

Draft, January, 2015

Please do not cite to or circulate without permission of the author

### **Facing Up To Risk**

Barbara H. Fried  
William W. and Gertrude H. Saunders Professor of Law  
Stanford University  
bfried@stanford.edu<sup>1</sup>

In 1972, a new Ford Pinto stalled in the middle of a highway, causing a car to rear-end it. The force of the collision pushed the Pinto's rear-mounted gas tank forward and into the differential housing, puncturing it. Fuel spraying from the punctured tank ignited inside the car, engulfing the car in flames. One of the two occupants died from the resulting burns. The other, a 13-year-old boy, was severely disfigured for life, requiring multiple skin grafts over the next decade.

In the lawsuit the heirs and survivors subsequently filed against Ford for defective product design, it emerged in pretrial discovery that before the Pinto ever went into production, Ford knew about the danger that a rear-end collision would explode the gas tank. Indeed, its own crash tests revealed the engine was vulnerable to puncture and explosion at crash speeds as low as 20 MPH. Internal memos from Ford engineers to top management estimated that at a cost of \$5 to \$11 per car, the Pinto could be redesigned to decrease the probability of explosions significantly, saving an expected 180 lives. Ford management, estimating that the redesign would

---

<sup>1</sup> Earlier versions of this paper were presented at the Center for Ethics and Public Affairs at Tulane University, University of Toronto Law School, Yale Legal Theory Workshop, American Studies Workshop at Princeton, Stanford Law School, and the University of Amsterdam. I am grateful to the workshop participants for their very helpful comments.

cost the company a total of \$137 million but save it only \$49.5 million in projected damage awards from the incremental deaths, injuries, and property damage caused by the design defect, declined to make the change.

The jury found against Ford, awarding the plaintiffs roughly \$3 million in compensatory damages, and the boy an additional \$125 million in punitive damages, an amount subsequently reduced on appeal.<sup>2</sup>

In the years since, Ford's conduct in the Pinto case has come to stand for all that is wrong with cost/benefit analysis, the decision procedure that Ford allegedly employed in deciding not to redesign the Pinto. While the 'received' account of Ford's behavior continues to be disputed, taking that account to be more or less accurate, one can easily understand the outrage it triggered.<sup>3</sup> Even on a straight cost/benefit analysis, Ford almost certainly made the wrong decision in not redesigning the car.<sup>4</sup> The fact that the Pinto was significantly less safe than most other cars on the road was probably sufficient to make its behavior legally negligent even without a faulty cost/benefit analysis. The crash engineers at Ford strongly urged redesign, but were overruled by top management, which was notoriously indifferent to safety concerns. Finally, Ford knew that federal safety regulations then pending would require it to protect the integrity of the gas tank in crashes up to 30 MPH, but chose to wait until the regulations were adopted to comply with the standards. Add to all of that the fact that

---

<sup>2</sup> Grimshaw v. Ford Motor Co., 119 Cal.App.3d 757, 808 (1981), quoting Dawes v. Superior Court 111 Cal.App.3d 82, 88 (1980) hg. den.

<sup>3</sup> The best-known of these revisionist accounts is Gary T. Schwartz, "The Myth of the Ford Pinto Case," Rutgers Law Review 43 (199): 1013

<sup>4</sup> See, e.g.,

<http://www.engineering.com/Library/ArticlesPage/tabid/85/articleType/ArticleView/articleId/166/Ford-Pinto.aspx>; Douglas Birsch, "Product Safety, Cost Benefit Analysis, and the Ford Pinto Case," in *The Ford Pinto Case: A Study in Applied Ethics* (Birsch and Fiedler, eds., 1994)

Ford withheld information about the risks for fear of bad publicity, and you have a poster child for bad corporate actors.

But the moral that many have drawn from the Ford Pinto case goes well beyond the particular facts of the case. Put simply, it is that trading off lives for any amount of money or other lesser 'goods' is wrong. As one commentator put it, "It is difficult to understand how a price can be put on saving a human life.... [I]t seems unethical to determine that people should be allowed to die or be seriously injured because it would cost too much to prevent it."<sup>5</sup>

In one form or another, that sentiment has been a cornerstone of Anglo-American moral philosophy over the past half-century. Its intuitive appeal is obvious. But it cannot guide action in a world of indeterminate consequences, which is to say, in the world in which we live. This article is about why that is so, why nonaggregative principles nevertheless continue to be offered as a plausible alternative to some form of aggregation, and the social costs when those same nonaggregative instincts are let loose in the policy arena.

I use "nonaggregation" here and throughout the paper to refer to any decision rule that evaluates our conduct at least in part without regard to the numbers of people potentially at risk or the probability of harm occurring. To the extent that nonaggregationists hold hybrid views, for current purposes the only part of interest to me is the portion that is ostensibly governed by nonaggregative rules. The rest has

---

<sup>5</sup> <http://www.wfu.edu/~palmitar/Law&Valuation/Papers/1999/Leggett-pinto.html>. Or as another commentator put it more provocatively, "Essentially, Ford argued before the government that it would be cheaper just to let their customers burn!" <http://www.engineering.com/Library/ArticlesPage/tabid/85/articleType/ArticleView/articleId/166/Ford-Pinto.aspx>

already been ceded to aggregation.

### **A. Facing Up to Scarcity.**

When Rawls wrote *A Theory of Justice*, utilitarianism was the game to beat in political philosophy, and Rawls made no bones about his intention to beat it. In his words, “Justice as fairness... offers, I believe, an alternative to the utilitarian view which has for so long held the preeminent place in our moral philosophy. I have tried to present the theory of justice as a viable systematic doctrine so that the idea of maximizing the good [meaning aggregate welfare] does not hold sway by default.”<sup>6</sup>

Three years after the publication of *A Theory of Justice*, Robert Nozick’s *Anarchy, State and Utopia* appeared, reinvigorating libertarianism as a serious contender in academic philosophy. For all their political disagreements, Rawlsians and libertarians are united in their opposition to utilitarianism, and in particular to its methodological commitment to allow “the hardships of some [to be] offset by the greater good of others” through some form of aggregation.<sup>7</sup> (For ease of exposition, I will use the term “deontology” to refer generally to all principles that take an individual’s right to be free from serious harm as presumptively trumping any amount of qualitatively lesser benefits for others. For my purposes, the differences among these principles are either indeterminate or immaterial.)

For Rawlsians, opposition to aggregation is grounded in the Kantian principle that “human beings have dignity and not mere price,” and that the ideal Kantian

---

<sup>6</sup> *A Theory of Justice* (1999), p. 513.

