a large group dying. Moreover, if “identifiability” is in whole or in part a species of “salience” or vividness, it may well be that seemingly unambiguously statistical deaths strongly associated with some past set of focal, salient dying persons will be treated more like deaths of identifiable victims: I suspect, for instance, that most costly steps designed to reduce the risk of death-from-terrorism are in significant ways thought by many subjects to redound not to *mere* “unidentifiable victims” or statistical lives but to the projected-to-the-future and therefore psychologically vivid lives of iconic 9/11 victims.

It might matter, too, even if we are in the *ex ante* decision-making position (Can we release the toxin? Must we take expensive steps to avoid its release?), whether we believe that we *will* know, *ex post*, that some particular deaths were attributable to the toxin. Victims will obviously be more “identifiable” *ex post* if the autopsy clarifies what caused their death than if, even *ex post*, we must rely on epidemiological evidence to tell us that *some* of the deaths among a large number would almost surely not have occurred had we acted differently. Note, then, that yet another reason preventive measures to avert death-by-terrorist may involve saving semi-identifiable persons is that we know, *ex ante*, that if some do die, it will be known, *ex post*, to be a result of terrorism, though it will not so clearly be a result of terrorism that could have been averted by the steps we might have chosen to forego. But, once again, whether the foreknowledge that this will be the case makes us evaluate the initial decision to a greater extent as killing the identifiable victims depends in part on how one *explains* the psychological processes behind the so-called “identifiable victim” effect. See Karen E. Jenni & George Loewenstein, *Explaining the ‘Identifiable Victim Effect,’* 14 J. RISK & UNCERTAINTY 235 (1997). I am skeptical that we have especially good explanations of the identifiable victim effect. Thus, I am not at all certain how to classify all victims.

I point out that it is also not easy to tell whether a person, identifiable or not, is already a victim or is merely in jeopardy of becoming one. See *infra* note 3.

3. “Cure” and “prevention” are hardly clear, binary categories either, but I offer some thoughts about how they might be profitably distinguished. At core, we are most clearly engaged in “cure” or “rescue” when several conditions are simultaneously met. First, we must be “substantially certain” that the party at risk could die as a direct result of an injury already sustained or a disease already developed. Second, we must believe that the measures we are considering are intended to make the party’s death substantially less likely. We are engaged in prevention, in my view, when the person is not yet injured or sick with the disease that would be listed as the cause of death, and when it is merely risky that he will be injured or develop the dangerous disease that poses a risk of death. Why are the categories plainly not analytically binary? First, there is no bright line distinction between an event being “substantially certain” and “highly probable.” Second, the notion that we are preventing rather than curing until the party has developed the last fatal complication in a disease course is problematic: are we not plausibly engaged in life-saving “cures” when we try to treat already-developed diabetes even though the person will likely die more directly from a vascular disease that is a “complication” of diabetes? Also, while we might typically think that those we take steps to cure or to rescue are sunk without our help, this need not be the case: it is clear that we can try to rescue people who are plainly *not* otherwise certain to die. The lifeguard may jump into the pool to save someone who might have struggled to the side of the

The two distinct issues are indeed each at play in the Baby Jessica case: when we direct resources to save Jessica, we both know she is already in jeopardy (i.e., it is a rescue case) *and* know precisely who she is in the strongest or most vivid sense (she is personalized, pictured, and named). But the issues could be completely separated: we might choose to expend more money to reduce the prospective risks that a known person or known persons would have an accident or develop a fatal malady than we would spend to reduce the prospective risks faced by unknown persons. In this regard, consider a decision to install expensive safety devices for all of the named and known racers at a particular NASCAR event rather than highway dividers that will save some unknown drivers in the future. We might also expend more resources researching how to *cure* a disease, not knowing who will be cured, than we would spend to *prevent the development (or spread)* of the same or another equally fatal disease in other unknown persons. Consider, for example, spending more to research better anti-virals than to develop vaccinations, as well as spending more to develop better medications to limit the impact of Type II diabetes rather than devising public health measures to shift diet in such a way that fewer cases will emerge.

It is not simply a stickler's analytical point that we should differentiate the choice between "cure" and "prevention" from the choice between curing or preventing the development of disease in an "identified person" and an "unidentified person." In fact, this paper aims to show that the willingness to spend on saving those already in jeopardy rather than to prevent peril from developing is significantly, and perhaps even dominantly, a reflection of our distinct reactions to *death*, on the one hand, and to *dying* and the *dread of death*, on the other. I believe that many of us have these distinct reactions and might rationally manifest them, even in thinking about what we would want done in our own cases. The preference for "curing" or "saving" over "preventing" is not, then, merely about selective empathy for the "identified"—justified or not. Nor is it solely about mistaken perceptions of the effects of our actions on what some mistakenly see as the sole outcome variable of interest—actual premature mortality rates.⁴ By contrast, I will argue that our comparatively strong sympathy

pool without her aid. But once we see that those who are saved or cured might merely be at risk (it is substantially certain that they are already suffering from a fatal *vulnerability*, not that they will die without intervention), then we can more readily think of those who, for example, have a genetic predisposition to develop cancers if exposed to certain environmental carcinogens (classic "prevention" cases) as already-in-jeopardy.

4. Some mistakenly see actual premature mortality rates as the sole outcome variable of interest. See, e.g., George Loewenstein, Deborah A. Small & Jeff Strnad, *Statistical, Identifiable, and Iconic Victims*, in BEHAVIORAL PUBLIC FINANCE 32, 37 (Edward J. McCaffery & Joel Slemrod eds., 2006) (arguing that the identifiable victim effect may be properly exploited to increase aid that would not otherwise be forthcoming but the underlying "first best" goal is increase social welfare

for identified over unidentified persons predominantly arises from non-utilitarian *moral* views—whether justified or not—that are largely of interest in thinking about our reactions to others.⁵

Resolving these issues is certainly of practical significance. For one, public resource allocation decisions are sensitive to the resolution of these issues. Finite resources could be invested, at the margin, in safer highways, or they could be invested in better emergency rooms that save the inflated number of crash victims we generate by not making the roads safer in the first instance. Moreover, though there are obviously a host of empirical issues about the deterrent effect of the death penalty,⁶ death penalty proponents could surely argue that policy debate in the area is “distorted” by the distinct reactions people have to killing an already identified convicted murderer rather than saving unidentified statistical lives who would never be killed if their potential murderers are deterred.⁷ Moreover, we may even, in making our private judgments, overvalue the reduction of *identifiable symptoms* at the expense of accepting higher but unidentified, diffuse risks of mortality and morbidity for reasons that parallel—under certain but not all conceptions—our reactions to saving identifiable victims. Under some obviously medically contested views, hormone replacement therapy is just such a treatment.⁸

by equalizing social benefits, e.g., lives saved, per aid dollar).

5. Because we invariably confront the decision to rescue in the context of rescuing a relatively identifiable person, it is possible that our “intuitions” about the choice between “rescue” and “prevention” are rules of thumb that build upon more basic feelings about some form or other of identifiability. (Even if we are saving a trapped earthquake victim whose name we do not yet know—thus he is not strongly or thickly identified—he will surely be known to be a particular person and we can plainly imagine his peril as a particularized sufferer.) Perhaps we would choose to expend resources on institutions that would cure rather than those that would prevent (e.g., emergency rooms rather than safer roads) because we know at some level that by the time the emergency room is “used,” it will be used on an identifiable victim, even though at the time we expend the resources, the party to be saved is not yet identified. We will never learn the identity of those who did not get killed because we built a safer road.

6. For a good summary of the literature—albeit a summary from the vantage point of authors skeptical that substantial deterrent effects exist, see generally John Donohue & Justin Wolfers, *Uses and Abuses of Statistical Evidence in the Death Penalty Debate*, 58 STAN. L. REV. 791 (2005).

7. This is precisely the sort of argument that is made in Cass R. Sunstein & Adrian Vermeule, *Is Capital Punishment Morally Required? Acts, Omissions, and Life-Life Tradeoffs*, 58 STAN. L. REV. 703, 710, 741 (2005), which argues that the capital punishment debate is affected by the saliency of persons sentenced to death compared to the namelessness and facelessness of statistical lives saved by capital punishment.

8. Hormone replacement therapy may alleviate both short-term symptoms of menopause (e.g., hot flashes, fat redistribution) and longer-term problems (e.g., osteoporosis, muscle aging). But it may be associated with increased risk of heart disease, stroke, breast cancer, and, when estrogen is administered without a progestin, endometrial cancer. The most significant piece in the public

But while I am quite interested in resolving these difficult practical problems, all I realistically hope to do in this piece is to make certain observations about the problems that will do no more than spur a more nuanced consideration of the practical options. Reflecting on these issues also raises far more widespread and general conceptual difficulties we face in evaluating both public policy formation and private decision making, forcing us to consider carefully certain disputes over the propriety or completeness of typical consequentialist methodologies.

My more specific goal is both to embrace, in part, and resist, in part, the mainstream intuition among fundamentally utilitarian “policy wonks” that subjects who spend more to save an identified life than they would choose to spend to prevent the death of an unidentified person are, at core, the poster children for the persistence of irrationality and error. This is the primary intuition of those schooled in the notion that rational decision making is *normative*, but that actors frequently are cognitively *incapable* of making rational choices, not simply because they lack information about the expected effects of their choices, but because their internal processing capacities are too limited to process the information that they do have. What lies just one baby step behind this intuition is the claim that we can impute only one sensible end to actors in these settings—a consequentialist commitment to minimize premature death. (It is critical to note that I am setting aside, now and for the remainder of this piece, issues that arise from the fact that differential physical morbidity might be associated with distinct deaths and paths to death. The fact that people might physically suffer more if we extend their lives once they are ill is clearly important to rational-choice (“rat-choice”) consequentialists. Obviously such consequentialists might have different value schemes, and thus make distinct trade-offs between extending the total number of years lived and the *quality-adjusted* number of years lived.)⁹

health literature criticizing the therapy is Writing Group for the Women’s Health Initiative Investigators, *Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women: Principal Results From the Women’s Health Initiative Randomized Controlled Trial*, 288 JAMA 321 (2002).

9. For a lucid discussion of Quality Adjusted Life Years, a concept developed in significant part to facilitate various sorts of cost-benefit analyses of distinct medical interventions, see Erik Nord, *Methods for Quality Adjustment of Life Years*, 34 SOC. SCI. MED. 359 (1992); and John Broome, *Qalys*, 50 J. PUB. ECON. 149 (1993).

There are obviously a host of practical and conceptual difficulties inherent in trying to measure losses from different levels of morbidity or to measure morbidity at all relative to lost years of life, but these are all beside the point for the purposes of this piece. For one of the classic early discussions of how consequentialist policy analysts ought to make trades between longevity and quality-of-life, see Joseph S. Pliskin & Clyde H. Beck, Jr., *A Health Index for Patient Selection: A Value Function Approach with Application to Chronic Renal Failure*, 22 MGMT. SCI. 1009, 1010 (1976) (“The two factors, increments to longevity and quality of life, are the

If increasing the number of years people, including ourselves, live is the only rational goal to seek in this domain, then it follows readily that these actors could better meet this goal if they transferred an incremental dollar invested in death-decreasing projects from a death-decreasing project that saved fewer lives to one that saved more lives. If the subject is not following this policy, he must be making one of two forms of cognitive error: He might misperceive the effect of his actions, or to put the point more narrowly, he might misperceive the number of lives that will be saved by taking each action because he has a distorted understanding of the probabilities that particular outcomes follow from particular actions. Alternatively, he might not see that outcomes that are in fact the same are really the same because the outcomes are framed differently, or, perhaps more frequently, his views of the relative desirability of the options are elicited using particular methods that influence how desirable each outcome seems.

What I hope is manifestly clear is that this sort of critique of decision-maker capacity is at the heart of the heuristics and biases (hereinafter, H&B) literature in psychology,¹⁰ associated with Nobel laureate Daniel Kahneman and his long-time collaborator, Amos Tversky. The H&B literature is significantly incorporated in the policy analytical world under the label “behavioral economics.”¹¹ People may try to compute expected values but they cannot estimate the probability that distinct outcomes will eventuate if they make certain concrete choices, and they have trouble coming to stable evaluations of the outcomes whose probabilities of eventuating they have already miscomputed. A subject *wants* to take risk-minimizing actions or to insure against risks that are more prevalent, but mistakenly perceives, for instance, that those risks readily “available” to him—that is, easily recalled or brought to mind (airplane crashes are the canonical example) are more commonplace than other risks, factually more prevalent, that are less salient and harder to recall (bathtub falls are the canonical example here).¹² A subject believes he can directly assess how much he values a nice pen, relative to a sum of money, but he values the pen more

predominant dimensions when evaluating response to treatment....”). For my purposes here, though, it is sufficient to note that those who think subjects should have simple consequentialist ends would believe that caring about physical morbidity, however measured, was perfectly sensible, but would find it as senseless to care about the morbidity of the identifiable person rather than the statistical person as to care more about the identifiable person’s death.

10. The best summary of work in this tradition remains HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT (Thomas Gilovich, Dale Griffin & Daniel Kahneman eds., 2002).

11. The most lucid single summary of the quite varied behavioral economics tradition, as applied to law, remains Christine Jolls, Cass R. Sunstein & Richard Thaler, *Behavioral Responses to Law and Economics*, 50 STAN. L. REV. 1471 (1998).

12. The canonical work on biased probability estimates that can grow out of using availability as a proxy for frequency is Amos Tversky & Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, 5 COGNITIVE PSYCHOL. 207 (1973).

highly when he is presented with an irrelevant third alternative—a pen that is far inferior to the pen whose value he is realistically assessing relative to the money.¹³ I am quite sympathetic to the H&B literature,¹⁴ but also try to remain aware of the standard problems within the tradition.

What *are* the standard problems I will need to deal with? The first problem is the problem I am most concerned with, and it is the topic of Part II. Even if our subjects are or should be simple consequentialists, it is possible that we are profoundly wrong when we say that they have failed to meet the simple end we have imputed to them. We may be ignoring the possibility that they are seeking alternative or supplementary ends efficaciously. What I will argue in this regard is that rational subjects may seek not only to minimize deaths or, to put it another way, maximize either their own or collective life expectancy, but that they seek to alter *how people die*. In discussing this point, recall that I am completely setting aside significant distinctions in physical pain and morbidity associated with distinct paths to death and will focus solely on the possibility that what they are trying to do is not simply to maximize how long they live but to dampen certain aspects of the existential despair associated with *dying*, rather than the quite distinct—and actually quite elusive, perhaps even non-existent—well-being losses associated with *being dead*.¹⁵ It is possible to interpret what I am doing as merely *refining* or amending our conception of quality of life by noting that those who know they are dying suffer in ways we need to be cognizant of, even if they are not suffering in ways that those conventionally described as physically ill are suffering.

13. The seminal works on this sort of context-dependent decision making are Itamar Simonson & Amos Tversky, *Choice in Context: Tradeoff Contrast and Extremeness Aversion*, 29 J. MARKETING RES. 281 (1992); and Amos Tversky & Itamar Simonson, *Context-Dependent Preferences*, 39 MGMT. SCI. 1179 (1993).

14. I started writing in this tradition more than thirty years ago, and I got my first job at Stanford on the basis of a third-year law school paper that independently worked through two of the issues that the early heuristics and biases researchers were working on: the endowment effect and problems of self-governance over time (and meta-preferences). See Mark Kelman, *Choice and Utility*, 1979 WISC. L. REV. 769; Mark Kelman, *Consumption Theory, Production Theory, and Ideology in the Coase Theorem*, 52 S. CAL. L. REV. 669 (1979). I give a sympathetic summary and reconstruction of some of the major themes in the literature in Chapter 2 of MARK KELMAN, *THE HEURISTICS DEBATE: ITS NATURE AND ITS IMPLICATIONS FOR LAW AND POLICYMAKING* (forthcoming Oxford University Press 2011) [hereinafter *THE HEURISTICS DEBATE*].