<sup>7</sup> In his foreword to the 2013 edition of *Anarchy, State and Utopia*, Thomas Nagel traces the common origins and common assumptions that bound Nozick’s version of libertarianism to the Rawlsian strain of liberalism unfolding contemporaneously. For further discussion of this issue, see B. Fried, “The Unwritten Theory of Justice: Rawlsian liberalism versus libertarianism,” in *The Blackwell Companion to Rawls* (Jon Mandle and David Reidy, eds., 2013).

legislator “unconditionally attribute[s] a worth to persons that cannot be quantified and is not subject to trade-offs.”<sup>8</sup> As Jay Wallace put it, “[I]t is virtually a paradigm of moral thought to allow the grave objections of a single individual to defeat a project or plan that would otherwise be expedient for many; a willingness to defer to the standpoint of individuals in such cases seems to be required by their standing as ends in themselves (as Kantians would put the point).”<sup>9</sup> For classic liberals or libertarians, opposition to aggregation is grounded in a stringent version of rights that presumptively trumps any benefits others would realize from curtailing them. To quote the most famous contemporary articulation of the view, “Individuals have rights and there are things no person or group may do to them (without violating their rights).”<sup>10</sup>

It is not hard to understand the appeal of that paradigmatic moral thought. When harnessed on behalf of something other than narrow self-interest, it has inspired much good in the world. But in a world of scarcity, it cannot serve as a general guide to action, because it brings the wrong tools to the job. I use the word ‘scarcity’ not in its conventional sense-- to denote some absolute level of deprivation-- but in the sense in which economists would use it: to denote any situation in which the demand for a ‘good’ exceeds its supply, with the consequence that we cannot satisfy all competing demands for it.<sup>11</sup> In this sense of ‘scarce,’ most goods in society are necessarily scarce,

---

<sup>8</sup> Thomas E. Hill, Jr., “A Kantian Perspective on Moral Rules,” in *Philosophical Perspectives*, 6, Ethics, 285-294 (1992), 292.

<sup>9</sup> Jay Wallace, “Scanlon’s Contractualism,” *Ethics*, 112 (3): 429-470 (2002), at .

<sup>10</sup> *Anarchy, State and Utopia*, Preface.

<sup>11</sup> More precisely, it denotes any situation where, if a good were free, people would consume more of it than is available. This qualification is meant to set aside the market solution to scarcity: raise the price of a good until demand no longer exceeds supply.

either because the material resources needed to produce them are finite (e.g., 'goods' like health, product safety, national defense) or because one person's pursuit of his or her life projects necessarily compromises others' (e.g., activities that impose some risk of harm on others). If a good is not scarce in this sense (breathing the air around us, thinking our private thoughts), in a liberal society there is no reason for government to get involved at all.

In the face of scarcity, virtually all choices require us to trade off one person's interests against others'. That is the universal tragedy of social life. The essentially optimistic premise on which nonaggregative principles rest-- that tragic choices between the deepest interests of individuals are the exception and not the rule--cannot tell us what to do about it.

To be clear, that claim does not extend to interests we decline (on deontological grounds) to count at all—e.g., the interests of criminals in pursuing most criminal conduct (killings, rape, theft, fraud, vigilante justice, etc.); the interests of noncriminal actors that threaten the fundamental values of a society (antidiscrimination norms, free speech, the right to a fair trial, etc.). Precisely because we choose not to count the interests on one side at all, we do not need to aggregate conflicting interests to come up with a decision rule.<sup>12</sup> My concern here is with the domain of potentially harmful conduct in which we take there to be prima facie legitimate interests at stake on both side. The discussion that follows should be understood to be limited to such conduct.

---

<sup>12</sup> I am setting to the side some deep questions about whether the scope of the rights we protect against utilitarian objections is itself determined by rule-utilitarianism. For present purposes, I am happy to stipulate that some fundamental norms in society can be explained and defended only in nonconsequentialist terms. In addition, every comprehensive moral theory, including utilitarianism, is vulnerable to the critique that it smuggles in other norms at critical junctures.

To elucidate the central argument here, I borrow Philip Tetlock's useful psychological taxonomy of tradeoffs. Tetlock divides peoples' interests into sacred and nonsacred values, which correspond (roughly) to the distinction in the nonconsequentialist literature between fundamental and nonfundamental interests. He identifies three different perceived categories of tradeoffs, based on the interests implicated on each side: routine tradeoffs, taboo tradeoffs, and tragic tradeoffs.<sup>13</sup>

*Routine tradeoffs* are tradeoffs between two non-sacred values. Crucially, in the nonconsequentialist literature they are taken to include not just tradeoffs in which the underlying interests at stake on each side are unimportant (e.g., allocating tickets to an oversubscribed event) but also tradeoffs in which sacred values are implicated on one or both sides, but the probability of harm to those values is low (e.g., fly paths that cross over residential neighborhoods, putting the population at risk in the extremely improbable event of a crash). Most nonconsequentialists have ceded routine tradeoffs to aggregation, either because they have concluded the choice among the interests at stake is a 'non-moral' one, and hence can be handled by any noninvidious decision rule (including aggregation), or because they have reluctantly concluded that, in the case of risk, there is no alternative to aggregation.

At the other extreme are *taboo tradeoffs*, which are tradeoffs between sacred and nonsacred values. Taboo tradeoffs are the principal redoubt of nonconsequentialism. They describe the realm in which interpersonal tradeoffs are forbidden on deontological grounds, because the sacred value(s) on one side always trump the

---

<sup>13</sup> Tetlock et al., "The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals," *Journal of Personality and Social Psychology* (2002).

nonsacred value(s) on the other, irrespective of the numbers involved. Scanlon's 'Greater Harms Trump Lesser Harms' principle targets one set of taboo tradeoffs; libertarian property rights as an absolute side constraint targets another. Nonconsequentialist principles have a significant role to play in determining principles to govern potentially harmful conduct only if the domain of taboo tradeoffs is significant.

In the middle, between routine tradeoffs and taboo tradeoffs, are *tragic* tradeoffs: tradeoffs between two sacred values (a life v. a life, a life v. ten lives, a life v. a lifetime of pain, etc.). May we (must we?) kill one to save a hundred? May we bomb enemy targets to protect innocent civilians, knowing that some number of other innocent civilians will be collateral damage? May we (must we?) save one person from certain death, at the cost of a lifetime of pain for a million? May we prefer family or friends to strangers? Known victims to unknown ones?

As T.E. Hill has commented, cases like these are "notoriously hard" to resolve by appeal to the principle that "one must not trade or sacrifice the dignity of anyone for 'more dignity' for others," because *neither* of our available choices "fully respect[s] the dignity of all."<sup>14</sup> Consistent with that view, a substantial portion of the nonconsequentialist literature over the past forty years has been given over to figuring out if these one-off tragic choices can (or should) be resolved by nonaggregative principles, or instead must (or should) be turned over to some form of aggregation. Dedicated trolleyologists have tried to keep tragic tradeoffs in the nonconsequentialist fold by deploying a secondary arsenal of nonconsequentialist distinctions (e.g., doing v.

---

<sup>14</sup> Hill, "A Kantian Perspective on Moral Rules" at 293.

allowing harm, foreseeing v. intending harm) to break the tie. Others, most notably Scanlon, have argued that if the sacred values arrayed on both sides are commensurate in importance, the numbers should count as a tiebreaker, relegating the tradeoff to aggregation. (If they are not deemed commensurate, then they are taboo tradeoffs).

My concern here is not with how different nonconsequentialists resolve tragic tradeoffs, but rather with the negative pregnant that justifies the lavish attention they have received over the past fifty years: Most tradeoffs are either taboo tradeoffs or routine tradeoffs, and nonconsequentialists already know how to handle both of (in the former category we may aggregate and in the latter we may not). That leaves only the small, impossibly difficult, category of tragic tradeoffs to resolve.