15. In this regard, I am trying to give more content to a rather casual observation made by Thomas Schelling in the first academic article that considered the identifiable victim effect, though the observation was never followed up on in any of the subsequent psychological literature. Schelling noted that “the pain associated with the awareness of death—with the prospect of death—is probably often commensurate with the costs of death itself.” T.C. Schelling, *The Life You Save May Be Your Own*, in PROBLEMS IN PUBLIC EXPENDITURE ANALYSIS 127, 144 (Samuel B. Chase, Jr. ed., 1968).

Next, I discuss in Part III that it is possible that we are wrong to assume that our subjects are bumbling consequentialists, or even consequentialists with a surprisingly complex array of goals, rather than *non-consequentialists*. They may be non-consequentialists because their moral or decisional code is agent-relative, rather than agent-neutral. They may be agent-relative in the sense that they are interested in what they as individuals *do* (and whether they cause certain outcomes purposely, knowingly, recklessly, or negligently), and in particular whether they violate the “rights” of other agents or treat each of them with the respect owed separate autonomous rational agents. They are not just interested in the results that are likely or necessary causal outcomes of their entire course of acts and omissions (sometimes, but not always, including the violation of rights by third parties). Think, in this regard, about the standard, if hardly uncontroversial, deontological intuition that an actor A could not justly “punish” an innocent V to save another set of innocent Victims X, Y and Z from a mob that would be satiated by V’s punishment, because A’s *conduct* in such a case would impermissibly violate rights, even though it would arguably causally prevent the violation of rights by others. Or think about the less commonplace Kantian intuition that one should not lie to another, even if doing so would deprive the lied-to person of information he would use to harm another.¹⁶ Are

16. There are a host of such dilemmas for strong deontologists. My intuition is that one can probably most readily assess just how strong a subject’s attraction to the view that one can never do anything “wrong” oneself, whether or not one’s “wrongdoing” might interfere with a separate moral agent’s plans to cause even really bad consequences, by whether he thinks it worth more than a second’s reflection to resolve the cherished Kantian dilemma of whether it is morally permissible to lie to a would-be killer about the whereabouts of his intended victim. Some neo-Kantians—like Korsgaard—ultimately argue that one can lie to the would-be murderer only on the supposition that he is trying to deceive you about his plans, but seem to feel that lying in other cases is always categorically impermissible, because doing so fails to show “respect” for his autonomous decision-making capacity, since one deprives him of the information he needs to make autonomous choices that meet his considered goals. See CHRISTINE M. KORSGAARD, *CREATING THE KINGDOM OF ENDS* 133-58, 335-62 (1996). Others, like Schapiro, believe that one is obliged to tell the truth only when the purposes of truth-telling are served: she argues that they may not be served when the person to whom one lies lacks the capacity to make autonomous choices in any case (this justifies paternalistic deception, not in any situation in which it could improve decision making and never with that goal in mind, but only in cases in which one is dealing with someone who lacks autonomous rational agency) or when, as in the case of the would-be murderer, he has betrayed the communicative relationship by “withdrawing” from a commitment to engage in joint legislative projects (“defensive deception”). See generally Tamar Schapiro, *Kantian Rigorism and Mitigating Circumstances*, 117 *ETHICS* 32 (2006).

For most of us, I suspect the hard question is whether anyone *giving* the true information about the victim’s whereabouts, knowing how it is to be used, is complicit with the murderer, or whether mere knowledge that one is aiding him—rather than a purpose to do so—is insufficient. The question of whether one is actually obliged to lie rather than to shut up simply reduces to a question

those who devote more resources and energy to saving identifiable lives parallel to, though not precisely like, those who believe that their duties to avoid harming or violating rights directly are considerably more extensive than their duties to save or prevent others from violating rights? If the failure to take precautions that decrease global risk is most like “failing to act” and killing a person is most like unambiguous rights-violating action, then perhaps failing to take steps to save an already-identified person occupies a moral midpoint, neither as “bad” as killing nor as “acceptable” as failing to save, for an agent relativist focused on his or her own conduct. I will make what I view as a very tentative argument to that effect in Part III. At core, the argument is that the clearest form of problematic conduct or rights violation involves a dyadic relationship between a uniquely situated identifiable perpetrator and an identifiable victim whom the perpetrator injures. One stays closer to the core case so long as there is identifiability on at least one side of the dyad.

But our decision makers might also be non-consequentialists because they make judgments that are agent-relative in a distinct sense. They might see no reason to treat all deaths as equally bad even if they accept the abstract moral equality of those who will die because they are entitled to demonstrate partiality and preference for some subset of persons who might die. Again, thinking about the canonical example should help. A father might act justly in saving his own child from drowning after a boating accident, even though he could have saved two other equally worthy children had he not devoted his energies to saving his own child. While obviously the relationships with identified strangers are weaker than with kin or friends, it is possible that we establish non-trivial “relationships” with those we know are dying merely by virtue of learning their identity—relationships that justify at least a weak form of partiality.¹⁷

I will return briefly, in Part IV, to consider whether the preference for prolonging the life of identifiable persons, separated out as much as possible

of whether a stranger has an affirmative duty to save. But if she chooses to answer questions about the victim’s whereabouts at all, she may have no choice but to lie if knowing aid is a form of complicity.

There are also plainly those whose deontological intuitions are stronger in situations in which one set of unwanted consequences involves not only actor conduct but also impermissible rights violations by the actor-in-question while another does not. Thomson, for instance, believed at one point that the reason one cannot push the man off the drawbridge to block a trolley that will otherwise kill five innocents while one could divert a trolley about to run five people over so that it kills just one is that in the former case, the pushed party has a “right” to be free from the homicidal assault while none of the possible “accident victims” has such a right. Her view was that in the divert-the-trolley case, none of the six possible victims has a right be free from being accidentally run over. See Judith Jarvis Thomson, *The Trolley Problem*, 94 *YALE L.J.* 1395, 1403-06 (1985).

17. Sarah Clark Miller makes a more complete variant of this argument in relationship to the duty to rescue. See generally Sarah Clark Miller, *Need, Care and Obligation*, 80 *ROYAL INST. PHIL.* (Supp. 57) 137 (2005).

from the frequently associated question of whether we are *saving* those already in jeopardy rather than preventing jeopardy from developing, can be understood in some significant part by reference to this second form of agent relativity. I will briefly comment as well on whether we should foster this seeming consequentialist “error” if and perhaps only if it serves consequentialist ends to do so. In this regard, I will make note of empirical findings that indicate that even when people are alerted to the fact that many people expend more resources to save identifiable victims than statistical lives, they do not become more generous towards unidentified strangers. Instead, their altruistic concern for identified people diminishes.

I. DO SUBJECTS MEET THEIR GOALS INEFFICACIOUSLY OR DO CRITICS MISCOMPREHEND THEIR GOALS?

A. Conceptualizing Distinct Hesitations about Claims of Cognitive Error

Heuristics and biases (H&B) researchers claim that they have experimentally demonstrated that subjects are frequently unable to meet a set of stable ends because of cognitive distortions.¹⁸ It is fruitful to begin by distinguishing two distinct sorts of arguments made by those suspicious of this claim. Those taking the first approach—most associated with Gerd Gigerenzer’s “fast and frugal heuristics” (hereinafter, F&F) school¹⁹ and its critiques of Kahneman and Tversky’s work on biased judgment²⁰—argue that subjects are and should be less conventionally rational. Instead, they typically use nonrational judgment and decision-making tools that take advantage of the readily available information in natural environments to make behavioral decisions that meet the organisms’ proximal goals. The theory is that subjects’ decisions are not conventional rat-

18. It is easiest to conceive of H&B scholars as criticizing the descriptive realism, rather than the normative desirability, of expected utility theory: H&B theorists think we may both miscalculate the probability of the events that might occur if we take any particular action and cannot consistently and stably evaluate how we would value the end-states whose probability we may well misconstrue. I discuss these points at length in Chapter 2 of *THE HEURISTICS DEBATE*. See *supra* note 14. A similar sort of argument about the most basic message of the H&B school can be seen in Richard Thaler, *Toward a Positive Theory of Consumer Choice*, 1 *J. ECON. BEHAV. & ORG.* 39 (1980).

19. Some of the classic works in this tradition are *BOUNDED RATIONALITY: THE ADAPTIVE TOOLBOX* (Gerd Gigerenzer & Reinhard Selten eds., 2002); *GERD GIGERENZER, ADAPTIVE THINKING: RATIONALITY IN THE REAL WORLD* (2002); *HEURISTICS AND THE LAW* (Gerd Gigerenzer & Christoph Engel eds., 2006).

20. There are many such critiques. Among the most prominent are Gerd Gigerenzer, *How To Make Cognitive Illusions Disappear: Beyond Heuristics and Biases*, 2 *EUROPEAN REV. SOC. PSYCHOL.* 83 (1991); and Gerd Gigerenzer, *On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky*, 103 *PSYCHOL. REV.* 592 (1996).

choice, logical decisions, meeting conscious goals that the H&B researchers simply failed to take into account. Rather, they are decisions that are ecologically rather than logically rational, meeting sensible, even obvious, ends whether or not they follow canons of rationality.²¹ The second approach²²—more associated with rat-choice resistance to the findings of H&B research that people fail to act rationally—is that subjects who seemingly fail to meet ends ascribed to them by experimenters actually are consciously or semi-consciously meeting a separate, less obvious end.²³

Here is an example of the distinction between these two types of critiques of researchers who believe subjects are failing to pursue their ends: Assume that

21. I discuss and critique this approach at length in *THE HEURISTICS DEBATE*, *supra* note 14, Chapters 3 and 5. I discuss the criticisms of H&B theory in Chapter 4.

Here is an example that should help illustrate the distinction between ecological and logical rationality: people often fallaciously believe that a player who has made many recent shots in a basketball game would continue to make them (the so-called “hot-hand fallacy”). See Bruce D. Burns, *Heuristics as Beliefs and as Behaviors: The Adaptiveness of the ‘Hot Hand,’* 48 *COGNITIVE PSYCHOL.* 295 (2004). Burns acknowledges at least for argument’s sake that conventional H&B researchers were correct to assert that people are mistaken to believe in the “hot hand” in basketball. That is to say, people are wrong to believe that the probability that a player will make his next shot is dependent on whether that particular player has made or missed his previous shots. For the initial statement of this finding, see Thomas Gilovich, Robert Vallone & Amos Tversky, *The Hot Hand in Basketball: On the Misperception of Random Sequences*, 17 *COGNITIVE PSYCHOL.* 295 (1985).

But Burns argues that it is nonetheless *ecologically rational* to take the action that belief in the hot hand “fallacy” suggests one should take: to pass the ball to the “hot shooter” and try to ensure that the shooter with the hot hand gets as many shots as possible. It is a rational belief because it entails adaptive action: Even if the *reason* that one does the right thing is that one mistakenly believes that his shooting percentage will be higher if he is “hot” than “cold,” so long as following the hot-hand heuristic meets the overarching rational ends of the team—to score more points—the “fallacy” is ecologically rational.

22. The second of the two arguments is much closer to the one I ultimately want to emphasize when I make an effort to distinguish between subjects seeking to avert death and those seeking to cushion the existential despair associated with the awareness that they are dying.

23. I emphasized that this sort of rat-choice critique of the findings of researchers who purported to demonstrate bias or incompetence was ubiquitous—and sometimes, though not always, credible—in Mark Kelman, *Law and Behavioral Science: Conceptual Overviews*, 97 *Nw. U. L. REV.* 1347, 1364-72 (2003). Many H&B experiments are subject to the critique that subjects may have sought to meet different ends than the experimenters believed they had pursued irrationally.

Many are subject to yet a different criticism; H&B researchers may claim that subjects evaluate the *same* option differently depending upon how preferences among options are elicited; rat-choice critics may respond that the subject is not really evaluating the same option in each setting, but that the option he rejects given one “elicitation mode” is actually more costly to pursue. *Id.* at 1370-71.

experimental subjects are shown an urn with seven red and three green balls and told that a ball will be drawn ten times, with replacement; each time the subject picks the color of the ball that is drawn, she wins \$5. A subject seeking to maximize the amount of money she wins would pick red all ten times, but most subjects engage in what is usually dubbed “probability matching”—picking red seven and green three times because this is the most likely aggregate outcome. They do so even though the expected value of that choice pattern is $0.7 \times 7 + 0.3 \times 3$ or 5.8 correct guesses rather than 7 (0.7×10).²⁴

Theorists who believe we make irrational decisions note that those who seek to maximize their winnings are not doing so—and are not doing so because they are miscomputing expected values. Rat-choice theorists making the sort of critique I am interested in here are prone to argue that the subjects simply have a more complicated utility function than the H&B researchers suggest: arguably, at least, the subjects want not only the money they can earn but want to keep the game-playing task amusing, and choosing red each time is a bore.²⁵ F&F theorists, on the other hand, note that probability matching—even if counterproductive to this specific task—evolved in the context of developing optimal foraging strategies. Strategies in which subjects select a mix of more and less probable “winners” have evolved to meet organism needs, even if the actors never *compute* the advantages of the strategy in either the foraging or the ball-selection context.²⁶

24. See, e.g., Amos Tversky & Ward Edwards, *Information Versus Reward in Binary Choices*, 71 J. EXPERIMENTAL PSYCHOL. 680 (1966); Richard F. West & Keith Stanovich, *Is Probability Matching Smart? Associations Between Probabilistic Choices and Cognitive Ability*, 31 MEMORY & COGNITION 243 (2003).

25. This is one interpretation of the finding that subjects engage in probability matching less frequently when the monetary stakes rise. Experiments finding that people engage in this behavior less frequently when the incentives to maximize expected return are stronger are summarized in Nir Vulkan, *An Economist's Perspective on Probability Matching*, 14 J. ECON. SURV. 101 (2000). At any rate, one could surely argue that the implicit price of keeping oneself amused rises as stakes rise, and people “buy” less amusement as its price goes up. Of course, it is possible that rising monetary stakes simply cause subjects to focus on a task that they otherwise do carelessly and without focus, not because the desire to act “carelessly” is best considered a genuine alternative end, but because attentiveness and care are always effortful.

26. The evolutionary advantage of the strategy in the foraging context is (purportedly) that one should not merely maximize current food gathering, but also learn more about environments in which finding food is less likely, so that one will have alternative sites to find food in the future. Thus, it is suboptimal in meeting the fuller set of goals to maximize expected current winnings. See William S. Cooper & Robert H. Kaplan, *Adaptive Coin-Flipping: A Decision-Theoretic Examination of Natural Selection for Random Individual Variation*, 94 J. THEORETICAL BIOLOGY 135 (1982); Morris P. Fiorina, *A Note on Probability Matching and Rational Choice*, 16 BEHAV. SCI. 158 (1971); Catrin Rode et al., *When and Why Do People Avoid Unknown Probabilities in Decisions Under Uncertainty? Testing Some Predictions from Optimal Foraging Theory*, 72

I do not mean to dismiss the F&F critique; to the contrary, I believe the F&F researchers have done a good job noting a number of significant problems in the H&B findings,²⁷ but have, at the same time, downplayed conceptual and empirical problems in their own work.²⁸ But, as I said, what I want to highlight

COGNITION 269 (1999).

A mainstream rat-choice theorist would argue that the optimal forager—who genuinely learned information by “sampling” sites where the probability of success was lower—was not acting irrationally; instead, he simply had the foresight to maximize the discounted present value of the long-term income stream. Even if F&F theorists were willing to acknowledge that translation (which treats the behavior as logically rational, not just ecologically rational), he would argue that the strategy was not consciously chosen by a subject attempting to maximize discounted expected income.

27. I discuss some of the strengths and weaknesses of these critiques in Chapter 4 of *THE HEURISTICS DEBATE*, *supra* note 14. The most significant, basic critique that F&F theorists level at H&B research is that subjects *seem* to perform sub-optimally in H&B experiments only because they are given problems in these experimental settings that do not accurately mimic problems that they would confront in natural environments. What ultimately *causes* the gap between performance on “real world problems” and laboratory problems is that the mental capacities that evolved are the capacities to solve recurring problems that increase inclusive fitness, not the more diffuse capacity to be an abstractly better calculator (e.g., of expected values). In this view, what is wrong with H&B research is that the H&B investigators have fashioned lab problems that merely test formal problem-solving capacity and then interpret formal failures on these problems as functional failures of cognition.