This view of the moral terrain, I argue, gets things exactly backwards. The high-stakes, hypothetical ‘tragic tradeoffs’ that populate the literature may be difficult to resolve without resorting to aggregation. But under any realistic assumptions about our ex ante state of knowledge, the tradeoffs we actually face—whether classified as taboo or routine-- are impossible to resolve without resorting to aggregation. As a result, far from offering a comprehensive alternative to aggregation, nonaggregative principles can operate only at the margins of real-world choices we face—indeed, arguably outside the margins entirely—relegating the vast literature on trolleyology to a luxury indulgence of our nonaggregationist instincts.

Thus, far from offering an “alternative, systematic account of justice” to replace aggregation, nonaggregative principles are viable in the minimal sense of operationalizable only as long as they are confined to a very small and rarified set of tragic choices. And they will *appear* to present a systematic alternative to aggregation

only as long as attention is confined to that set. That describes virtually all of the vast nonconsequentialist literature over the past fifty years on the problem of harm to others.

## **B. The Moral Terrain of Trolleyology.**

The typical hypothetical used to develop and test our moral intuitions about tragic choices in the Kantian literature goes something like this:

We have a limited supply of a certain vaccine to treat a rare and usually fatal disease afflicting eleven people in the country. One of those eleven has a very acute form of the disease and will die without the entire vial of vaccine. Each of the other ten has a more mild form, which can be cured with only one-tenth of the vial. Without the vaccine, all ten will survive but in a radically diminished physical state. Do we save the one from death or save the ten from a life compromised by extreme disability?<sup>15</sup>

The hypothetical has three key features. We are required to choose between two persons or discrete groups of persons with important but conflicting interests at stake. The affected parties are all on stage. And the consequences of our available choices are both determinate and known at the moment of choosing. On the basis of that information, we are asked to choose between the parties' fundamental interests in dignity, rights, autonomy, freedom from harm, freedom from constraint by another's choices, self-determination, etc. I will refer to hypotheticals that take this form as "trolley problems," after their most (in)famous instantiation.

As T.E. Hill has commented, cases like these are "notoriously hard" to resolve by appeal to the principle that "one must not trade or sacrifice the dignity of anyone for 'more dignity' for others," because *neither* of our available choices "fully respect[s] the

---

<sup>15</sup> See Frances Kamm, *Intricate Ethics*, p. .

dignity of all.”<sup>16</sup> The negative pregnant of Hill’s remark is that, once nonconsequentialists figure out how to handle these notoriously hard cases without resort to aggregation, the rest will be easy. Over the past fifty years, an enormous amount of high-octane intellectual talent has been devoted to figuring out whether these ‘notoriously hard’ cases can in fact be resolved by nonaggregative principles, or instead have to be ceded to some form of aggregation.<sup>17</sup>

The vast majority of the tragic choices we face daily in our private lives and in our collective policy-making roles lack all three of these features. They rarely involve a simple dyadic choice between two individuals or small groups of individuals. The universe of people who are potentially affected can run to the hundreds, thousands, even billions. We typically don’t know *ex ante* the identities of the ultimate winners and losers. And, most significantly, we must choose with at best probabilistic knowledge about the types, magnitudes and frequencies of harms and benefits that will result from our available choices. When choosing under conditions of epistemic uncertainty, it is not easy to avoid aggregation. It is impossible.

How then could it be that deontologists (Kantian, contractualist, and libertarian alike) continue to think otherwise? The answer, I suggest, lies in two central features of the nonconsequentialist literature that have led philosophers to overestimate vastly the potential domain of nonaggregative principles.

The first is the disproportionate amount of attention given to simple trolley-type problems. While trolley problems are notoriously hard—indeed, impossible—to

---

<sup>16</sup> Hill, “A Kantian Perspective on Moral Rules” at 293.

<sup>17</sup> For a more detailed discussion of this literature, see Barbara Fried, “What Does Matter? The Case for Killing the Trolley Problem (or Letting it Die),” *Philosophical Quarterly* 62:248 (July 2012).

resolve without trading off one person's important interests against another, typically the choice between persons can be made by reference to other nonaggregative distinctions (act v. omission, intending v. foreseeing the harm, identifiable v. nonidentifiable victims, etc.). However uncertain or morally unpersuasive those distinctions may be, and however morally unappealing the consequences of relying on them, they often *can* resolve tragic choices on nonaggregative grounds. They can, that is, if we do not inquire beyond the stipulated boundaries of the problem.

But the objective of deontologists is not to show that nonaggregative principles are viable in one-off, carefully curated tragic choices. It is to show that such principles supply a viable, *systematic* alternative to aggregation. To do that, the 'right' (nonaggregative) answers to one-off tragic choices must be able to tell us what to do once we scale up the numbers of parties directly affected, or widen our focus to take into account all the offstage consequences of that isolated choice (on family members, bystanders, and other interested third parties, on the many unidentified individuals whom we could have saved if we had diverted the same resources to their rescue instead). Thomas Nagel trenchantly pressed this problem with respect to Nozick's principle of the inviolability of human beings.<sup>18</sup> The problem is no less acute in kinder and gentler Kantian versions of inviolability.

Many of the problems entailed in widening our focus or scaling up the numbers have been widely discussed in the deontological literature, in the context of debates (inter alia) over threshold deontology, the Paradox of Rights, agent-centered prerogatives, the Doctrine of Doing v. Allowing, the Doctrine of Double Effect, the Duty

---

<sup>18</sup> "Libertarianism Without Foundations," p. 140.

of Easy Rescue, and the moral relevance of proximity and identifiability.<sup>19</sup> In general, I share the view of critics that the nonaggregative solutions offered, far from supplying a viable, systematic alternative to aggregation, amount to a morally unconvincing patchwork of ad-hoc rules and exceptions that must be further qualified when faced with the least perturbation of the facts. The result, to borrow A.R. Ammons's lovely phrase, is a "too-adequate relationship" between moral principles and hypothetical fact, and between moral principles and desired results.

I add just two observations here. First, the challenge that scaling-up poses to nonaggregative decision rules has, if anything, been understated in the literature, because of the prominent role given to life/life or life/harm tradeoffs. However unappealing it may be to sacrifice 100,000 in order to save one, it is possible to do so, because there is no limit to the numbers of people we could kill, let die, or let live severely compromised lives (assuming we have no intention of caring for the latter). To put it another way, killing people or letting them die is a low- or no-cost proposition. But the moment we have to pay to indulge the paradigmatic moral thought that no person should be sacrificed for the lesser interests of others—as, for example, we must do if we really believe that no amount of money is too much to spend to save a person's life-- the thought cannot be indulged very far before deontologists must, and will, fold.

---

<sup>19</sup> For a very useful overview of this terrain, see Larry Alexander and Michael Moore, "Deontological Ethics," <http://plato.stanford.edu/entries/ethics-deontological/>. For discussions of many of these problems in the context of Scanlon's Relative Complaint Model, see Elizabeth Ashford, "The Demandingness of Scanlon's Contractualism," *Ethics*, 113 (2003), pp. 273–302; Alastair Norcross, "Contractualism and Aggregation," *Social Theory and Practice* 28 (2): 303-314 (2002); Norcross, "Comparing Harms: Headaches and Human Lives," *Philosophy & Public Affairs* 26 (1997): 135-67.