Among the key works in the F&F literature that raise these concerns are GIGERENZER, *ADAPTIVE THINKING*, *supra* note 19, at 92-123; Gigerenzer, *On Narrow Norms*, *supra* note 20, at 593; Ralph Hertwig & Gerd Gigerenzer, *The ‘Conjunction Fallacy’ Revisited: How Intelligence Looks Like Reasoning Errors*, 12 *J. BEHAV. DECISION MAKING* 275 (1999). Strong versions of these critiques can also be found in Leda Cosmides & John Tooby, *Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions from the Literature on Judgment Under Uncertainty*, 58 *COGNITION* 1 (1996); Gerd Gigerenzer & Klaus Hug, *Domain-Specific Reasoning: Social Contracts, Cheating and Perspective Change*, 42 *COGNITION* 127 (1992); and Catrin Rode et al., *When and Why Do People Avoid Unknown Probabilities*, *supra* note 26.

28. I detail some of the basic criticisms of F&F work in Chapter 5 of *THE HEURISTICS DEBATE*, *supra* note 14. The most contentious criticism leveled against F&F theorists is that they are simply wrong when they declare that their descriptions of the heuristics people use are more detailed and accurate than those that H&B theorists provide. Instead, the critics suspect, the heuristics the F&F people identify are frequently inaccurate idealizations of actual capacities or cognitive strategies—ungrounded both in behavioral observations and in neurobiology—that merely posit a solution to some imputed adaptive *goals*, as if these solutions were observed capacities. To put that point another way, H&B scholars arguably believe that the F&F theorist typically describes a cognitive process without regard to its real nature, but only as the projected solution to the adaptive problem the F&F theorist *imagines* the organism both needed to solve and must have solved in the fashion the theorist posits. This criticism of F&F scholarship is not widespread but can be seen to some extent in Ben R. Newell, *Re-Visions of Rationality?*, 9 *TRENDS COGNITIVE SCI.* 11 (2005).

Critics also note that F&F theorists cannot adequately account for the variability across

are situations in which the theorist examines researchers who posit that subjects are making irrational judgments because they do not appear to be meeting the ends the researcher has posited they are seeking, ignoring the possibility that they might be seeking other ends.

One more example might help: H&B researchers have identified multiple forms of hindsight bias—the problematic tendency to overestimate the ex ante probability that results that ultimately occurred would occur, the tendency to overestimate one’s own prior estimates of the probability that the actual outcome would occur, and the tendency to believe others should have known what actually occurred even when the outcome was not readily cognizable ex ante.²⁹

H&B experimenters note, in this regard, that experimental jurors directed to judge whether police were justified in engaging in a search—in the experimental instructions, the propriety of the judgment to search is *explicitly* grounded only in the ex ante probability of finding relevant contraband or evidence—is altered by learning that contraband was found.³⁰

persons in the use of heuristics: variability is not especially readily reconciled with the claim that adaptive pressures dictate the use of relatively mandatory heuristics. This particular critique is made especially sharply in Keith E. Stanovich & Richard F. West, *Individual Differences in Reasoning: Implications for the Rationality Debate*, in *HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT* 421-440 (Thomas Gilovich, Dale Griffin & Daniel Kahneman eds., 2002).

Finally, and most importantly, there is evidence in many settings that subjects who purportedly use lexical, non-compensatory “fast and frugal” heuristics in fact use additional cues in reaching judgments. Thus, for instance, Gigerenzer (and Goldstein) assert that people will *only* use information about whether or not they recognize a city in deciding that one city is more populous than another. Their assertion was best spelled out in Daniel C. Goldstein & Gerd Gigerenzer, *Models of Ecological Rationality: The Recognition Heuristic*, 109 *PSYCHOL. REV.* 75 (2002). It is rejected in Mark Kelman and Nicholas Richman Kelman, *Revisiting the City Recognition Heuristic* (unpublished working paper, 2008), much of which is repeated in Chapter 5 of *THE HEURISTICS DEBATE*, *supra* note 14. *See also* Ben R. Newell & David R. Shanks, *On the Role of Recognition in Decision Making*, 30 *J. EXPERIMENTAL PSYCHOL.: LEARNING, MEMORY & COGNITION* 923 (2004); Daniel M. Oppenheimer, *Not So Fast! (and Not So Frugal!): Rethinking the Recognition Heuristic*, 90 *COGNITION* B1-B9 (2003); Rudiger F. Pohl, *Empirical Tests of the Recognition Heuristic*, 19 *J. BEHAV. DECISION MAKING* 251 (2006); Tobias Richter & Pamela Spath, *Recognition Is Used as One Cue Among Others in Judgment and Decision Making*, 32 *J. EXPERIMENTAL PSYCHOL.: LEARNING, MEMORY & COGNITION* 150 (2006).

29. The distinction among these three forms of hindsight bias—which are too often inadequately distinguished in the literature—was highlighted in Mark Kelman, David E. Fallas & Hilary Folger, *Decomposing Hindsight Bias*, 16 *J. RISK & UNCERTAINTY* 251 (1998). The seminal work on hindsight bias is Bernard Fischhoff, *Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty*, 1 *J. EXPERIMENTAL PSYCHOL.: HUM. PERCEPTION & PERFORMANCE* 289 (1975). A good summary of the literature can be found in Jay J.J. Christensen-Szalanski & Cynthia Fobian Willham, *The Hindsight Bias: A Meta-Analysis*, 48 *ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESS* 147 (1991).

30. Jonathan D. Casper, Kennette Benedict & Jo L. Perry, *Juror Decision Making, Attitudes,*

The cognitive error explanation is that they overestimate the ex ante probability of finding contraband if it is indeed found. But, of course, it is possible that they are implementing an additional or substitute norm beyond the one they are explicitly directed to attend to and that the experimenters impute as their only conceivable end. They may believe, at least to some extent, that only the factually innocent—those without contraband—should be protected from intrusive searches.

B. Death and Dying

The basic case I will deal with grows from my own experience as a cancer patient. In the spring of 2008, I was diagnosed, after an eight-month monitoring period, with a rare form of cancer, a retinal melanoma. The disease is more complicated than I will explain, in large part because it is considerably more complicated than I understand. I will also largely omit the details of the space-age treatment I received. Suffice it to say that a doctor looking suspiciously like the Wizard of Oz first implanted tiny metal balls in my eye. Having the little metal balls in one's eye permits the relevant radiologists to transform X-ray pictures of the tumor into something that seems awfully like an accurate 3D video game target. The target is briefly zapped by "external beam" proton radiation administered not in a medical facility but in a physics department building housing a mildly out-of-date linear accelerator that pretty closely resembles a warehouse in which terrorists on the TV show *24* would be assembling a suitcase bomb or Jack Bauer would be torturing these same terrorists off-site. What I want to account for is an intuition that helps illuminate the difference between dying and death. To simplify, I will assume completely counterfactually that there is a single follow-up visit in which one can not only learn that the treatment failed but also learn that it "succeeded." My intuition is that I would have readily made the following trade: Instead of facing my actual low odds of a "bad" follow-up (again, to simplify, treat them as 1 in 100), I would, if I could, have chosen to reduce those odds to zero in exchange for a substantially higher increase of, say 1 in 50, in my background global risks of dying suddenly and without warning (think lightning, think wayward buses, think an instantly fatal heart attack or stroke) during the same period in which the melanoma realistically could metastasize and kill me. If I am right that I would have made that trade,³¹ the question that arises for me is whether I failed to meet

and the Hindsight Bias, 13 L. & HUM. BEHAV. 291 (1989).

31. I am absolutely confident that I would have made the trade but for the fact that I have a wife and children. Death obviously affects third party survivors (one hopes a great deal in the case of wives or kids), whether or not it affects me (At all? A great deal?). But had I been acting wholly selfishly, it would have been easy to accept the trade-off I posit between a higher risk of death and avoiding the dread of death.

the only rational end one could attribute to me—the desire to maximize my life expectancy. I might have manifested this precise action preference (take the objectively higher risk) solely because cognitive bias led me to misapprehend what my actual life expectancy would be making each choice. Is it rational, though, to dampen the existential anguish of becoming a *dying person* by getting the diagnosis that my treatment had worked, even at the cost of increasing the still-low odds that I would soon be *dead*? It is crucial to note that I am *not* claiming that everyone would make the trade I claim I would make, or even that most people would; I am simply claiming that it reflects a comprehensible subjective “taste,” rather than a cognitive error.³²

32. It is worth noting as well that the subjective taste may be a product not of the form of error that I discuss at length in the text—that I am miscalculating the probability of death. It is also possible that I improperly estimate how badly I will feel if I indeed learn that I am dying relatively imminently, basically because I, like most subjects, underestimate the degree to which people hedonically adapt to most circumstances. Many researchers have argued that our affective forecasting is poor—we overestimate how good we will feel when something good happens (e.g., we think it will make us happier than it does to get tenure or win a lottery) and how bad we will feel if something bad happens (e.g., if we don’t get tenure or have a disabling accident). The standard account, with explanations and data-based support, is given in Daniel T. Gilbert et al., *Durability Bias in Affective Forecasting*, in *HEURISTICS AND BIASES*, *supra* note 10, at 292. See also Shane Frederick & George Loewenstein, *Hedonic Adaptation*, in *WELL-BEING: THE FOUNDATIONS OF HEDONIC PSYCHOLOGY* 302 (Daniel Kahneman, Ed Diener & Norbert Schwarz eds., 1999). If there are indeed “real” biases in affective forecasting, there may well be significant policy implications. For instance, non-disabled people may underestimate the actual quality of life of those with disabilities, mistakenly projecting how they think they would react to disability on to people whose actual reactions are far less adverse; this might lead those who implicitly or explicitly ration health care to undervalue life-saving steps that risk disability or extend the life-span of those with disabilities or lead juries to overcompensate those whose injuries result in disabilities. These implications are explored, for instance, in Samuel R. Bagenstos & Margo Schlanger, *Hedonic Damages, Hedonic Adaptation, and Disability*, 60 *VAND. L. REV.* 745 (2007).

Again, it is not by any means central to my point here, but there are reasons to be skeptical about the empirical findings of those who claim that people hedonically adapt to bad news: for a variety of reasons that I have explored at length in other writing, it is at least plausible that Daniel Kahneman is correct that seeming hedonic adaptation is a reporting error—that people’s moment-by-moment experiences really are worse when bad things have happened and better when good things have but that they *report* reasonably high “overall satisfaction” levels because they report such levels relative to a downward shifting baseline of expectations. Kahneman’s basic argument is laid out in Daniel Kahneman, *Objective Happiness*, in *WELL-BEING: THE FOUNDATIONS OF HEDONIC PSYCHOLOGY*, *supra* note 32, at 3, 11-12. For a somewhat critical discussion of his view, see Mark Kelman, *Hedonic Psychology, Political Theory, and Law: Is Welfarism Possible?*, 52 *BUFF. L. REV.* 1, 52-54 (2004).

What seems more bothersome to me, though, about the hedonic adaptation literature is that, taken seriously, it tells us very little about the impact of shifts in life circumstances, whether policy-sensitive or not; its sole lesson, it seems, is that nothing much matters. To the extent that we

1. *The Possibility of Cognitive Bias*

There are a number of reasons to believe that while I am able to say, “I would trade a 1 in 100 chance of death for a 1 in 50 chance”—that I actually subjectively processed the 1 in 100 chance of death I was discarding as *higher* than the 1 in 50 chance I was ostensibly trading for. It is not at all implausible that I miscomputed the relative probabilities of dying if I took option one (sure cure from melanoma) rather than two (increased sudden death vulnerability) and that it was miscalculation, and only miscalculation, that made the irrational trade seem sensible. Let me just touch on three of the many possible reasons, drawing on the H&B literature, that I might have miscomputed the probability of death. People might make the precise same mistakes when thinking about the more general question I am addressing. People might miscompute the relative death rates that will result from taking life-saving rather than peril-preventing steps, even when given bottom-line probabilities to work with:

Availability heuristic: Death from retinal melanoma, or from any source so well-defined that we are considering saving a person from the peril she is in, had become incredibly salient to me. The thought or prospect of dying in that particular manner was therefore readily “available” or easily retrieved from memory. Ordinarily, of course, judging the frequency of events by reference to their availability is not just quick and easy, but accurate. We typically recall things most readily when we have been exposed to them more frequently, and we have typically been exposed to things more frequently when they indeed occur more frequently. But we may *substitute* availability for more considered multi-cue judgments of frequency even when events are available solely because they are emotionally salient.³³ Thus, I may have thought that death from retinal melanoma was far more common than I purported to “know” it really was, because it was the most readily pictured and recalled form of death. Whether we make this cognitive error because we have directly come to confuse availability for frequency and rationally base predictive probability judgments on past frequencies or

develop a measurement schema that tells us that all states are fundamentally the same (prison, cancer, good marriages, messy divorces, high income, squalor), it might be appropriate to question the schema—whether this entails moving from a focus on distinctions in hedonic states to a focus on distinctions in capabilities, for example, or to a focus on the satisfaction of preferences, or whether it requires us merely to do a better job reflecting on how hard it is to measure hedonic states.

33. The seminal work on the availability heuristic is Tversky & Kahneman, *supra* note 12. For a very good review essay on the pioneering work on availability, see Shelley E. Taylor, *The Availability Bias in Social Perception and Interaction*, in *JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES* 190 (Daniel Kahneman, Paul Slovic & Amos Tversky eds., 1982).

because we compute frequencies by sampling events retrieved from memory is controversial.³⁴ Whatever the reason, it is surely possible that my actual, *working* subjective estimate of the probability of death from retinal melanoma was much higher than what I ostensibly learned it to be. That is to say, I “thought” that to be cured of the melanoma was to defuse a risk considerably higher than 1 in 100, and I judged the trade I had proposed based on these distorted subjective probabilities.

Affect heuristic: Both academic psychologists in the H&B tradition and clinicians who use “cognitive behavioral therapy” techniques believe that people’s cognitive judgments often reflect the strength of their emotional reactions to situations. Whether we call this an “affect heuristic” (following Slovic and the H&B researchers)³⁵ or “emotional reasoning” (following modern-day cognitive behavioral therapists like Burns),³⁶ the mechanism is much the same. We treat the intensity of our fear of an event as diagnostic of the actual danger of the event. Since we are often afraid when events are objectively scary, we come to think that events must be objectively scary if we are afraid, even though our emotional reactions may have a multiplicity of sources. Again, to take my own case, it is perfectly plausible that my focus on and fear of both the follow-up visit itself and death by melanoma (availability may do its work through its impact on affect) made me believe that there was an “objectively” higher probability of dying from the disease than I ostensibly “knew” there was.

Distortions in Aggregating Alternatives: People may have a great deal of trouble comparing probabilities of aggregated outcomes, rather than comparing the probabilities of the most salient members of groups. Let me take an example from experiments on lottery choices: consider two possible lotteries in which the objective probability of drawing a winning ticket is 20% (15 of 75). In each case, one wins if one draws one of the fifteen “blue” cards from the pile and loses if one draws one of the other sixty cards, which come in seven different colors, that one might draw. Most experimental

34. For a fuller discussion of this controversy, concluding that ease of recall is the mechanism that dominates when subjects make “trivial” judgments and pseudo-sampling is the mechanism they use in making highly personally salient judgments, see Norbert Schwarz & Leigh Ann Vaughan, *The Availability Heuristic Revisited: Ease of Recall and Content of Recall as Distinct Sources of Information*, in *HEURISTICS AND BIASES*, *supra* note 10, at 103.

35. See Paul Slovic et al., *The Affect Heuristic*, in *HEURISTICS AND BIASES*, *supra* note 10, at 397. A similar argument is proffered in George F. Loewenstein et al., *Risk as Feelings*, 127 *PSYCHOL. BULL.* 267 (2001). An actor seeking to avoid health or safety risks might substitute his level of emotionally charged fear of X (rather than Y) for a (more relevant) considered judgment that X is actually the more serious threat.

36. DAVID D. BURNS, *WHEN PANIC ATTACKS* (2006).

subjects “feel” much better about entering a lottery with 15 winning blue tickets when the other 60 losing tickets come in bundles of 12 of one other color, 11 of another, 10, 8, 8, 8, and 3 rather than 29 of one other color, 10 of another, 10, 4, 4, 2, and 1.³⁷ Again, there are a variety of explanations for this phenomenon. I happen to be most drawn to the idea that we quickly compute the probability of drawing a winning ticket relative to the probability of drawing the single most probable alternative sort of ticket and then get “anchored” to the idea that we have a good or bad ticket, but the explanations for the finding are essentially beside the point.³⁸ The relevance to my case is fairly straightforward: all the “losing tickets” associated with death from retinal melanoma were of a single sort while the increased odds of death I was willing to take on were diffused across many, many alternative forms of “losing tickets,” none of which seemed common enough to be judged as especially probable relative to death from retinal melanoma. A single event with a 1 in 100 probability might seem more probable than a diffuse group of events with an aggregated probability of 1 in 50—internally labeled or constructed as the “alternative event”—so long as none of the constituent events constituting the alternative has a probability nearly as high as 1 in 100.

At the same time, it is possible that I not only misapprehended the actual relative probability of dying from the melanoma and dying from “increased diffuse risk”—even though I had stated the trade in terms that explicitly referred to the higher probability of dying from increased diffuse risks—but that I was subject to rationally indefensible framing effects. These framing effects arguably made me evaluate freedom from risk of my melanoma differently than I evaluated freedom from the hypothetically equivalent or greater risks I would voluntarily choose to take on.

So, again, here’s one of many plausible framing mechanisms I might have used that has been identified in the H&B literature that might account for this: subjects are told that a disease has been discovered that will kill 600 persons if

37. See Paul D. Windschitl & Michael E. Young, *The Influence of Alternative Outcomes on Gut-Level Perception of Certainty*, 85 *ORG. BEHAV. & HUM. DECISION PROCESS* 109 (2001).

38. One could argue that this phenomenon is not radically different from selections made in accordance with “contrast effects.” Subjects asked to pick between a Mark Cross pen and \$6 are more likely to pick the pen if also offered a markedly inferior pen (thus violating a cardinal principle of rational choice—that the presence of irrelevant alternatives should not impact a decision between two relevant ones). My intuition has always been that such contrast effects also arise from the fact that one judgment is easy—the Mark Cross pen is preferable to the cruddy pen—and that subjects then anchor to the choice of the Mark Cross pen when they have to make the more difficult choice between the pen and the cash. See Tversky & Simonson, *Context-Dependent Preferences*, *supra* note 13; Mark Kelman, Yuval Rottenstreich & Amos Tversky, *Context-Dependence in Legal Decision Making*, 25 *J. LEGAL STUD.* 287 (1996).

untreated, and asked to select between two alternative treatment programs. Most pick a program that will *save* 200 lives over one with a 1/3 chance of saving 600 and a 2/3 chance of saving none. At the same time, they pick a program with a 1/3 chance that no one will die and a 2/3 chance that 600 will die over a program in which 400 will surely die.³⁹ They make these distinct choices even though the programs are substantively identical; the outcomes are merely framed differently.⁴⁰ People may indeed be risk seekers when it comes to avoiding deaths (losses) and risk avoiders when it comes to being saved (gains). I may have construed myself as already dying from melanoma and constructed freedom from that form of death as a gain, as a form of being saved, and I was risk averse when it came to being saved, while I was a risk-seeker when thinking about dying from *new* diffuse risks, events that I would construe as losses.

In thinking about the conventional literature on the identifiable victims effect, it is important to realize that individuals who thought that those who would willingly devote some sum \$X to save the identified person rather than, say, 50 unidentified persons, were making just these sorts of computational errors. Thus, for instance, in their first set of experiments—which they subsequently substantially disclaimed for reasons I will return to in the discussion of whether our subjects are non-consequentialists rather than failed consequentialists—Loewenstein and various colleagues dominantly attributed the identifiable victims effect either to a computational error or an irrational projection from an individually rational decision to one that is irrational when

39. This well-known problem was originally presented in Amos Tversky & Daniel Kahneman, *The Framing of Decisions and the Psychology of Choice*, 211 *SCI.* 453 (1981).

F&F theorists believe that the finding is a perfect exemplar of misleading H&B experiments. The claim, at core, is that experimental subjects are interpreting the instructions with a better ear for the conversational implications of the words the experimenter utters than the experimenter have: When we tell the subjects that 200 people will be saved by treatment 1, it leaves open the possibility that more than 200 people will live even though they have the disease. When we say that 400 people will be dead after treatment 2, those 400 people are once-and-for-all goners. A program that “saves” 200 (and leaves open the possibility that others will not die) really is better than one that certainly results in 400 deaths. See Anton Kuhberger, *The Framing of Decisions: A New Look at Old Problems*, 62 *ORG. BEHAV. & HUM. DECISION PROCESSES* 230, 231 (1995).

The critique is certainly not implausible, a priori, but it is probably wrong: the fact that experimental subjects with generally superior cognitive skills more typically give the “rational choice normative” answer rather than the answer subject to “framing effects” makes it seem unlikely that subjects are interpreting the instructions in a more nuanced way than the experimenters intended. See, e.g., Keith E. Stanovich & Richard F. West, *Individual Differences in Reasoning: Implications for the Rationality Debate*, in *HEURISTICS AND BIASES*, *supra* note 10, at 421, 436-39.

40. In each case, one is deciding between a 1/3 chance that all 600 people will live and a 100% chance that 200 people, or 1/3 of the people at risk, will live. In the first case, we *frame* the certain outcome as a gain (*saving* 200), while in the second, we frame it as a loss (400 *dying*).

made for a group.⁴¹ It might seem, at first blush, that people would recognize that one saves more lives with an \$X investment if one saves 50 people from developing a fatal malady or dying in an accident than if one saves a smaller number of known persons in peril. Most people reason, however, about the efficacy of their actions by looking at the *proportion* of good results they achieve.⁴² A public health preventative measure might save 50 of 5 million lives (a high number, but a very low proportion) while saving Baby Jessica saves “one of one.”⁴³

41. See Jenni & Loewenstein, *supra* note 2.

42. This point is also made, without reference to the identifiable victims effect, in James Friedrich et al., *Psychophysical Numbing: When Lives Are Valued Less as the Lives at Risk Increase*, 8 J. CONSUMER PSYCHOL. 277 (1999). When asked to indicate how many lives they felt must be saved to justify an \$850 million investment in improved braking systems, 62% of respondents required more lives to be saved when the number of lives at risk was larger. *Id.* at 285. See also David Fetherstonough et al., *Insensitivity to the Value of Human Life: A Study of Psychosocial Numbing*, 14 J. RISK & UNCERTAINTY 283 (1997).

43. “According to the proportion of the reference group at risk explanation, there is not a strict dichotomy between identifiable and statistical lives. Instead, identifiable victims lie at one end of a continuum running from low probability risks spread over the entire population (statistical deaths) to certain deaths for every member of the population ‘identifiable deaths.’” See Jenni & Loewenstein, *supra* note 2, at 239.

Thinking in terms of proportions might be a simple cognitive error in one of two ways. Jenni and Loewenstein highlighted the degree to which individuals find it rational to spend more to reduce risks in their own cases by a fixed amount from a higher baseline rate than from a lower baseline—perhaps echoing Weinstein et al.—because doing so involves the loss of bequest-directed rather than own-consumption-directed income. See, e.g., Milton C. Weinstein, Donald S. Shepard & Joseph S. Pliskin, *The Economic Value of Changing Mortality Probabilities: A Decision Theoretic Approach*, 94 Q.J. ECON. 373, 374-75, 385-89 (1980). But in spending to save others, one is spending own-consumption-oriented income in either case, so the projection from self-interested certainty effects to outward-directed ones would be problematic. It is also possible that people simply overestimate probabilities close to one relative to more modest probabilities.

What is striking, too, regarding the possibility that we are dealing with cognitive error in thinking about proportions, rather than that there is something special about the salience or affect-provoking qualities of the identifiable is that psychophysical numbing—the tendency to value some fixed number of lives X to be saved as more valuable when X is a high proportion of the lives at risk—seems to closely parallel a well-known phenomenon in marketing: people react less to the absolute number of dollars they will save when offered a discount than they react to the percentage discount they will receive. These parallels are explored in detail in James Friedrich et al., *Psychophysical Numbing*, *supra* note 42, at 279-80, 296. For the standard study on the perception of price discounts, see Timothy B. Heath, Subimal Chatterjee & Karen R. France, *Mental Accounting and Changes in Price: The Frame Dependence of Reference Dependence*, 22 J. CONSUMER RES. 90 (1995). Similarly, people value *time* inconsistently in discount-shopping: Far more people will spend a half hour to get a 50% discount on a \$100 item (worth \$50) than will spend the same half hour to get a 5% discount on a \$1000 item (also worth \$50). See, e.g., Richard

Similarly, while the identifiable victims effect seems to persist even in cases in which the victims are not described vividly (no thick identifiability), it is possible that the effect is often created in part by a miscalculation of risk based on either the affect heuristic (vividly described persons evoke strong emotions that create a sense that we are responding to something urgent) or availability (the peril of vivid people is pictured most readily).⁴⁴

Thaler, *Mental Accounting Matters*, 12 J. BEHAV. DECISION MAKING 183, 186 (1999); Amos Tversky & Daniel Kahneman, *The Framing of Decisions and the Rationality of Choice*, 211 SCI. 453 (1981).

It is also possible that we care about the deaths of high proportions of reference groups for reasons that are wholly rational—if, for instance, the reference group that would be wiped out is not defined by simple identifiability but by virtue, say, of its cultural heritage. Jenni and Loewenstein acknowledge this point, but do not seem to realize that if one pushes it hard enough, one might become skeptical of the significance of identifiability per se:

Given that reference group size is often a matter of framing—a reference group of arbitrary size can be specified for virtually any hazard—a blanket endorsement of a policy that treats fatalities differently based on what proportion of the reference group they compose is normatively dubious. For example, it probably makes no sense to treat a disease that kills 100% of the 10% of the population susceptible to it differently from one that kills 10% of the 100% of the population susceptible to it. *However, some reference groups may be more normatively defensible than others. Thus, even after careful consideration, one might be more upset about a disease that kills an entire family or people in a small geographic area than one that kills a similar number of victims from around the country.*

See Jenni & Loewenstein, *supra* note 2, at 239-40 (emphasis added).

But then think about whether we would and should be more upset about a rare, fatal hereditary disease that will kill half the population of a small, culturally endangered tribe; even though the victims are not strongly identified, at least if identification requires *individuation* (we don't know which tribe members will die), we might rationally spend more to prevent those deaths than to prevent a similar number of deaths diffused over the nation's population, even if we could identify the diffuse victims because the death of a high proportion of some groups rather than the death of a high proportion of a wholly arbitrary "group" might have effects beyond the sum of the personal tragedies.

44. Jenni and Loewenstein raised the possibility that the identifiable victim effect is caused by the vividness of the identifiable victims, though they did not think the early experimental evidence strongly supported this explanation. See Jenni & Loewenstein, *supra* note 2, at 235, 237. Loewenstein (and colleagues) largely reject the explanation even more thoroughly in a set of later experiments that I discuss in talking about agent-relative judgments (see *infra* note 48); the gist of the experiments is that people will give more to a certain number of charitable beneficiaries selected from a pool of potential beneficiaries (because the pool remains the same size, the proportion-of-saving explanation drops out) *merely if they are told that the persons who will receive their gifts have already been selected, without any further information about them.* (They are thus—at least seemingly—no more vivid, as well.) See Deborah Small & George Loewenstein, *Helping a Victim or Helping the Victim: Altruism and Identifiability*, 26 J. RISK & UNCERTAINTY 5 (2003).

Loewenstein and his colleagues do not explore one last possible explanation, explicitly grounded in the H&B literature on scope neglect. Subjects making use of the representativeness heuristic typically neglect the scope of a problem. They are willing to pay little more to save 10,000 birds from an oil spill than to save one, because their view of the severity of the problem is grounded in thinking only about the representative instance, the bird pictured as stuck in the oil.⁴⁵ If that is right, we should not be surprised to find that people are willing to spend less *per life saved* on large-scope statistical deaths. If 1000 drowning deaths caused by the failure to shore up the levees seem no worse than 1 drowning death caused by the failure to save the representative victim—say, an elderly, immobile flood victim stuck up on the roof of her rapidly disappearing building—then it is no great shock that people would be willing to spend the same amount on each. If they do that, though, they will spend far less per life in the former case.

2. *The Possibility of Unacknowledged Goals: Death, Dying and Dread*

I acknowledge, then, that it is absolutely *possible* that my intuitive sense—that I would voluntarily choose to lower my life expectancy—was grounded in the very same sorts of irrationality that H&B researchers have always highlighted. But at the same time, I want to note the strong possibility that I had an entirely separate end. There are, in this view, two distinct things that I sought to eliminate, and they are not equally present in the situation I was willing to accept and the situation I was eager to put an end to. The one that we have already focused on, of course, is death, and the intuition that the trade I claim I would have made is irrational is grounded in the idea that it seems to reveal that I would have preferred a higher chance of undesired death to a lower chance. What I want to argue, though, is that there is an entirely separate, and conceivably more significant harm that I was trying to avert by accepting my imaginary trade: a focused awareness that I had tipped over into the world of persons who were

I am not fully convinced that identifiability can ever be experimentally dissociated from vividness, or at least I am not convinced that it has been in the existing experiments that purport to do so. As soon as we believe individuals are identified, we may fill in details of their lives and doing so makes them vivid, even though we are making them more vivid on the basis of imagined traits. (Recall, too, in this regard the problem I raised earlier—of whether we can think about the lives that might be saved by expensive counter-terrorism measures as purely statistical. I am dubious that we can know, first, whether those evaluating such expenditures treat these saved lives as purely abstract and statistical or as identifiable and vivid, because they may first project past well-known deaths from terrorism—Flight 92 victims, World Trade Center victims—to the purportedly abstract lives, and then give the putative victims—projection of past victims—vivid, if imaginary mini-biographies.)

45. See William Desvousges et al., *Measuring Resource Damages with Contingent Valuation: Tests of Validity and Reliability*, in *CONTINGENT VALUATION: A CRITICAL ASSESSMENT* 91 (Jerry Hausman ed., 1993).

dying. I sought to fight off the awareness that I had become a person who would be unable to escape the dread of death, unable to maintain the reassuring fantasy that we typically maintain until we confront the strong possibility of fairly imminent death that life stretches out, if not precisely infinitely, at least indefinitely.

First, let me emphasize the point that death and dying are *different* negative experiences. I need not believe, nor convince anyone else, that death itself is of no moment psychologically. Still, one should acknowledge that there is a perfectly defensible, if controversial and stunningly counterintuitive, position, dating back to Epicurus, that at least for those who do not believe in an afterlife, death is neither bad nor good, in welfarist terms at all. This is fundamentally true because, unlike dying, no subject ever experiences it.⁴⁶ Note that one of the many disquieting implications of this view is that killing causes no harm to anyone but grieving survivors.⁴⁷

The view that death has no welfare consequences is powerful. To the extent that one's intuitions about welfare are fundamentally Benthamite—a person's welfare is measured by the sum of his positive pleasures and his negative pains,

46. Here is one of many versions of the position typically associated in the metaphysics-of-death crowd with Epicurus, the source of this quote:

Death is nothing to us. For all good and evil consists in sensation, but death is deprivation of sensation. And therefore a right understanding that death is nothing to us makes the mortality of life enjoyable, not because it adds to an infinite span of time, but because it takes away the craving for immortality. For there is nothing terrible in life for the man who has truly comprehended that there is nothing terrible in not living. [Death] does not . . . concern either the living or the dead, since for the former it is not, and the latter are no more.