Second, as others have remarked, there is an air of desperation to many of the efforts to develop fixes for challenges facing nonaggregation. Some nonaggregationists have been quite explicit that their motivation for doing so is to rescue nonaggregation from itself.<sup>20</sup> In a subtle but important respect, this seems to get things backwards. Surely, before reaching the question how to rescue nonaggregation, deontologists ought to be asking themselves whether it deserves to be rescued.

***Tragic choices in an indeterminate world.*** The second, and in my view far more important, explanation for why so many philosophers remain optimistic that deontological principles can supply a viable *systematic* alternative to aggregation is that they have failed to take seriously the problem of decision-making under uncertainty (risk).<sup>21</sup> As I suggested above, for the most part the deontological literature has sidelined the problem of risk. When it has not, it has mischaracterized it in a fashion that effaces the insoluble problem it poses for nonaggregation.

The balance of the paper is devoted to this issue. Much of the discussion focuses on the academic philosophical literature. My ultimate concern, however, is not academic but practical. The same intuitions driving deontologists to reject aggregation feature prominently in commonsense morality. The disastrous public policy consequences of that piece of commonsense morality can be seen all around us. I return to this matter at the end.

---

<sup>20</sup> Thomas Hill, "A Kantian Perspective," p. . Scanlon, *What We Owe to Others*, pp. . I would include in the same spiritual camp expressions of relief, like the one from Rawls cited below, that what appears to be a very serious problem for nonaggregation can be ducked. Nonaggregationists are hardly the only intellectual sinners in this regard. Most of us are wont to switch into defensive mode once we have convinced ourselves of the rightness of our position. But there are important differences in degree here, reflecting the centrality and severity of the problems we are ducking.

<sup>21</sup> Risk and uncertainty have different technical meanings. Roughly, risk refers to gambles with known odds and uncertainty to gambles with unknown odds. For present purposes, that distinction is irrelevant, and for ease of exposition I use the terms interchangeably here.

### C. Facing Up to Risk

The deontological literature is hardly alone in minimizing the problem of risk. As one commentator has noted, “Throughout the history of moral philosophy, moral theorizing has for the most part referred to a deterministic world in which the morally relevant properties of human actions are both well-determined and knowable.”<sup>22</sup> Given that the consequences of our actions are virtually always epistemologically uncertain *ex ante*, deontologists’ fixation on choices with specified, determinate outcomes seems to require some explanation.

In his well-known article “Conspiracy and the Paranoid Style: Causality and Deceit in the Eighteenth Century,” Gordon Wood supplies an historical one, which traces back to the Scientific Revolution in the 17<sup>th</sup> century. The mechanistic conception of the physical world produced by the Scientific Revolution, Wood argues, “became the paradigm in which the enlightened analysis of all behavior and events now had to take place. Cause was something that produced an effect; every effect had a cause; the cause and its effect were integrally related. Such thinking created a new world of laws, measurements, predictions, and constancies or regularities of behavior all dependent on the same causes producing the same effects.”<sup>23</sup> People came to believe that causal relationships that lacked such regularity and predictability “did not

---

<sup>22</sup> Sven Ove Hansson, “Risk and Ethics,” in Tim Lewens, ed., *Risk: Philosophical Perspectives* (Routledge, 2007), p. 30.

<sup>23</sup> *The William and Mary Quarterly*, Third Series, Vol. 39, No. 3 (Jul., 1982), pp. 401-441, 413.

deserve the name of science.”<sup>24</sup> “Chance... was ‘only a name to cover our ignorance of the cause of any event.’”<sup>25</sup>

Applied to the realm of human action, the mechanistic story located the first mover in human intentions and motivations, from which the action and its consequences followed deterministically. In Wood’s words, “Although the new science tended to remove man from the center of the physical universe, at the same time it brought him to the center of human affairs in ways that even classical and Renaissance thinkers had scarcely conceived of. It promised him the capacity to predict and control not only nature but his own society, and it proceeded to make him directly and consciously responsible for the course of human events.”<sup>26</sup> God still hovered in the background as the theoretical First Mover in the physical and social realms. But in both realms, God was notably absent from the causal story that people really cared about: the story of cause and effect in the phenomenal world in which we live.

In the physical world, the Enlightenment’s mechanistic account of cause and effect has given way over the past two centuries to more complex and probabilistic conceptions of causation. But in the social world, the mechanistic account remains very much alive. In modern analytic moral philosophy, it is lodged in two foundational assumptions: that our intentions are the uncaused cause of our actions, and that the consequences of our actions follow deterministically from the act itself.

---

<sup>24</sup> Wood, “Conspiracy and the Paranoid Style,” quoting Adam Smith, “Lectures ... on the Subjects of Moral and Political Philosophy” I, 9, 122.

<sup>25</sup> Wood, “Conspiracy and the Paranoid Style,” p. 414, quoting Smith, “Lectures,” 11, 22. See also Wood, 420: “Just as devout Puritans believed that nothing occurred without God’s providence, so the liberal-minded believed that nothing occurred without some person willing it. Earlier, men had sought to decipher the concealed or partially revealed will of God; now they sought to understand the concealed or partially exposed wills of human beings.”

<sup>26</sup> Wood, “Conspiracy and the Paranoid Style,” pp. 412-413.

The first of these assumptions supplies the moral undergirding for the Kantian notion of a perfected will, and with it the congeries of negative 'reactive attitudes' that reign supreme in contemporary American politics and common sense morality: blame, retribution, guilt, etc. If we are not the cause of our own intentions-- if they are caused by God, or social forces, or biological ones-- then of what moral interest is the quality of our willing or the ends we 'seek' to bring about? The Enlightenment view of the individual will as an uncaused cause may eventually buckle under mounting evidence to the contrary from neuroscience, genetics, etc., but it has so far managed to ward off similarly powerful evidence of *social* determinism through the dodge of compatibilism.<sup>27</sup>

But my concern here is with the second of these assumptions: that the consequences of our actions follow predictably and deterministically from the actions themselves. In contrast to the first assumption, which has been debated endlessly in contemporary philosophical arguments over free will v. determinism and incompatibilism v. compatibilism, philosophers have said little about the second. Indeed, the central role given to actions with (ex ante) well-determined and knowable consequences is less an argumentative assumption of moral philosophy than a near-universal and unexamined practice.

Those who have explicitly defended moral philosophers' preoccupation with conduct stipulated to have (ex ante) determinate consequences have generally done so on one of two (mutually inconsistent) grounds. The first is that choices made under conditions of epistemological certainty are a different moral kind from those made

---

<sup>27</sup> B. Fried, "Beyond Blame," *Boston Review* July/August 2013.

under conditions of epistemological uncertainty, and that only the former are the proper subject of moral theory. The latter are taken to implicate less lofty considerations.<sup>28</sup> The second is that the ‘right’ way to handle risk can be derived from the ‘right’ way to handle certainty, and hence by resolving the latter we resolve the former as well.