The position associated with Epicurus is defended in Stephen E. Rosenbaum, *Epicurus and Annihilation*, in *THE METAPHYSICS OF DEATH* 267 (John M. Fischer ed., 1993).

47. The legal implications of this view are explored in considerable detail in Matthew D. Adler, *Risk, Death and Harm: The Normative Foundations of Risk Regulation*, 87 *MINN. L. REV.* 1293, 1299-1301, 1389-1444 (2002-03). Adler also reviews defenses and critiques of the claim that death is not a harm to the person who dies, concluding ultimately that death does harm the person who dies, so long as we think of a policymaker or the person P who might die herself comparing P's lifetime well-being to the lifetime well-being of some P' who is just like P but for the fact that P dies at an earlier age. It is less obvious to Adler that we can make any meaningful statements about whether the *momentary* well-being of P' is lower, higher, or the same as the momentary well-being of P at some point in time when P is still alive and P' is dead because it is inaccurate to describe P' as having some hedonic state at all. Her state is not "neutral;" rather, it does not exist. Nonetheless, Adler thinks P' is worse off whether we think of the party's welfare hedonically (defining hedonically superior states as desired states rather than as states that give rise to a single preferred emotion like "pleasure," but without regard to whether the desired states are experienced) in terms of her as having superior objective capacities or in terms of satisfying her preferences. *Id.* at 1303-10, 1321-40.

each of which must be experienced—it takes some not entirely easy or satisfactory work to get to the view that death “harms” the once-living person. Being dead is obviously neither painful nor pleasurable.⁴⁸ While a hedonic utilitarian trying to maximize *aggregate social welfare* could readily recommend a policy that minimized death if living people, on average, had lives that delivered more pleasure than pain, since doing so would improve the aggregate pleasure/pain balance, this sort of preference for death-averting practices would have no distinct force from the dubious utilitarian case for pro-natalist policies.⁴⁹ Dead S is not really distinct from never-born S. While replacing dead S with some alive S could increase social welfare, it is simply by virtue of creating a placeholder for a relativistic utility calculus that must be performed by others, rather than by virtue of his or her “own” hedonic state.⁵⁰ Moreover, even if all we

48. The most prominent attack on the position associated with Epicurus is an attack by Nagel, who argues that a person is harmed if he is deprived of the opportunity to have positive experiences. See Thomas Nagel, *Death*, in *THE METAPHYSICS OF DEATH*, *supra* note 46, at 67. It is clear that the aggregate quantity of social welfare is reduced if a positive experience does not occur. It is unclear, however, that any individual is harmed if he does not have an experience since there is no such thing as an individual incapable of having experiences with whom one could compare an individual who has them.

49. I hope the arguments in the text are clearly distinct from familiar arguments about whether a utilitarian should maximize total utility rather than average utility. For discussions of this distinction, see, for example, JOHN RAWLS, *A THEORY OF JUSTICE* 22, 161-64 (1971). Obviously, killing a person or preventing the birth of a person with below-average utility would increase average utility, but so long as she had positive utility, her birth or survival would increase total utility. I am assuming in the text for simplicity's sake that encouraging birth or survival as social welfare aggregative policies would increase both total and average utility. I mean merely to question whether a hedonic utilitarian can readily conclude from that fact alone that the once-embodied individual S who dies or the hypothetical S who was never conceived is “better off” hedonically, in some morally significant sense, than S who survived or some arbitrarily selected person S, matched as the never-conceived non-S's “comparison partner.” The question in some sense goes to whether aggregate utilitarianism requires giving moral meaning to an imaginary reified collective entity, and that it is not really possible to think morally about anything but real entities. The argument in the text can be partly translated in the following terms: we could know that a policy of “conceiving S” improves total or average utility but still have little idea what it would mean to say that “unconceived S” would be better off if he were conceived, since assigning him a comparison person is wholly arbitrary. It is similarly not clear that once-living S is any more rightly paired with still-living S; dead and never-conceived S may well be indistinguishable non-subjects.

Still, it is worth noting that from the vantage point of the individual, death will always lower his average and total utility if living hedonic states are positive, while if one is merely toting up social welfare functions, the death of someone who we conceive of as harmed by death could still *increase* average utility.

50. It would seem inadequate to say that death is precisely like (neither better nor worse than) never having been born in hedonic utility terms even if the only way we can understand each is that

are attempting to do is to make aggregate social welfare judgments, policy may be indeterminate in a fashion that reveals that the dead and never-born have no particular utility level. What if we could only increase the number of births by allowing some to die? Would we have any way of ascertaining whether death is worse or better than non-birth from a hedonic welfarist viewpoint when neither the dead nor the never-born feel anything?⁵¹

At the same time, the view that death does not harm is powerfully unpersuasive and almost gallingly paradoxical for those whose most fundamental utilitarian intuitions are framed by preference utilitarianism. When the hold-up man says, “your money or your life,” nearly everyone eagerly chooses to hand over his money and almost surely thinks he has never cared nearly so much about getting what he prefers. What we learn from that decision, though, is that death seems worse than the loss of money, and we certainly treat *this* as a substantial harm.

But the paradox is not so readily resolved for the sub-set of preference utilitarians who believe, at core, that satisfying preferences is not an intrinsic source of utility, but instead that our preferences are merely better or worse predictions about the hedonic satisfaction we will gain from the end-states we choose to achieve rather than those we choose to forego. In this view, preference utilitarianism is not so much a theory of the sources of or the definition of utility; rather, it emerged merely as a reactive response to the obvious flaws of early hedonic utilitarianism. The Benthamite hedonic utilitarians had not given a satisfactory response to the primary accusation against them: the charge that they assumed a very narrow, non-catholic set of plausible tastes. Here was the basic problem: Why should every subject care only about pleasure or pain? What about the dutiful? What about the masochistic?⁵² At the same time, preference

each state is less “utility-producing” than living. Surely we cannot explain the sorts of preferences we will soon explore about dying by comparing them to preferences about never having been born; as difficult as it may be to contemplate the *cessation* of one’s subjectivity and consciousness, the notion that we can contemplate a particular pre-conscious or pre-subjective “being” seems even more strained. What the comparison misses is precisely what is hard to capture in the hedonic utility framework, the need to associate sensations with subjects. I will not do more than avert here to the more uncompromising aggregative hedonic utilitarian position, best articulated by Parfit, that we ought to care dominantly about experiences—rather than who has the experiences—because subjects are far thinner and more fleeting than we typically imagine, that the sense of subjectivity/personal identity is significantly illusory. See DEREK PARFIT, *REASONS AND PERSONS* 346 (1984). Not surprisingly, Parfit finds it easy to think of non-existent people as fit place markers for a welfare calculus.

51. It is hard to avoid, at this point, contemplating the old joke in which a group of Talmudic scholars ruminate on whether it is better to have been born or never to have been born before the wisest one settles the argument with the pronouncement, “Gentlemen, I think we would all agree that it is better never to have been born. But how many are so lucky: one in a thousand?”

52. See JAMES GRIFFIN, *WELL-BEING: ITS MEANING, MEASUREMENT AND MORAL IMPORTANCE*

utilitarianism, defined once again as a mere negative reaction to the problems of Benthamite thinking, was not only more catholic about tastes (pleasure is good only insofar as a subject desires it more than, say, duty fulfillment), but was also not subject to the criticism that utilitarians could not say whether option 1 was “better” than option 2 because the pleasure and pain that would result from each option were not tractable, commensurable, nor measurable constructs. Hedonic utilitarians require unavailable cardinal utility measures. The preference utilitarian, however, could extol one option so long as subjects could provide ordinal rankings.

Still, preference utilitarianism, again, defined negatively and reactively, remains conceptually wedded to a more catholic, less pain-and-pleasure centered form of hedonism, or at least what might be better dubbed “experientialism.” Satisfying preferences is not necessarily intrinsically utility-enhancing. It is just the best way of guessing or estimating what made people satisfied, given their catholic attitudes about what ends are desirable. Understood negatively or reactively, it did not eliminate the requirement that utility must be experienced by a sentient subject. I may gain utility from the *prospect* that the money I direct in my will to be used for a particular purpose will get used this way, and other living people may gain utility from seeing that decedents’ wishes are respected, but there is no subject who gains utility when my preference that the money be used in a particular way is satisfied after my death.⁵³

Assume, then, that one believes that preferences are merely predictors of future hedonic states. Assume further that one is sensitive to the fact that preferences might be poor predictors,⁵⁴ particularly if one is under-informed or imprudent. One could then readily argue that the manifest preference for life over death is merely an imprudent or under-informed preference. This would be a hard conclusion to avoid unless it is indeed true that in *some* sense, we could reasonably say that good or even bad experiences turn out to be preferable to no experience. It is not obvious, though, from the vantage point of a hedonic

8 (1996), for one of the most lucid of many statements of this critique of hedonic utilitarian thought.

53. See generally L.W. SUMNER, WELFARE, HAPPINESS AND ETHICS (1996) (explicating this negative or reactive view of preference utilitarianism).

54. The fact that people may predict hedonic outcomes poorly drove some philosophers and the “new hedonic psychologists” back to hedonic utilitarianism. See, e.g., RICHARD BRANDT, A THEORY OF THE GOOD AND THE RIGHT 246-53 (1979) (expressing one such resulting philosophical position); Daniel Kahneman, Peter O. Wakker & Rakesh Sarin, *Back to Bentham? Explorations of Experienced Utility*, 112 Q.J. ECON. 375 (1997) (presenting one such resulting psychological view emphasizing not only that predictive error was possible, but that it was likely to be ubiquitous given our cognitive limitations). For an excellent summary of the psychological literature explaining how poorly people predict the affective outcomes of their choices, see George Loewenstein & David Schkade, *Wouldn't It Be Nice? Predicting Future Feelings*, in WELL-BEING: THE FOUNDATIONS OF HEDONIC PSYCHOLOGY 85 (Daniel Kahneman, Ed Diener & Norbert Schwarz eds., 1999).

utilitarian what it would mean to say that “no experience” is “less pleasurable” than any sort of experience, good or bad. It is also not obvious for a utilitarian interested in attainment of objective goods, one who believes people are better off when they maximize their “capacity” to achieve whatever ends they happen to seek rather than to maximize their enjoyment of experiences, that a person who is dead is less capable in the relevant sense than someone who is alive because she cannot clearly meet her ends more readily. Once again, though, if preferences are merely predictors of hedonic states, then they are of little moment if they are just bad guesses. Naturally, it is more likely that they are just bad guesses, in relationship to this particular choice, because we are unable to choose between options in a fashion that mimics our typical choice pattern at all (deciding which of two end-states is likely to be more desirable or satisfying along whatever dimensions we care about).⁵⁵

Certainly, it is plausible that when we make the choice between life and death, we have no information or real basis for making a prudent judgment; for the secular, at least, death is a state that is utterly beyond our concrete contemplation. Presumably, the fact that our own death is so incomprehensible is part of what triggers our dread of death. It is almost surely the case that when we contemplate our deaths, we are doing so in a fashion we would acknowledge is significantly grounded in error. As Herbert Fingarette put it, “[S]ince I can’t imagine being dead, I imagine, unwittingly, that I am conscious, a presence in that future world, conscious of it and yet utterly cut off from participation in it. That way of imagining my ‘death’ is confusion because in reality I would not be there to suffer such alienation and longing.”⁵⁶ Even if we could contemplate what our own death is “like” to some extent, it is plainly a state we have not learned much about through repeated trial choices, nor—given dread and anxiety about

55. It might be instructive to return for a moment to note that certain sorts of anti-consequentialists might claim that our “choices” are not made by a utility-computing brain, but by a modularized brain that manifests certain choice or action tendencies because such tendencies are the ones that increased inclusive fitness. In this sense, say, if death-avoidance were instinctive (those who didn’t seek to avoid death would not have reproduced as successfully as those who did), it would be odd to describe death-avoidance as utility-maximizing (or as “normative” in any sense) rather than as merely an inevitable proclivity that each member of our species has. The idea that it is daffy to think about the taste for life over death as a preference—rather than as something more like a central biological defining feature of beings, the pre-desiring core that makes preference possible or meaningful—far pre-dates not only modern evolutionary psychologists, but even Darwin. See, e.g., Arthur Schopenhauer, *The World as Will and Idea*, reprinted in HERBERT FINGARETTE, *DEATH: PHILOSOPHICAL SOUNDINGS* 143-44 (1996) (“[T]he fear of death is independent of Reason. Animals have the fear; yet they do not know death. Every creature born into the world brings this fear with it. This a priori fear of death is simply the other face of the Will to Live, which is what we all are. Just as the concern for self-preservation is innate, so too, therefore is the fear of one’s own destruction.”).

56. See FINGARETTE, *supra* note 55, at 9.

death—is the choice to avert it likely to be one that we would make in a calm, detached, and prudent way.

No matter how one resolves the thorny question of whether death is a welfare-reducing harm at all, it is clear that the desire not to be dying in the perceptible short-term reflects a wish to avoid a *distinct* loss of welfare. The desire not to experience an intense and focused dread of death—an experience that I felt I would have if my particular follow-up visit went badly—is simply *different*. While it may well be the case that the existential dread is itself imprudent unless death is itself a bad outcome, that need not be the case. The person who dreads knowing she is dying soon may dread the life one leads and *plainly* experiences once one knows that.

We may powerfully and prudently dread a period of time in which long-term plans seem pointless, and believe that the experience of recognizing that planning is pointless strips us of central aspects of our identity.⁵⁷ Furthermore, we may dread the inability to think that days hold the possibility of significant favorable surprises or variation, and may dread dealing with the discomfort of those around us who know that we are dying. We may dread the loss of the recalcitrant fantasy of indefinite life; the incapacity to imagine or make sense of what nothingness is may be profoundly disquieting. There are reasons that existential psychiatrists press patients to try to recognize that they have all “experienced” nothingness before, in the infinitely long period *before* they were born.⁵⁸ But it is hardly

57. Fingarette argues that what most defines us as people *in the present* is how we exist given a particular conception of both the past and, even more significantly, the future. While we cannot but “live in the present” (in the sense that it is all we experience), it is hardly paradoxical to be focused on plans:

“[L]ive in the moment.” How odd that advice is. Taken literally, it’s utterly superfluous. As if there were any alternative! But what about the future? It’s our nature to live with purpose, to have goals. Shouldn’t we be seriously committed to these? It seems a contradiction to say: Be whole-heartedly committed to the goals you set yourself—but live in the moment.

The paradox is resolved when we examine the inward experience of time. The goals I set for myself, and the future as I see it ahead of me, are present to me now. The past, too, insofar as it exists for me, insofar as it registers in my consciousness, is present to me now.

Id. at 73. “[T]o be without purpose is to lack something essential to the fully human mind.” *Id.* at 70.

58. See IRVING YALOM, *STARING INTO THE SUN* 81-2 (2008). Many philosophers who write about the metaphysics of death recognize that if death is merely non-experience, there is no interesting distinction between the time after we die and the time before we are born. Schopenhauer put it elegantly: “Were it the thought of our nonbeing that made death seem so frightening, we ought to shudder, too, at the thought of the time before we ever came to be.” Schopenhauer, *supra* note 55, at 144. But it is plain that as a matter of fact, people dread the loss of experiences they either expect or can at least imagine, not the simple absence of experience. Trying to account for

intuitive to do so, nor is it obvious that it is helpful to describe pre-birth as an experience at all. Freud famously argued it was simply impossible.⁵⁹ Above all, perhaps, what is almost surely the most powerful aspect of our dread of death—the dread that we will miss events and people, that something like a magnified version of “curiosity” will go unsatisfied—is something we can experience, in the present, once dying. Take the familiar, illustrative paradigm case: the parent who dreads that she will not be alive for her child’s wedding next summer. The pain that one *knows* one’s curiosity won’t be satisfied is experienced while *dying*: it is neither experienced when dead nor when one does not strongly fear imminent mortality. If I know I will die before my kid gets out of high school, I *know* right now, that there are a host of questions I am anguished to know I won’t have answers to; this is a concrete lived experience.⁶⁰ It is likely that a good deal

that distinction between lost and absent experience preoccupies those who think death is in some ways harmful. For attempts to explore or explain the distinction, see, for example, FINGARETTE, *supra* note 55, at 10-11, which emphasizes that one has a more detailed sense of the events that one will miss in the future—they involve people and situations you can picture in detail—while not having experienced periods of history one can only vaguely imagine is relatively painless; PARFIT, *supra* note 50, at 165-85; and Anthony L. Brueckner & John Martin Fischer, *Why Is Death Bad?*, in *THE METAPHYSICS OF DEATH* 221-22 (John Martin Fischer ed., 1993), which argues that there are situations in which it is rational to care more about what will happen than what has happened.

Plainly, these are very difficult issues to get one’s head around: it seems plain that even if the harm of death somehow involves lost opportunities, we need to account for distinctions between lost opportunities that the subject herself was able to imagine having but for death (death of a competent adult with a sense of her future versus, for example, the death of a baby) and the loss of opportunities that a competent adult realistically imagines she might have in the “ordinary course” of events. (Are we able to muster as much sadness at dying at all as we are at dying young, even though, hypothetically, we could imagine experiences stretching out into the indefinite future if we were not to die?)

59. “We cannot, indeed, imagine our own death; whenever we try to do so we find that we survive ourselves as spectators . . . [A]t bottom no one believes in his own death, which amounts to saying: in the unconscious every one of us is convinced of his immortality.” SIGMUND FREUD, *REFLECTIONS ON WAR AND DEATH* 114 (A.A. Brill & Alfred B. Kuttner trans., 1918).

Yalom argues—on behalf of existential psychiatry—that Freud so radically underestimated the centrality of death anxiety in the intra-psychic lives of patients precisely because he thought neurosis arose from the conflict between the conscious and unconscious mind. Since Freud thought the unconscious mind did not fear death, he thought it unlikely that the “buried” or “repressed” fear of death produced intra-psychic distress. See YALOM, *supra* note 58, at 18-19.

60. Fingarette makes this familiar point well, but makes it in the context of arguing that the preference for life over death is rational, even if death is not a (good or bad) experience.

The truth is that appraisal should depend on how I feel *now* about any particular future, not how I’ll feel then . . . suppose I look into the distant future and see my little grandson, John, as an adult, mature and in mid-life. Since my grandson is only a five-year-old now . . . I can have only the faintest and most uncertain glimpses of what he will be like as an adult. But I hunger to know—and I know I never will. Yet it makes sense for me to

of our dread of death—not dying—comes from falsely projecting that same sense of longing for more knowledge, more experiences, more contact into some post-death future where such longing is not really possible.

I do not mean to downplay the possibility that *positive* hedonic states may also be uniquely available to those who know they are dying soon: capacities to reflect on what gave one's life meaning, to garner sincere honor from others, to get one's practical affairs in order, to say proper goodbyes and to reconcile and accept for instance. I merely need to assert that it is not irrational to believe that those gains could be outweighed by the negative experiences for at least some non-trivial sub-set of subjects. Some people might well prefer to die more slowly and knowingly than simply to be dead, having died quickly and without significant consciousness of impending death, but so long as *this* taste is not universal, my point remains unaltered.

To try to explore further how we might react to *dying*, and to distinguish reactions to dying from those to death, I think it is helpful to reflect on the familiar paradoxes that Broome, among others, has raised in criticizing how policy analysts conventionally value lives.⁶¹ Generally, economists and cost-benefit analysts value lives by measuring the inverse of the shadow price placed on accepting increased or foregoing decreased ex ante risks of death. If we will pay only X dollars more for a product that reduces the risk of death by 1/100, or demand only a Y dollar wage premium to tolerate an increased risk of 1/100, we value life, if risk-neutral, by 100X or 100Y.⁶² If “not dying” were like winning a prize in a lottery, this would seemingly be a perfectly reasonable procedure. Broome noted that this valuation method seemed paradoxical since if we knew we were just about to die unless we took some costly step, we would pay so much more than 100X to prevent our death that it would be difficult to explain

wish now that I'd still be alive to see and be with him. My death means to me now that I will never know his future All such doomed wishes are real elements of the present subjective meaning of my future death. They are a source of legitimate *present* sadness.

Fingarette, *supra* note 55, at 69. And Fingarette recognizes, to some degree, what I argue in the text: that until one is dying, one maintains a somewhat false belief that planning is always sensible: “As a practical matter, I have always assumed an indefinite period ahead in which life will continue as usual. I *must* assume this. I must look to tomorrow and tomorrow as days of life, even though I know that at some unspecifiable point there will be no tomorrow.” *Id.* at 72.

61. The two basic works are John Broome, *Trying To Value a Life*, 9 J. PUB. ECON. 91 (1978); and John Broome, *Trying To Value a Life: A Reply*, 12 J. PUB. ECON. 259 (1979).

62. For more sophisticated versions of this basic procedure see, for example, Kevin M. Murphy & Robert H. Topel, *The Value of Health and Longevity*, 114 J. POL. ECON. 871, 884-88 (2006), which accounts for distinctions in the value of additional years of life over the life cycle; and W. Kip Viscusi, *The Value of Life: Estimates with Risks by Occupation and Industry*, 42 ECON. INQUIRY 29 (2004), which attempts to account for distinctions in fatality rates within industries and occupations in assessing wage premia that workers must be offered to accept greater risk.

the ex ante prices in terms of conventional attitudes towards risk.⁶³ Broome's narrow argument, in terms of public cost-benefit analysis, was that the acceptability of public plans was unduly held hostage to random exigencies if one employed the conventional policy analytical procedure: we would accept or reject the precise same program depending on an informational accident. Were victims identifiable? Did we have limited or fuller information about the mortality impact of our proposed action? More particularly, could we find individuals who would know the project would kill *them*, in particular, so that we would have to ask them what they would need to be paid to tolerate the project going forward? Would few, if any, projects pass muster then? But more profoundly, his point was that all estimates of the value of life based on risk tolerance are inadequately informed. In a world where we were smarter and had more determinate information, we could arguably always identify the people who would die from projects now described as merely risky. So either the ex post measures were the apt ones, or we needed to rethink the issue of how to value a life.

I think there are two related points worth raising here, the second perhaps more closely tied to explicating the importance of attending to existential dread. The first, though, is that *neither* the ex ante risk-averting nor the ex post saving perspective are the only ones to take on the value of life. We may better think of the problem as a problem of evaluating *extra time*. We could at least *imagine* taking a whole life utility-maximizing perspective in which people trade off additional lifetime consumption and additional bequests for a defined and certain increase in life years, and make them stick to their judgments. What is critical to note is that the hypothetical trade we are making is *not* between a certain defined long life and lower income, but between lower income and a particular certain *increase* in life years, added incrementally to a still-unknown base life expectancy. That is to say, we can imagine—conceptually if not practically—trading off extra income for one extra year of life without knowing whether that extra year will mean we live for 81 rather than 80 years or for 51 rather than 50 years. What is certain in making the trade-off is the incremental *increase* in time;

63. See Weinstein, Shepard & Pliskin, *The Economic Value of Changing Mortality Probabilities*, *supra* note 43, at 373-96 (arguing that people rationally spend more to save their own lives when they are at higher risk of death because they are more likely to be “merely” spending money they will bequeath to others, rather than money they would be spending themselves, which they value more highly). This “rational choice” explanation, I suspect, has a germ of truth to it even though, at first blush, it seems rather daffy: to keep safe from all low-level risks would be to change one's lifetime consumption pattern a great deal. But I am not sure it is wholly right to characterize the person who would spend a lot when she faces a high probability of death as feeling free to do so because she is “only” spending money she no longer cares about rather than that she is suddenly unable to evaluate alternative spending plans as anything but trivial and small and/or that she is unable to spend in the way that people able to maintain the fantasy of indefinite life are able to.

what remains uncertain is life expectancy. (If we imagine instead a trade in which life span were made certain at a particular income cost, the trade would permit us not only to extend the amount of time we spent alive—to trade consumption goods and bequests for time—but to eliminate the anxiety that we might die over a prolonged period in which we now must suffer that anxiety.)⁶⁴

As I noted, I do not believe as a practical matter that any actual market judgments mimic this procedure. The point is that it is distinguishable from attempting to assess whether our distinct reaction to the *risk of death* is more “authentic” or “revealing” than our reaction to *certain death*. Such judgments of the value of extra time were they possible, likely would not reveal a constant dollar value for additional life years at every point in the life cycle. The marginal value of incremental years would almost surely decline quite dramatically—adding years as a younger or middle-aged person (i.e., not dying “young”) would likely seem radically more valuable than adding years in old age. This would almost surely be the case not merely because old age years are likely less enjoyable but, perhaps more significantly, because to the degree most of us intuitively resist Epicurus’s claim that our own death does not harm us,⁶⁵ we do so, above all, because we yearn to experience conventionally expected life cycle markers.⁶⁶ What is perhaps most critical to recognize is that is quite plausible if this is right that neither the conventional *ex ante* nor conventional *ex post* procedures give us much information about the “value of life.” Judgments about *ex ante* risk-acceptance do not so much measure the value of living (i.e., the value of extra time alive, measured in terms of foregone consumption and bequests) as they measure the value of modest reductions in death-anxiety. And

64. The fact that anxiety over the possibility of death—the psychically costly fear that any terribly bad headache could be a brain tumor or aneurism, that one must check one’s body for potentially fatal lumps—is different still, distinct both from death and dying—can be seen readily by “playing” an introspective game. Consider whether the following set of life expectancy lowering preferences (that I find totally comprehensible) are coherent: would you prefer an 85-year life expectancy with risk of death at any point in your life to an 83-year life expectancy with 77 years certain and uncertainty thereafter?

But lest you think that dying does not matter, that the concern is *only* about the reduction of mortality *anxiety* (from our normal existential state “indefiniteness” to an unimaginably superior one, “non-anxiety” for some period of certain survival), think about whether you would, like me, pick the 83-year expectancy with 77 years of certain life over an 84-year life expectancy with a certain date of death at 84. (In this scenario, as one approaches 84, one is not *anxious* out of uncertainty—one knows what is coming—but one shifts existential frames from living in denial to dying.)

65. *See supra* notes 47-48.

66. Those with children, for instance, might especially dread not seeing what becomes of them as grown-ups, or seeing grandchildren. Perhaps because we have no models around us of people who live more than eight to ten decades, we don’t have a standard set of experiences in mind that would be wiped out by death in old age.

ex post judgments (especially hypothetical ones in which people give honest responses to what they would now pay—if they only could, by going back and saving what they did not save by taking expensive precautions—to cure them of the incurable diseases or rescue them from the hopeless situations their prior risk-taking caused) are typically merely attempts to have one’s cake and eat it too. It is far too easy to say one would now undo choices to consume more or live a more exciting, risky life that may indeed have been perfectly reasonable.⁶⁷

But here is the second point—the point closer to the one I have emphasized throughout Part II about the death/dying distinction. The risk of death is ubiquitous in our lives. We are typically able to background it if we are to avoid paralyzing despair. At most times, our futures are indefinite: a shift within the domain of diffuse, reasonably low risk simply does not alter the sense that our lives still stretch out *indefinitely*. This sense that life stretches out indefinitely has many correlative benefits. We can still care about planning for the future, care about the trivial present without contrasting it with the weightiness of soon-arriving death, feel a rush from the sense of invulnerability or freedom that days are not so finite and precious that any particular day can be badly misspent. We may believe that we will learn the answers to the typically small and personal questions about which we are most curious, that we will not have to soon experience the rupture of relationships. From the ex ante low-risk perspective, it is even plausible to me that most of us truly believe in some hazy, not fully cognized sense that we are immortal: we have only experienced consciousness and a sense of the self, and simply cannot imagine what it would mean for that to end.⁶⁸ The ex post perspective kicks in when we have shifted existential frames, when nothing but our death seems of any moment. When we shift existential frames, we would do nearly anything for life chances to appear meaningfully indefinite again.

If it is right that we are rational to dread dying at least as much as we seek to avoid death, then we can readily comprehend a whole range of private and public decisions that seem to increase or do little to decrease the *risk* of death at any point in “normal” time, but try at great cost, to eliminate or dampen the experience of having to shift existential frames.⁶⁹

67. I have articulated the perspective that it is worth thinking about the value of extra time, not of “life.” This perspective is not enormously distinct from the perspective articulated—in quite different form—by Allan Feldman in a series of unpublished working papers in the late 1990s that I became aware of after I had written this section. See, e.g., Allan Feldman, *The Value of Life Revisited* (Brown Univ. Dep’t of Econ., Working Paper No. 96-2, Jan. 1997).

68. Whether we are maintaining the infantile fantasy that Freud describes, *supra* note 59, that our lives stretch out infinitely, that we are immortal in our unconscious, in part because we cannot imagine what it would be like to be dead, is a separate question. The sense that life’s end is not on the table, not worth thinking about, seems *distinct* from the sense that one is truly immortal.

69. The first social scientist who highlighted the “identifiable victim effect,” Thomas

And now think back to our public policy problems: Isn't it plausible that seemingly excessive investment in life-saving cures rather than prevention is in significant part an effort to reduce the number of occasions in which diagnosis switches people into the state of thorough existential dread? Developing cures, or treatments that make cure seem more possible can at least frequently delay dread until it is accompanied by a level of morbidity that makes the existential dread secondary to physical pain and morbidity. This is definitely *not* to say that the trade-offs we observe *in fact* serve the more complex ends. If we look at the standard sorts of lists of life-saving government programs that note we spend far more money to save lives in some distinct contexts than in others, one does not exactly discover the hidden rationality that I seek. Instead, it appears that a variety of perceptual problems and political pathologies drive death-averting spending levels: undue attention to highly publicized and available death-causing practices⁷⁰ and interest group politics.⁷¹

Schelling, seems to be the last one who intuitively recognized (though did not pursue) the point that our reactions to identifiable victims invoke our anxieties about our own deaths and dying, as well as feelings of responsibility for the death of others (an issue we will return to in discussing anti-consequentialism), while we could face increases in statistical risk while remaining in the world in which we are largely in denial of death. *See* Schelling, *supra* note 15. Here was Schelling's basic, canonical take on the identifiable victim's effect:

Let a 6-year-old girl with brown hair need thousands of dollars for an operation that will prolong her life until Christmas and the post office will be swamped with nickels and dimes to save her. But let it be reported that without a sales tax, the hospital facilities of Massachusetts will deteriorate and cause a barely perceptible increase in preventable deaths—not many will drop a tear or reach for their checkbooks.

Id. at 129.

To evaluate an individual death requires attention to special feelings. Most of those feelings, though, involve some connection between the person who dies and the person who has the feelings; a marginal change in mortality statistics is unlikely to evoke these sentiments. Programs that affect death statistically . . . need not invoke these personal, mysterious, superstitious, emotional, or religious qualities of life and death.

Id. at 131. And here is yet another of Schelling's remarks on the difficulty of thinking about death: "Death is an awesome and indivisible event that goes but one to a customer in a single large size . . . For many people it is a low-probability event except on special occasions when the momentary likelihood becomes serious." *Id.* at 158.

Schelling was also quite aware that the reduction of anxiety over death could lead people to take steps that would increase their chances of dying. The example he gives is someone who knows a life-saving operation will become necessary at some point, and that the operation itself has some risk of being fatal. He will sometimes prefer to undergo the operation immediately; he thus "raises the stakes against himself" (lowers his life expectancy) to avoid the anxiety associated with the non-trivial, already-present risk of death in the future. *Id.* at 145. It is thus clear, once more, that he recognizes that the dread of death and the harms of death itself were separate.