As I have argued elsewhere, both positions are in my view indefensible. Certainty is just the limit case on the logically possible range of probabilities. If a 1.0 probability strikes us as a wholly different kind from, say, a 0.99 probability, it is for reasons that reside in the psychological and not the moral realm. And the moral principles extracted from extensive philosophical engagement with choices stipulated to have determinate consequences (trolley problems, etc.) cannot tell us what to do about choices made under uncertainty. As a result, sidelining risk is essential to maintaining the surface plausibility of the case for nonaggregation as a viable, systematic alternative to aggregation.

The threat that risk poses to nonaggregation is easily demonstrated. Suppose you are planning to drive to the movies. You always drive carefully (holding in suspension for the moment what exactly we mean by ‘carefully’). But you know that even driving carefully, there is a 1 in 50,000 chance that in the course of the trip you will seriously injure or kill someone. If you could know with certainty that, of the tens of thousands of times that you are going to drive over your lifetime, *today* is the day when your driving will result in serious injury or death to another, then you could simply stay home *today*. Doing so entails a personal sacrifice, to be sure—you’ll have to put off

---

<sup>28</sup> See B. Fried, “What *Does* Matter? The Case for Killing the Trolley Problem (or Letting it Die).”

going to the movie or visiting your mother until tomorrow-- but it is hardly a tragic one, and it is one that aggregationists and nonaggregationists alike would make. This is the seductive fantasy world into which trolleyology invites us: a world in which all tradeoffs can be made by pairwise comparisons of the actual harms that *will* result, allowing us to ignore all the other instances of identical, risky conduct in which, as the odds would have it, no one ends up being harmed.

But in a world in which you cannot know, *ex ante*, which (if any) of the tens of thousands of times that you drive over your lifetime you will end up injuring someone, the chance that you take with others' lives *today* is the same as the chance you take every other day you get in your car and drive somewhere. As a result, whatever we decide about the permissibility of your driving *today* will apply equally to all instances of driving. That reality leaves deontologists with two unacceptable alternatives, whether they are proceeding under something akin to Scanlon's Greater Harms Trump Lesser Harms principle or the typical libertarian prohibition on boundary crossing: If the mere possibility that an act will result in death or grave harm to others is sufficient to rule it out, then no act is permissible (except perhaps when necessary to save others from harm more severe than the proposed act threatens to impose).<sup>29</sup> On the other hand, if the mere possibility that an act, undertaken with due care, could result in serious harm to others is not sufficient to rule it out-- if harm must be absolutely certain to eventuate *on this particular occasion*-- then in a world of indeterminate

---

<sup>29</sup> Scanlon acknowledges the absurdity of this result, but believes it can be avoided without resort to discounting harms by their probability of occurring (one form of aggregation), if we allow probabilities to be factored into the definition of the reasonable precautions that "the agent has to take to avoid causing harm." *What We Owe* at . . . . As I discuss below, this maneuver simply relocates aggregation; it doesn't diminish its role.

consequences, all actions are permitted until it is too late to avoid their bad consequences.

Writ large, the same two unacceptable choices present themselves when we make the identical moral calculus collectively-- for example, in choosing rules of general application to govern whether and under what conditions people may drive. The only difference between the two cases is the mode of presentation. In large-number cases, the risk is conventionally stated in frequentist rather than probabilistic terms. That is, rather than asking whether a 1 in 20,000 chance that Ms. Jones will injure or kill someone when she drives X miles today is too great a risk for her to impose on others, we ask whether the statistical certainty that 1 out of every 20,000 car trips of X miles will result in serious injury or death is a sufficient reason to prohibit all driving. If we refuse to let the numbers count when the choice is posed in frequentist terms, we end up with the same two unacceptable alternatives. If we regard statistical certainty as morally equivalent to absolute certainty (whatever we take that to mean), then driving is never permissible, because over the course of a year, however prudently we drive, collectively we are certain to kill or injure thousands of people. On the other hand, if statistical certainty isn't enough-- if we have to know with absolute certainty that a particular instance of driving is going to result in grave injury or death before we can rule it out—then we cannot say anything about the permissibility of driving at all, or under specified conditions, until the harm has already materialized.

Consider the Ford Pinto case again. As I stated at the outset, the moral many have drawn from the Pinto case is that it is “unethical to determine that people should be allowed to die or be seriously injured because it would cost too much to prevent it.”

But what exactly does that mean Ford should do? However safe a car design, for some amount of money it could be made safer. And however much money Ford spends on making it safer, it will never eliminate the possibility that a design feature will be the but-for cause of someone's death. Does that mean that Ford may not manufacture and sell cars? If not, what does it mean? Suppose that by investing \$5 billion in additional safety features, Ford could reduce expected deaths or serious injuries by .01 percent. Should Ford (must Ford) make that investment? If so, how about \$50 billion? \$500 billion? In a world of finite resources, we have to draw the line somewhere. I do not think any deontologist would disagree with that, and I imagine all would draw the line considerably short of \$500 billion, given that such a sum—which in the end consumers will bear-- would price most consumers out of the car market and hijack scarce societal resources that could save thousands if not millions more lives if deployed in a more cost-effective manner. But wherever we draw the line, we will knowingly be choosing to allow Ford to increase the number of preventable deaths 'merely' to save money (to be used for some other generally offstage purposes). The rule that determines where to draw the line cannot come from nonaggregative principles.

The most telling evidence that nonaggregative principles cannot deal with risk has come from deontologists themselves. When confronted with garden-variety tragic choices that must be made under conditions of uncertainty, most deontologists fold, ceding the territory to aggregation.

A surprising number have done so explicitly. Consider this, from Nozick: "It is difficult to imagine a principled way in which the natural-rights tradition can draw the line to fix which probabilities imposed unacceptably great risks upon others. This

means that it is difficult to see how, in these cases, the natural-rights tradition draws the boundaries it focuses upon.”<sup>30</sup> Or this from Rawls: In discussing how to reconcile natural duties (which include “the duty not to harm or injure another”) when they “come into conflict, either with each other or with obligations,” Rawls reluctantly concedes that with utilitarianism off the table, “I do not know how this question is to be settled, or even whether a systematic solution formulating useful and practicable rules is possible.” Happily, he concludes, he needn’t answer the question in order to derive the general principles governing the basic structure of society.<sup>31</sup> Rawls’s dismissal of a utilitarian solution is brief but telling. He considers and rejects use of the “utilitarian principle to set things straight” not because it isn’t a viable solution, but because it would very quickly lead to utilitarianism full stop: “Requirements for individuals so often oppose each other that this would come to much the same thing as adopting the standard of utility for individuals”—a solution that he has already ruled out on the grounds that it leads to “an incoherent conception of rights.”<sup>32</sup>

Many more have folded implicitly, by qualifying our prima facie unconditional duties (not to sacrifice the one for the good of the many, not to cross others’ boundaries) in terms that sound in deontological values but get cashed out as aggregation manqué. Consider this from Charles Fried: “[I]t is wrong to expose the person or property of another to undue risk of harm, but what risk is undue is a

---

<sup>30</sup> *Anarchy, State and Utopia*, p. 75.

<sup>31</sup> *Theory of Justice* (1971), pp. 114, 339-340. For other representative statements ceding the problem of risk to aggregative techniques, see Jules Coleman, *Risks and Wrongs* (Oxford University Press, 2002), p. 210; Feinberg, *Harm to Others*, pp. 190-93; Scanlon, *What We Owe to Each Other*, pp. 204, 205.

<sup>32</sup> *Theory of Justice* at .

function of the good to be attained and the likelihood and magnitude of harm.”<sup>33</sup> Or this from Joel Feinberg: the culpability of an actor should turn on three factors: the probability of expected harm, the magnitude of expected harm, and the independent value of the risky act to the actor himself and to society at large.<sup>34</sup> Or this from Thomas Scanlon: While the ex ante likelihood that harm will occur cannot be used to “diminish[] the complaint of the affected parties [by] discounting the harm by the likelihood of their suffering it,” it is relevant to determining the reasonable precautions that “the agent has to take to avoid causing harm.”<sup>35</sup>

This is surely a case in which covert tools are not good tools. If concepts like “undue risk” or “reasonable precautions” have an operational meaning distinct from aggregation, it needs to be put on the table so it can be assessed side by side with aggregation. If they don’t—if (as I believe) they boil down aggregation manqué—then we can cross of the list that purported disagreement between deontologists and consequentialists and move on.

But, as Rawls foresaw with respect to harm more generally, ceding risk of harm to aggregation comes at a very high price for the nonaggregationist project, given that virtually everything we do poses at least some remote risk of seriously compromising the interests of others. Every time we jog down a city street, or decide to release a prisoner early, or decide not to breach patient confidentiality to warn someone else that he or she may be at serious risk, however prudent our decision, we are taking others’ lives in our hands. As a result, ceding epistemological uncertainty to

---

<sup>33</sup> *Right and Wrong*, at 12.

<sup>34</sup> *Harm to Others*, 190-93.

<sup>35</sup> *What We Owe to Each Other*, pp. 209, 236.

aggregation saves nonaggregative principles by rendering them largely irrelevant.

Not surprisingly, then, those deontologists who have grappled with risk have backed away from the concession, searching for ways to keep risk in the nonaggregationist fold.

One approach I have mentioned already: smuggle in aggregation by the back door in cashing out the qualifications on our absolute duty not to harm (e.g., except when we act with “due care” “or take ‘reasonable precautions’; except when the person we harm is liable to being harmed.<sup>36</sup>) A second approach is to distinguish among cases based on factors (e.g., identifiability, proximity, doing versus allowing, upstream versus downstream harms) that—whatever their independent appeal-- would seem to have nothing to do with the paradigmatic thought that every person has a worth “that cannot be... subject to trade-offs.”<sup>37</sup> A third approach is to seek a meta-principle that provides *nonaggregative* justifications for embracing aggregation (the retreat to reasons in the Kantian literature; reliance on what Nozick describes disparagingly as “a utilitarianism of rights” before proceeding to deploy it himself to derive the just minimal state).<sup>38</sup>

Many of these strategies to rescue risk from aggregation have been discussed extensively in the literature. For the balance of the discussion, I want to focus on a fourth strategy that, in my view, has played a much larger (if largely unnoticed) role in

---

<sup>36</sup> Seth Lazar, “Risky Killing.”

<sup>37</sup> For a trenchant critique of Scanlon’s attempt to distinguish between certain and uncertain harms on this basis, see Norcross, “Contractualism and Aggregation.” I have discussed the same issue in “Can Contractualism Save Us From Aggregation?” *Journal of Ethics* (July 2011).

<sup>38</sup> *Anarchy* at 14. For an excellent summary of many of the arguments, see Larry Alexander and Michael Moore, “Deontological Ethics,” <http://plato.stanford.edu/entries/ethics-deontological/>

the nonaggregationist literature: recharacterizing ex ante *uncertainty* as *certainty* through what amounts to simple hindsight bias.<sup>39</sup>

It would be an exaggeration to say that the case against aggregation is built on hindsight bias. But it would not be that much of an exaggeration. Strip away hindsight bias, and the conclusion that nonaggregative principles cannot supply a viable, systematic alternative to aggregation becomes unavoidable.

#### **D. Is non-aggregationism just one large case of hindsight bias?**

Hindsight bias refers to the logical error of judging the prudence of a decision through the lens of what happens to happen as a consequence of it.

The standard trolley-type problem obviates the need for explicit hindsight bias by building perfect 20/20 hindsight into the definition of the problem. Because we are endowed, ex ante, with perfect knowledge of the ex post consequences of our actions, the ex ante and ex post epistemological perspectives converge. As suggested above, that convergence allows us to indulge the fantasy that the costs of strictly applying the Relative Complaint Model or treating rights as inviolable will be limited to those few occasions in which serious harm *will* result.

But in an indeterminate world, the two epistemological perspectives necessarily come apart. Yes, the ex post consequences of our choices can yield important information to help guide similar choices in the future (e.g., the discovery turned up in the investigation of the Challenger explosion that O-rings crack in freezing temperatures). In some small number of cases, the consequences can shed some light on the ex ante prudence of the particular decision that produced them. But our

---

<sup>39</sup> 'Strategy' is something of a misnomer here, because resort to hindsight bias in the deontological literature is surely unconscious.

reactive attitudes towards bad outcomes are, alas, not that discriminating. They generally take the much cruder and logically indefensible form of concluding that if something bad has happened and some human agent is a but-for cause of it, that agent is to blame (in the moral, not causal, sense).

Our impulse to reevaluate conduct in light of its consequences is hardly surprising. We live in the world that *is*, not in the alternative worlds that, *ex ante*, might have been much more likely to come to pass. When one is a but-for cause of some tragic consequence, it is almost impossible not to run the film backwards and think, “if only I had gone to the store an hour later, none of this would have happened.” That retrospective recharacterization of the tragic choice we *actually* faced-- how to trade off the convenience of driving against the remote possibility that on any occasion we drive we will harm uncountable and unidentifiable others— creates the illusion that the choice needn’t have been tragic at all. After all, would it really have been a great inconvenience to go to the store an hour later? That is to say, it invites us to reconstruct every choice as a trolley problem, in which we can choose with full knowledge of the consequences of available choices.

But the illusion is based on simple hindsight bias. As I suggested above, faced with the actual decision every time we get into a car whether to drive at all, and if so, with what level of precaution, all we know is that there is a remote possibility that *this* is the day when, driving prudently, we will nonetheless kill someone. Of course that remote possibility gets factored into our rational *ex ante* choice. But it is only one factor to weigh against many others, and is necessarily weighed through some form of aggregation. It manifestly does not prove dispositive for the hundreds of millions of

Americans who drive, or the government charged with the task of deciding under what conditions they are permitted to do so.

I turn now to a closer examination of hindsight bias at work in the deontological literature on harm to others.