70. Obviously, the reaction to Baby Jessica herself reflected not just an "identifiable victim" effect—the outpouring for her exceeded the outpouring for other identifiable victims—but the high

Why do we pour so many resources into saving those who are trapped in mines or drifting at sea? Is it in some significant part because there is scarcely anything more dreadful than the acute awareness of imminent, preventable death unaccompanied by other forms of morbidity and suffering? Or, to put the more significant health policy point more provocatively, can we explain our desire to cure rather than prevent illness (to eliminate “diagnosis shock”) but think we do not do enough to insure that most people die from heart attacks, strokes, and accidents, the sorts of deaths that occur without people having to experience a time where they are *dying*? If we could dissociate mortality and morbidity enough to do that, should we? Is the real problem of under-spending on prevention rather than cure that we can’t raise cure rates enough to make it rational to spend less on preventing the development of incurable disease, not that we spend too little on preventing death? To put that last point another way, is it possible current policy is irrational not because it does not focus exclusively on maximizing life expectancy, but because it does not focus enough on preventing lingering, foreseen deaths?⁷²

The chief problem, looked at through this lens, is that we cannot realistically hope to funnel far more deaths into sudden unexpected death, and people typically discount—or fail to imagine—the possibility that the steps they take now will put them in the position of *dying*, not just being dead, later. The problem is not that it is irrational to want to avoid life after a grim diagnoses, it is the failure to realize there are preventative measures we can take to reduce the number of such diagnoses. Conceivably, we could better funnel deaths towards “sudden death” by reducing the number of people who *develop* cancer (e.g., by reducing the presence of environmental carcinogens) than we do by developing as many cures for cancer as we are capable of developing. At the same time, though, the fact that typical public opinion poll subjects are willing to spend roughly four times as much to prevent a death from cancer as they are willing to

degree of media attention devoted to her case (and the associated extreme salience of her particular jeopardy). *See supra* notes 1-2.

71. The canonical list of selected regulations with wildly disparate costs/death averted is presented in CASS R. SUNSTEIN, *RISK AND REASON* 30-31 (2002). Generally speaking, though, the list does reveal something of the pattern I suggest in the text is rational though it may well reveal it to an irrational extent: far more is spent preventing lingering fatal illnesses—by and large—than is spent preventing sudden accidental deaths. (But the variation within the “avert lingering illness” regulations is still enormous.)

These lists are partly difficult to interpret, too, because it is unclear in each case what the marginal efficacy of increasing death-averting expenditures would be: it is not implausible, for instance, that some of the cheap death-averting programs are set at the technologically maximally feasible levels.

72. This could be done by spending more to develop either successful or plausible cures for diseases that patients will otherwise know are incurable or by spending more to prevent the development of diseases known to be incurable, rather than diseases that kill suddenly.

spend to prevent a sudden accidental death⁷³ (a pattern that seems to understate the value placed on averting cancer deaths in our actual regulations⁷⁴) strikes me as reflecting wholly rational decision processes, even without regard to the hardly trivial fact that cancer deaths will almost surely be preceded by substantial *physical* morbidity.⁷⁵

II. CONSEQUENTIALISM AND AGENT-RELATIVITY (A): CONDUCT V. CAUSE

I will discuss only briefly whether our attitudes about identifiable lives reflect a more wholesale rejection of conventional consequentialist reasoning. I am almost wholly unsympathetic toward one of the two standard forms of anti-consequentialism: the one in which subjects are morally bound above all to evaluate their own conduct and/or mental states, to determine whether the conduct itself accords due respect to the rights or interests of others rather than to evaluate the expected value impact of their action or inaction on all those affected by that action or inaction.⁷⁶

Some of us find this sort of focus on the virtue of a particular actor-in-question puzzling or infuriating. Why should I care only about what I do—or want, or know I will cause, or self-consciously or carelessly risk causing—rather than what states of the world will exist, or are more likely to exist, given alternative courses of action I might take? Isn't this just infantile, narcissistic ego-centrism masking as morality? Others find it incomprehensible to think

73. See Richard L. Revesz, *Environmental Regulation, Cost-Benefit Analysis, and the Discounting of Human Lives*, 99 COLUM. L. REV. 941, 962-74 (1999).

74. See SUNSTEIN, *supra* note 71, at 30-31. It is almost surely the case, too, that data on per death prevention costs are hard to interpret because we don't know, looking at the raw data, whether incremental spending on those risks we spend little on would be as efficacious as spending on those risks we currently spend a lot on. There is no reason to believe there is a linear relationship between spending and deaths prevented: while we might save two people at \$10 million per life saved to eliminate exposure to a carcinogen, we may save two people by spending \$1 million on avoiding an industrial accident and no more by spending \$19 million more.

75. Of course, even if it would be rational for "first parties"—those who can elect between their own sudden death and "dying" before being dead—to funnel mortality towards sudden deaths, this would not inevitably be the right social policy. Surviving family and loved ones might *gain* in significant ways if they had the chance to come to terms with the loss of the person who passes away and might thus prefer slow deaths to sudden ones. If this is right, there is a purely "utility-based" argument against funneling deaths towards sudden deaths.

76. Recall the Kantian reluctance to countenance lying about the whereabouts of a would-be killer's potential victim: all the putative liar can attend to is whether she "shows respect" for the autonomy of the person she might deceive. Lying invariably shows disrespect for another's autonomy since it contributes to that rational agent's inability to direct his own choices appropriately, with good information about the effects of his choices. The fact that the party one does not deceive may well use the information to interfere with another's autonomy, as killing him surely will, does not affect the own-conduct centric's obligations. See *supra* note 16.

about ethical behavior without staying focused on what our appropriate *obligations* might be, obligations consistent with establishing a social order giving adequate reign both to our own autonomy and our consideration for others' autonomy. Individuals are separate entities; gains to one do not necessarily counterbalance losses to another. We rightly tolerate only a certain sort of "loss" or "restriction" on our own freedom; the sorts of losses we and those around us should accept are those consistent with our *recognition* of others as similarly and equally autonomous agents. While our interests cannot in the final analysis ever be fully protected—suffering is ubiquitous; bad stuff happens; we all die at some point—we *can* live in a world in which we are all respected as equally autonomous beings. But I have absolutely nothing to add to the familiar and well-worn debate on that issue. Thus, my far more modest goal is simply to highlight that the debate over why subjects seem to care so much more about identifiable lives and whether it is irrational or immoral to do so might implicitly invoke the debate over why people focus more on what they directly do than what they either diffusely cause or "merely" allow to happen.

The preference to save "identifiable" lives doubtless has many roots. In addition to those widely discussed in the existing literature,⁷⁷ it might track the commonplace preference to avoid harmful acts and more direct harm causing by the agent herself, rather than harmful omissions thought to lead to equivalent or worse consequences through a more diffuse causal chain or only in conjunction with the conduct of others. This might be true even in cases in which we are considering failing to save either the identifiable person or an unidentified person. I am not saying that in the identifiable victim experiments, researchers have carelessly conflated cases in which subjects are comparing scenarios where they actively harm identifiable persons with situations in which they merely fail to protect unidentified persons. Nor am I claiming that the standard deontological distinctions many experimental subjects act upon are the same as the distinctions drawn between failing to rescue identifiable persons and failing to take steps that prevent the death of as-yet unidentified ones. This is true whether we are considering distinctions between acts and omissions, between rights-violating acts and non-rights-violating acts that result in identical bad effects, or between intended and merely known negative effects of conduct.⁷⁸

77. See *supra* notes 41-45.

78. These are all distinctions that might be drawn in the standard Trolley Problems. In one of the canonical Trolley Problems, generally dubbed "Fat Man" in the literature, subjects embracing standard deontological positions refuse to push a heavy person off a bridge to stop a trolley that will otherwise kill five people, though they will pull a switch to direct a trolley off-course, knowing it will kill one, but save five. In the first case, they (a) *directly* act upon the person who is killed, rather than allowing an intermediating agent, the trolley, to do all the direct dirty work; (b) they arguably violate bodily integrity *rights*—the right of the person pushed off the bridge not to be killed by another agent is purportedly distinguished, in a non-circular fashion, from the absence of

The argument I want to make instead is one I offer quite tentatively: This is true in part because I am not convinced it would prove to describe accurately subjects' reactions to distinct cases if it were appropriately experimentally tested. It is also true because even if it were descriptively accurate that subjects are responding as they do for the reasons I suggest, this would not affect my views of the *acceptability* of the identifiable victim effect. That is the case because the effects I am describing strike me as arising from the use of difficult-to-justify moral heuristics that would aptly frame morality only in a world in which harm-causing was invariably simple (X batters/kills Y) but would serve us poorly in a world in which harm is caused in a more complex fashion (X makes Y more vulnerable to a host of environmentally present toxins that might kill Y even if X had not acted as he did, though his actions hurt Y's *chances* of surviving exposure to the toxins). In our ancestors' world—when we presumably developed some of our most readily accessible heuristic intuitions—we neither had the post-agrarian technology to cause diffuse increases in the risk of harm on a regular basis, rather than to damage another identifiable person directly, nor did we understand in any sophisticated ways probabilistic epidemiological inference.⁷⁹

a right not to be hit by a diverted trolley since neither the five victims nor the single victim on the side-track have any trumping right to be free from accident; (c) they arguably create a novel risk rather than merely redistributing something like a pre-existing risk; and (d) it is their *purpose* that the victim suffer injury or death though their *motive* might be to save others. Similarly, killing necessarily temporally precedes saving in the “Fat Man” case but comes after the multiple victims have been saved in the case in which the trolley is diverted. Thus, in the “divert-the-trolley” case, it is at least arguably merely a known side-effect of their direct decision to save the five in jeopardy that someone else will die. To put the double effects issue in more conventional Kantian terms, the person who pushes the Fat Man over the bridge “uses” him as a means to save the others while the person who merely diverts the Trolley onto a track where a person happens to be would be thrilled if no victim were there and (arguably) does not use him in any ways to help others: it is using someone as a means that demonstrates the particularly troublesome form of disrespect.

For some, of many, classic deontological defenses of drawing a distinction between these cases, see, for example, Michael J. Costa, *The Trolley Problem Revisited*, 24 S. J. Phil. 437 (1986); F.M. Kamm, *Harming Some To Save Others*, 57 PHIL. STUD. 227 (1989); John Mikhail, *Universal Moral Grammar: Theory, Evidence, and the Future*, 11 TRENDS COGNITIVE SCI. 143 (2007); James A. Montmarquet, *On Doing Good: The Right and the Wrong Way*, 79 J. PHIL. 439 (1982); and Judith J. Thomson, *Killing, Letting Die, and the Trolley Problem*, 59 THE MONIST 204 (1976).

79. Cass Sunstein has forcefully made the argument that we frequently use problematic moral heuristics. See generally, e.g., Cass Sunstein, *Hazardous Heuristics*, 70 U. CHI. L. REV. 751 (2003); Cass Sunstein, *Moral Heuristics*, 28 BEHAV. & BRAIN SCI. 531 (2005); Cass Sunstein, *Moral Heuristics and Moral Framing*, 88 MINN. L. REV. 1556 (2004).

Sunstein self-consciously analogizes the “moral heuristics” he describes to the often-mistaken heuristics H&B theorists argue that people use in making factual judgments, especially judgments about the probability that certain events will occur. He is also fairly explicit that the sorts of framing/elicitation effects that H&B theorists have argued are in play when people evaluate end-

Here, though, is the descriptive hypothesis that I want to suggest is plausible. People might intuitively conflate judgments of identifiability in situations in

states generally are operative when we try to elicit considered moral judgments, which are, in this respect, merely a particular variety of evaluative judgment.

In each sort of case in which a subject makes use of a heuristic, the agent has some limited set of goals—to assess the probability of an outcome, to judge the permissibility of certain conduct—and the rule of thumb will be “accurate enough” to assess probability or to make the judgment about permissibility most of the time. Thus, the agent substitutes the “heuristic attribute” for the true “target attribute” in making a judgment. What can often be problematic about the use of heuristics is that these heuristics are, like any rule, inapt to the full range of situations to which they may apply. Thus, to return to the canonical case of judgment bias, the availability heuristic, it indeed is *usually* the case that one will judge probability accurately if one follows the “rule” that events that are readily available to memory have occurred more frequently than unavailable events, but sometimes one will not because one occasionally recalls events because they are salient rather than frequently encountered. Similarly, in the realm of “moral heuristics” it will usually be the case, for instance, that omissions are less culpable than commissions. For instance, those who omit to take steps are less likely to intend harm and intending harm is relevant; one cannot discern whether or not omissions are deliberate or not and placing blame without “proof” is a poor idea; or because when X omits to save Y, some Z or Zs may still do so while when X aggresses against Y, harm is more certain. Sometimes, though, omissions are not less culpable. *See, e.g.,* Sunstein, *Moral Heuristics, supra*, at 540.

Sunstein is not especially clear whether he thinks of the moral heuristics as consciously adopted rules designed to meet known ends; whether he believes instead that they are essentially the sort of cognitive routines that we develop because in trying consciously to meet a particular end over a range of situations, we subconsciously develop a habit of substituting one or a small number of attributes that we see often in analyzing a situation for a fuller analysis of the situation; or whether individuals are predisposed to process the simpler heuristic cues, completely unaware that they might be predisposed to do so because, somewhere in our evolutionary history, processing these simple cues was sufficient to meet our ends in most “similar” situations. Another way of putting this point is that he is not clear whether people consciously know the “target attribute” at all, and if they do, whether they are consciously aware that they are substituting a “heuristic attribute” for the target attribute because they know it is easier to do so.

Thus, for instance, it is not enormously clear whether Sunstein believes that those who distinguish omissions from commissions ever think that, say, discerning the intention of the party whose course of conduct they are evaluating is their true purpose. It is likewise ambiguous whether he thinks that they did so before developing a conscious rule of thumb that they would merely ascertain whether they were dealing with an act or failure to act or whether the “rule of thumb” simply developed as an increasingly automatic reaction because in a large number of situations in which they were attempting to discern intention, they unconsciously noted that they so rarely found it in omissions cases that the simpler search for an act became a habit. It seems that he is more drawn to a different story: People have no conscious idea of why they distinguish omissions from commissions. Doing so is not a method of meeting a purpose (e.g., to ferret out intentional actors) that they have consciously developed nor does doing so remain as a residue of having gone through many iterations of trying to ascertain intent and settling on an economical way of getting there. Rather, it is a “rule” whose connection to its original purpose is unrecoverable.

which we merely fail to act to prevent a death with judgments about causation and, in a related fashion, judgments about the presence or absence of independent conduct. This is true of both situations in which we are failing to save and situations in which we are failing to avert the development of peril. More specifically, they might conflate judgments of identifiability with judgments about whether causation is “relatively direct” rather than “relatively diffuse” and with judgments about whether other actors are more or less likely to be seen as “superseding,” not necessarily because they are temporally causal interveners, but because their causal roles are thought of as more central. The familiar way in which one form of “identifiability” matters is that we find it more natural to think of omissions as causes when we can *single out* and thus identify the actor-in-question from some mass of actors who are either equally able or equally duty-bound to take affirmative steps to prevent peril or to rescue. In one view, duty or obligation (which might flow simply from being uniquely situated to avert harm but more readily flows from status or contract) is primary. Once we decide that a party is independently duty-bound to avert a harm, only then do we attribute causal power to that party, and only then with some discomfort.

But it is also possible that, at least to an extent, making intuitive judgments about which sorts of parties are the more direct cause of an outcome pushes us away from consequentialism (a focus on the net result of any course of action or inaction). It pushes us towards a focus on actor-centered moralism, focusing merely on avoiding taking steps that more directly cause harms. If, though, it is not the judgment that the actor-in-question owes some prior duty to the victim driving the conclusion that he is the “cause,” but identifiability or relative uniqueness of at least some actors in the story that drives the judgment that the actor-in-question is the cause, then *victim identifiability* may increase the sense that the actor-in-question causes harm as well.

In this view, the “prototype” causal story has both readily identifiable perpetrators and identifiable victims: X shoots Y, and Y dies. The claim I am making here is that causal/conduct salience can be preserved in part either by singling out an X *or* a Y. When neither is singled out, and all we observe is an increase in the number of dead or injured Ys, and some diffuse group of Xs who might have stopped those bad outcomes from happening, conduct/cause-centered anti-consequentialists will be least moved to condemn.

Judgments that omissions are “conduct” and can “cause” results are made most readily in situations in which the actor-in-question is singled out from the mass of potential life-savers by the fact that he has an independent duty to prevent harm. But I think most people would believe that an omission is also more conduct-like and more apt to be thought of as causally relevant if the party has a unique or at least distinguishable *identity* as more readily able to avert harm than unnamed others, even in the absence of duty. The first distinction—between those with duties and those without—is raised clearly in a hypothetical offered by

Patricia Smith, who implies that our intuitions about cause are not strongly parasitic on prior judgments of duty. She implies that in the context of arguing that omissions are never comfortably considered causes:

Suppose that a train switchman, Charlie, and his buddy, Frank, are about to watch a playoff basketball game when Charlie says to Frank, "I have to pull that switch at 9:02. Don't let me forget or the L&W will crash straight into the Boston Flyer." The game is riveting, and at 9:03, the L&W crashes straight into the Boston Flyer. Charlie is responsible for the crash But what should be said about cause?

One possibility is that Charlie didn't really cause the wreck So we just say he caused it because we want to hold him responsible

So, the second possibility is that Charlie did cause the crash But if that is admitted, then it should be noticed that there is no causal difference between Charlie and Frank.⁸⁰

To the degree that observers are agent-centric only in the sense that they focus solely on breaches of legal duties or failure to respect the rights of others, not on conduct or cause, then (1) what distinguishes Charlie from all other persons is simply and exclusively that he alone has a duty to save (in this case established by contractual assumption of the duty to save) and that he has breached the rights of those who benefited from that contractual assumption of a duty to care and (2) *no* legal stranger's failure to save—whether it is a failure to save an identifiable or non-identifiable victim—is problematic. But to the degree that observers distinguish Charlie from his friend Frank in part because they single out or identify him as a unique perpetrator (active and causal) and that identifying him makes us especially revolted by his conduct, we would also expect that his friend Frank would seem more active, causally relevant, and therefore culpable than the typical person: Frank is more readily able to stop the wreck without much change in his life plans than one of an unidentified mass of persons who might have learned about the possibility of the wreck but failed to take similar steps to go out of her way to avert it.⁸¹

Assume then that we think Frank is more culpable than unidentifiable persons who might have the same opportunity to stop the disaster. The question, at that point, becomes whether the classic unidentified bystander who must go out of his way to save seems more active or causally relevant if he knows whom he might save by taking steps than if the people to be saved are also abstract and unidentified. Do we single out a bystander from a world of bystanders simply because they have been entwined in a particular relationship by *knowing* precisely whom they might save? So we are all bystanders to, say, a failure to install barriers or warning signs that might save some kids in general from

80. Patricia Smith, *Legal Liability and Criminal Omissions*, 5 BUFF. CRIM. L. REV. 69, 101-02 (2001).

81. For instance, any one of many people who walks by a switchman's office and sees the switchman has had a fatal heart attack and is not able to deal with the trains.

drowning, but if one of us knows the identity of a particular child we could save from drowning, does he—by virtue of that fact alone—become distinct from the rest of us? In becoming distinct, does he get thought of as more causal and more directly active in the death if the child indeed drowns?

It is conceivable that we could test this account experimentally. If my point is right, we would expect experimental subjects to differentiate between Neighbor 1's failure to install a fence around his pool, which might protect unnamed kids in the neighborhood, from some Neighbor 2 who fails to build the fence when he is or could be concerned with a particular identified neighborhood toddler, even holding constant the probabilities that a child will drown. I am fairly certain that the ultimate moral judgments would be different—that is merely a restatement of the identifiable-victims effect, in the prospective harm, rather than rescue, case. I am radically less confident that one could tease out whether they are different *because* subjects construct Neighbor 2 either as more active or causally direct. One would need to ascertain whether they are more likely to use the word “kill” instead of the phrase “fail to save” in Neighbor 2's case *and* that using that word *drives* their judgment, rather than being the word they use because they have made a harsh moral judgment. To test for “causal salience,” one would also have to test whether they are less likely to recognize the causal contributions of others—the drowning child herself, that child's caretakers, etc.—when the jeopardized child is named.

As I said, though, I would be unmoved normatively by learning that Neighbor 2 is more frequently harshly judged by experimental subjects: conduct-centric anti-consequentialism is no more compelling if it is intuitive or easily mentally accessible.⁸² (Unless we believe that our moral reactions are so

82. For a clear version of the familiar argument that intuitions are of little moment since they are framed in response to evolutionary pressures that may now seem irrelevant, see, for example, Peter Singer, *Ethics and Intuition*, 9 J. ETHICS 331, 331-51 (2005). The following quote expresses a quite commonplace reaction relevant both to Trolley Cases and cases in which we seem more forgiving of those who “kill” by raising global death rates:

For most of our evolutionary history, human beings have lived in small groups In these groups, violence could only be inflicted in an up-close and personal way—by hitting, pushing, strangling, or using a stick or stone as a club. To deal with these situation, we have developed immediate, emotionally based responses The thought of pushing the stranger off the footbridge elicits these emotional responses. Throwing a switch that diverts a train that will hit someone bears no resemblance to anything likely to have happened in the circumstances in which we and our ancestors lived . . . the standard trolley case describes a way of bringing about someone's death that has only been possible in the last century or two, a time too short to have any impact on our inherited patterns of emotional response.

Id. at 347-48.

Singer's argument that we can condemn certain moral arguments by learning of their provenance, and more particularly by discovering that they developed as useful rules of thumb in

modularized that we simply *cannot* subject them to scrutiny and self-critical reflection, it is not clear that we should make much of initial intuitions or ease of cognition.) And, of course, as Sunstein and Vermeule noted, collective bodies cannot use a conduct-centric moral metric to judge policy in any case: *everything* the state does or fails to do is, at core, an action taken by the relevant “agent” relative to some alternative policy.⁸³

III. CONSEQUENTIALISM AND AGENT-RELATIVITY (B): PARTIALITY

People may believe they “know” identifiable people to some limited extent. They may be more drawn to saving these identifiable people than they are drawn to saving unidentified people for some of the same reasons that they are drawn towards saving those whom they know *best*, like most, and care most deeply about (such as kin and friends) rather than identifiable persons who they like less, feel less attached to, and spend little time thinking about. Consider the canonical case: a parent manifests what many see as a legitimate preference to save his own child rather than to save a larger number of strangers. The purest act-utilitarian may believe that, at least at first blush, the parent is acting improperly. Assuming that each stranger’s life and his child’s life are of equal intrinsic value, he is taking a step that diminishes overall utility. Most respond to the example either by noting that “pure” utilitarianism is either too stringent a moral code or that it is a defective one.

We expect two conventional utilitarian responses to the accusations that utilitarianism cannot cope with the legitimacy of this sort of “partiality.” Partiality can be understood as valuing the interests of some more than the interests of others without denigrating their abstract moral worth or the claim that their interests, as a general matter, should be considered no less significant. First, we might expect the response that the anti-utilitarians are misconstruing the *domain* of utilitarianism. It is arguably not meant to govern all personal morality, but merely public policy-making. It may be right for me to care more about my kids than I care for people generally, but for the state, it is best to presume that everyone is equally special to someone.⁸⁴

Second, we expect utilitarian theorists to argue that *nurturing* this sort of agent-relative partiality has beneficial consequences. If we demanded that each party act as an impersonal consequentialist in each of her decisions, it would

an environment in which a different set of moral problems typically arose, has been sharply criticized. See Selim Berker, *The Normative Insignificance of Neuroscience*, 37 PHIL. & PUB. AFF. 293, 319-21 (2009).

83. See Sunstein & Vermeule, *supra* note 7, at 705, 707, 719-24.

84. An argument that utilitarianism was historically meant as a guide to policy-making, rather than as a more inclusive set of ethical rules, is made in Guyora Binder & Nicholas J. Smith, *Framed: Utilitarianism and Punishment of the Innocent*, 32 RUTGERS L.J. 115 (2001).

weaken utility-promoting institutions like families. The argument takes two subtly distinct, though not mutually exclusive, forms. One could argue that if we nurture deep, utility-maximizing ties to our family members, we will simply be stuck with partiality in the life-saving decision. As a matter of psychological fact, then, we cannot expect people encouraged to form special bonds to ignore them when it might maximize utility on a particular occasion to do so. Alternatively, one could argue that partiality-in-saving is not just an *inevitable* side effect once we nourish a helpful institution, but also that encouraging such partiality affirmatively fosters the utility-increasing institution.

These arguments, of course, resonate in typical rule-utilitarian—as opposed to act-utilitarian—understandings of the dictates of utilitarian philosophy.⁸⁵ It might be good to abjure an act-utilitarian decision on a particular occasion if following a rule that lowers utility on the particular occasion will increase utility overall. Thus, we may promote *rules* such as “care most for your kids even when doing so will result in a loss of utility” or “don’t take others’ property even when you value it more highly than they do” because doing so will solidify crucial institutions like the family or secure private property. Obviously, this is not one of those situations in which the most powerful argument for *rule* utilitarianism is present—where there are utility gains to be had simply from following a practice without exception because it is “costly” for people either to decide when to make or to deal with others who make exceptions. Thus, in the canonical case, the rule utilitarian could argue it might well be worthwhile to follow a rule that one will tell the truth *always*, even if lying on a particular occasion would seem to increase utility, because if people know that all that they hear is true, at least as best as the speaker knows, it will increase overall utility. One might say that it would increase the utility-promoting institution of “trust,” or, to put the point in more conventional economic terms, it will reduce the costs of information-acquisition because one will never have to expend resources to verify that what others tell you reflects their actual beliefs. Following the “rule” that one should save those one is partial to rather than figure out on each occasion whether that is utility-maximizing does not increase utility by virtue of its being a “rule” (i.e., easily followed by the decision maker herself, or followed without exceptions that force others to absorb the costs of the presence of exceptions, etc.). But it might be a good rule to follow because doing so promotes other good attitudes or because it is the only rule we can realistically follow if we have developed the attitudes that utility-generating institutions promote.

What is crucial to highlight in this regard is that one of the most robust findings of the literature on the identifiable victim effect is that it is fragile in a

85. For some seminal, standard accounts of the distinction between act and rule utilitarianism, see, for example, John C. Harsanyi, *Rule Utilitarianism and Decision Theory*, 11 *ERKENNTNIS* 25 (1977); John Rawls, *Two Concepts of Rules*, 64 *PHIL. REV.* 3 (1955); and J.J.C. Smart, *Extreme and Restricted Utilitarianism*, 6 *PHIL. Q.* 344 (1956).

very particular way. When people are told that people generally exhibit it, they *will* stop spending *more* to save or aid identifiable victims than they will spend to save or aid statistical lives. But the sad truth is that they do so not by increasing their altruistic outpourings towards statistical lives but by cutting back on their generosity to identifiable victims.⁸⁶ One might, in a precisely parallel fashion, believe that if parents are directed to show no more concern for their drowning kids than for others, they will show little or no concern for anyone. The permission of partiality frees up the manifestation of a utility-increasing practice that would simply not exist if we demanded a less partial, more neutral practice.

Altruistic concern may not depend *directly* on identifiability so much as it depends on the putative altruist having a strong affective reaction to the victim's suffering. Generally, we are more prone to have affective reactions when we react to named individuals. When we deliberate more, we lose affect as we see that we are over-reacting to a person who is merely representative of a larger group. But the claim here is not the rather depressing one that deliberation always breeds callousness. It might well be the case, for instance, that our typical low-affect reactions to distant strangers (whether they be distant physically, ethnically distinct, or socioeconomically) may be emotionally intensified when we deliberate more, and see the plight of these strangers as more compelling.⁸⁷ More generally, while high cognitive load—which reduces deliberation and increases the significance of affect in decision making—makes both self-identified political liberals and conservatives less prone to be generous in offering subsidies to those AIDS patients who they believe are more responsible for being sick than they offer those they believe are not responsible, liberals, but not conservatives, who deliberate more because under less cognitive load are just as generous to those they judge as blameworthy in acquiring the disease.⁸⁸

Still, what seems fairly clear is that we may purchase a certain version of rational consistency—equality of spending on preventing statistical deaths and saving identifiable lives—only by driving altruistic spending levels towards zero. If doing so would decrease aggregate utility—as it would if the utility costs of altruistic actions are lower than the utility gains to the objects of altruism—then merely wiping out the distinction in the ways we treat identifiable and statistical victims might not be a good thing. Whether we should try, though, to justify non-consequentialist judgments solely by reference to their consequences is another very thorny question. But it is not a question that I have any novel thoughts

86. See Deborah A. Small, George Loewenstein & Paul Slovic, *Sympathy and Callousness: The Impact of Deliberative Thought on Donations to Identifiable and Statistical Victims*, 102 *ORG. BEHAV. & HUM. DECISION MAKING PROCESSES* 143 (2007).

87. *Id.* at 151.

88. See Linda J. Skitka et al., *Dispositions, Scripts, or Motivated Correction? Understanding Ideological Differences in Explanations for Social Problems*, 83 *J. PERSONALITY & SOCIAL PSYCHOL.* 470 (2002).

about: I merely thought it important to note that one could imagine justifying a seemingly non-consequentialist partiality towards the identified in broadly rule-utilitarian terms.

CONCLUSION

People typically manifest a willingness to pay a good deal more to save an identifiable person in jeopardy than to prevent the deaths of persons not yet known. It is possible that, in so doing, they are failing to meet their considered ends. Perhaps our only goal is to achieve the best end states or consequences that we can, and the only end-state we should seek is higher life expectancy.

While it is important to recognize that people do not always rationally evaluate whether the steps they choose to take best meet their ends, it is also important to recognize that we, as observers, may both improperly assess what their ends are and may also underestimate the degree to which they resist making judgments that are exclusively sensitive to the consequences of their conduct. These two general observations are critical in thinking about whether the “identifiable victim” effect is unambiguously troubling, representing a clear breakdown in rational decision making.

It is not at all clear that people do or should seek to maximize their own or others’ life expectancy, even setting aside issues of morbidity and physiological life quality. Death is likely a bad consequence (in very complicated ways), but *knowing one is dying* is *different* hedonically (and arguably a good deal worse). It might be fully rational to trade additional years of life for a reduction in periods in which the dread of death is pronounced; preferring cures and rescues over prevention or preferring accidental deaths over cancer deaths may imperfectly manifest a rational preference to avoid acute death-dreading periods of life, rather than to maximize life expectancy. While it is certainly possible that we miscalculate the life-expectancy-extending efficacy of cure compared to prevention, miscalculation might not be all that drives the preference for cure and rescue.

At the same time, decisions either to save or prevent the imperilment of identifiable victims might be grounded in two forms of agent-relativity. We may simply not be as concerned with any sort of consequences—reduction of mortality, reduction of the experience of death dread—as we are concerned with either the appropriateness of our own conduct or with protecting a subset of persons to whom we are more connected. I expressed both descriptive and normative hesitations about the first form of agent-relativity. I am not at all sure that deontologists in fact believe that they are something more akin to rights violators or that they have failed to manifest appropriate respect for other autonomous agents only when they fail to save those they can identify; even if they do, I argue that the belief is grounded in an inapt moral heuristic. I am more sympathetic to the idea that in the absence of a social rule or practice that we are

entitled to attend more to the identified, we will simply reduce altruism excessively so that partiality is justified as a practice, even if any particular act of partiality seems to violate precepts against maximizing utility.

We may indeed mistakenly and irrationally overvalue rescues compared to prevention and identifiable lives to statistical ones. Given the complexity of attitudes both about the distinction between death and dying and the proper scope of consequentialist ethics, though, I am reluctant to conclude so readily that what we are observing is straightforward confusion and error.