**1. Contractualism.** The contractualist forms of justification deployed most famously by Rawls and Scanlon seem an unlikely place to find hindsight bias at work, given that contractualism is by its nature committed to an *ex ante* epistemological perspective and, in Rawls's case, one of radical uncertainty. But hindsight bias does in fact play a critical role for both Rawls and Scanlon in diverting Kantian contractualist thought experiments from where they naturally lead (which, at least in Rawls's case, is to some form of aggregate consequentialism).

(a) *Rawls's maximin solution.*<sup>40</sup> In principle, the requirement of unanimous consent in contractualist thought experiments rules out aggregation, provided that the parties have divergent interests *ex ante*. But Rawls's Original Position famously requires representative persons to choose under conditions of radical uncertainty (as to their position in society, the preferences they hold, etc.). As a consequence, everyone stands in the same epistemological position *ex ante*, effectively converting all *interpersonal* tradeoffs into *intrapersonal* tradeoffs. As John Harsanyi famously argued, from that epistemological perspective, the representative individual's rational choice is to maximize expected utility, given their risk preferences, etc. Given a society of *n* individuals, each making that same choice, we end up with average

---

<sup>40</sup> I borrow this discussion from B. Fried, "*The Unwritten Theory of Justice.*"

utilitarianism.<sup>41</sup>

Rawls, rerunning Harsanyi's thought experiment from the same epistemological perspective and with the same assumption of rationality, ends up rejecting the utilitarian answer in favor of a maximin rule, famously instantiated in the Difference Principle.

As many have observed over the years, it is wildly implausible that rational actors would display the infinite risk aversion necessary to make maximin the rational choice, in preference to some less extreme version of sufficientarianism.<sup>42</sup> What then explains Rawls's contrary conclusion?

There are a number of factors, in addition to Rawls's empirical hunch that faced with a high-stakes choice like this, people would in fact display a high (even if not infinite) degree of risk aversion.<sup>43</sup> Among them are Rawls's rejection of Harsanyi's Principle of Insufficient Reason as the rational response to a gamble with unknown odds,<sup>44</sup> and his apparent misunderstanding of how risk aversion is factored into the von Neumann/Morgenstern utility function Harsanyi deployed.<sup>45</sup>

---

<sup>41</sup> John Harsanyi, "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking," *Journal of Political Economy* (1953); Harsanyi, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy* (1955).

<sup>42</sup> Rawls repeatedly denies he is resting the case for the Difference Principle on that assumption. *Theory of Justice* (1999), pp.149, 512-13. But it is difficult to see how else to interpret comments like the following: "[T]he two principles are those a person would choose for the design of a society in which his enemy is to assign him his place"; the Difference Principle would be chosen by someone who "has a conception of the good such that he cares very little, if anything, for what he might gain above the minimum stipend that he can, in fact, be sure of by following the maximin rule," *id.* at 132-34; Each principle of justice "must suppose that the marginal utility of these fundamental interests is infinite; this requires anyone in the original position to give them priority and to adopt the two principles of justice," "Some Reasons for the Maximin Criterion," *The American Economic Review*, Vol. 64, 141-146, 143; or "The standard of utility approaches maximin as risk aversion increases without limit." *Id.* at 143. For further discussion of this point, see Fried, "The Unwritten Theory of Justice."

<sup>43</sup> *Theory of Justice*, p. 133.

<sup>44</sup> Rawls famously distinguished between risk (taking a gamble with known odds) and uncertainty (taking a gamble with unknown odds), the latter characterizing the Original Position. Harsanyi (1955, 1975) stipulated that, faced with the latter, the rational procedure is to treat all possible outcomes as

But the most important factor, by far, is simple hindsight bias. While maximin is almost certainly *not* a rational choice from the perspective of representative persons deciding on principles of justice *ex ante* from the Original Position, it *is* a rational choice from the perspective of those who find themselves to *be* the worst off *ex post*. Rawls's official position, of course, is that the choice of principles is to be made from an *ex ante* and not an *ex post* epistemological perspective.<sup>46</sup> But in defending the rationality of maximin, he steps out of the *ex ante* perspective and helps himself to the *ex post* perspective of the person who, in the event, will turn out to be the big loser. Positing two individuals, *A* (the more favored) and *B* (the least favored), Rawls counters the hypothetical complaint of *A* that "he is required to have less than he might since his having more would result in some loss to *B*," by noting that *A* is still better off than he would have been under a system of noncooperation.<sup>47</sup>

Nozick famously responded to that argument by pointing out that the same defense could be given for *any* scheme of cooperation, including one in which *A*—the

---

equiprobable (the Principle of Insufficient Reason), leading him straightforwardly to average utilitarianism. Relying on Daniel Ellsberg's famous paradox, Rawls rejected as irrational not just the Principle of Insufficient Reason, but any form of probabilistic reasoning when "knowledge of likelihoods is impossible, or at best extremely insecure" (TJ, 134), arguing that the rational strategy would instead be maximin. For further discussion of his choice and whether it can be justified, see Fried, "The Unwritten Theory of Justice."

<sup>45</sup> As Rawls recognized, a significant degree of risk aversion is built into the von Neumann/Morgenstern utility functions (TJ, 282; Harsanyi, 1975, 600). All other things being equal, this means that average utilitarianism, like the Difference Principle, will weight avoiding bad outcomes more heavily than achieving good ones. Rawls sought to distinguish the two on the grounds that the risk aversion built into von Neumann/Morgenstern utilities reflected peoples' *psychological attitudes* towards risk taking, which he deemed arbitrary from a moral point of view. *Id.* at 282. Rawls was simply wrong on this point. Von Neumann/Morgenstern utilities, like the Difference Principle, reflect the relative value that individuals assign to different outcomes, not a psychological attitude towards risk-taking. (Harsanyi, 1975, 600). Since Rawls (at least in his rational choice mode), like preference-based utilitarians, sought to model the actual degree of risk aversion the choices of those in the Original Position would reflect, any differences between the two should therefore reflect different empirical assumptions about risk aversion rather than any disagreements in principle.

<sup>46</sup> *Theory of Justice* (1999), p. 143.

<sup>47</sup> *Theory of Justice* (1971), p. 103.

more favored--demanded a *larger* share of the benefits than he would receive under the Difference Principle.<sup>48</sup> But as Nozick noted in a lengthy footnote, there is a deeper, structural problem with Rawls's argument: how to make sense of it from the "perspective of the original position."<sup>49</sup> In the original position, there would be no *A* and *B*; there would only be the representative person with some uncertain chance of being *A* and some uncertain chance of being *B*. If that representative person chose the Difference Principle as the principle, on balance, that would be most advantageous to his future self given the possibilities of turning out to be *A* or *B*, to whom is he now complaining? And when is now? Nozick concluded (rightly in my view) that the only sensible way to construe *A*'s and *B*'s complaints, as Rawls presents them, is to imagine that *A* and *B* have stepped out of the original position and are viewing matters *ex post*, after they know whether they have won or lost. From that perspective, *B*, knowing he is the loser and *A* the winner, will understandably be unmoved by the statement that he should have less so that *A* might have more.

But if *B*'s argument from the *ex post* perspective is decisive for Rawls, his contractualist thought experiment is not merely superfluous; it is misleading. For in that case, the notion of impartiality that motivates the Difference Principle is not that we should stand as equals in choosing the basic structure of society *ex ante*, but rather that we should stand as more or less equals (at least as regards primary goods) at every moment in our lives. In short, Rawls has converted choice under radical uncertainty (the Original Position) into continuous trolley-type pairwise comparisons between *ex post* winners and losers. Take away that illicit reliance on what (from a truly *ex ante*

---

<sup>48</sup> *Anarchy*, p. 196.

<sup>49</sup> *Anarchy*, pp. 196-97\*.

perspective) amounts to ex post regret, and Rawls's case for the Difference Principle as an ex ante rational choice, given the representative person's "assessment of which alternative is most likely to advance their interests... given their situation," falls apart.<sup>50</sup>

(b) *Scanlon's Relative Complaint Model*. Scanlon has built hindsight bias into his Relative Complaint model from the start, by stipulating that ex ante complaints should be weighted in accordance with the *actual* (ex post) harms that will befall actual victims:

Suppose that A is a principle which it would be rational for a self-interested chooser with an equal chance of being in anyone's position to select. Does it follow that no one could reasonably reject A? It seems evident that this does not follow. Suppose that the situation of those *who would fare worst* under A, call them the Losers, is extremely bad, and that there is an alternative to A, call it E, under which no one's situation would be nearly as bad as this. Prima facie, the losers would seem to have a reasonable ground for complaint against A.<sup>51</sup>

In *What We Owe To Each Other*, Scanlon fleshes out the case for an ex post epistemological perspective as follows: "In assessing the rejectability of [a] principle" based on relative burdensomeness, "we can begin...by taking the maximum level of burdensomeness and asking whether that would give a potential agent reason to reject the principle." Where the activity in question benefits many but "involve[s] risk of serious harm to some.... [i]t is obvious what the generic reasons would be for rejecting such a principle from the standpoint of someone *who is seriously injured*."<sup>52</sup> Most other contractualists writing in a Scanlonian vein have followed suit, adopting an ex post

---

<sup>50</sup> "Kantian Constructivism in Moral Theory," *Journal of Philosophy* 77: 515-572, 524 (1980).

<sup>51</sup> T. Scanlon, 'Contractualism and Utilitarianism', in A. Sen and B. Williams, eds., *Utilitarianism and Beyond* (Cambridge University Press, 1982), pp. 122-23 (italics added). See also Scanlon, *What We Owe to Each Other*, p. 207.

<sup>52</sup> *What We Owe*, p. 207.

perspective in weighting the relative complaints of the would-be Big Losers against the interests of all others.

As with Rawls, Scanlon, by adopting an ex post perspective on harms, turns what is purportedly an exercise in ex ante contractualism into a series of pairwise trolley problems with determinate consequences. When combined with the maximin rule built into the Relative Complaint Model, the result is the first of our two unacceptable alternatives: Whichever representative hypothetical person, peeking ahead, discovers she will be the one most seriously harmed by adopting a given principle may reasonably reject it on that basis, provided that the harm to her will be serious as an absolute matter and there is some alternative course of action available with a less-bad worst outcome. Assuming that her reasonable rejection is dispositive with respect to the principle in question, the result, writ large, is that virtually every action in the world can reasonably be vetoed by someone, producing moral gridlock.

Scanlon is hardly unaware of the moral gridlock problem. He raises the possibility of avoiding that untenable conclusion by assuming that the worst possible outcome of every action is death, and then resorting to his own tie-breaker rule: where the worst possible harm resulting from different principles is equally bad, we turn to aggregation as a tiebreaker. But using the tie-breaker rule to avoid the unpalatable implications of his Relative Complaint Model simply adopts aggregation in two steps rather than one, and Scanlon explicitly disavows the strategy.<sup>53</sup>

As suggested above, Scanlonians have sought other ways to avoid moral gridlock under the Relative Complaint Model. The gridlock problem, however, would never arise

---

<sup>53</sup> *What We Owe*, pp. .

from a truly ex ante version of Scanlonian contractualism, for the same reason that the maximin principle is exceedingly unlikely to be the outcome of a truly ex ante version of Rawlsian contractualism. From an ex ante perspective, Scanlon's representative persons, like Rawls's, will choose on the basis of expected outcomes, not actual ones.

Scanlon faces a more difficult challenge than Rawls in coming up with unanimous agreement from that ex ante perspective, because his representative persons (unlike Rawls's) calculate their expected outcomes with full knowledge of their particular situation and self-interest. But the challenge is still easier than the total moral gridlock that results from the ex post perspective that Scanlon *actually* adopts. More to the point, if we are trading off expected outcomes rather than actual outcomes, then the numbers will necessarily count, because expected outcomes automatically discount harms by their probability of occurring—something that Scanlon has ruled out.

(c) Combine ex ante and ex post perspectives. A number of contractualists have argued that the way to deal with the divergence of ex ante and ex post epistemological perspectives on the identical act is to combine the two perspectives in some fashion. Some have suggested doing so by requiring that conduct, to be permissible, must be permissible from both perspectives.<sup>54</sup> Others have suggested that the two perspectives should be weighed against each other in some form of reflective equilibrium, to come up with one 'right' answer.<sup>55</sup> As I have argued at length elsewhere, combining the two perspectives in either fashion is a logical impossibility.<sup>56</sup> Moreover, to the extent that the ex post perspective is given any weight in judging the

---

<sup>54</sup> Kamm, *Intricate Ethics*, pp .

<sup>55</sup> James Lenman, "Contractualism and Risk Imposition," *Politics, Philosophy and Economics* 7: 99–122 (2008).

<sup>56</sup> Fried, "Can Contractualism Save Us From Aggregation?"

permissibility of an ex ante choice, it amounts to simple hindsight bias.

**2. The deontological moral imperative: First, do no harm.** The problem of hindsight bias has gotten significant attention from philosophers in one part of the deontological literature: the paradox of moral luck. In the classic “moral luck” case posed by Bernard Williams, a bus driver sets out one day, following his accustomed route and driving with due care. A young child darts out in front of the bus. The driver, who does not see her and could not have seen her until it was too late to stop, hits and kills the child.<sup>57</sup> If asked ex ante (that is, before the child was hit) whether the bus driver was acting as we would have him act, the average person would say yes. But if asked the same question ex post, after things turn out badly, the average person (or so it is assumed) would conclude that the driver acted wrongly, after all.

At the risk of being accused of philistinism, I confess to puzzlement why Williams’s hypothetical has *ever* been thought to raise a logical conundrum. There is, to be sure, a tension here between ex ante and ex post perspectives. But the tension does not arise from a *logical* paradox. If deontologists really mean what they say when they say that the consequences of what we do determine whether we should have done it, they are simply elevating hindsight bias to high moral principle.<sup>58</sup> Generally, however, they mean something different. The judgment reached from the ex ante epistemological

---

<sup>57</sup> The original version of the hypothetical comes from Bernard Williams, “Moral Luck.” In some versions, the hypothetical is altered to make the driver a little bit careless (e.g., he takes his eyes off the road for a split second) but the consequences are massively out of proportion with that moment of carelessness. In my view that change does not change the central problem of outcome luck; it simply changes the degree of luck involved. For an astute discussion of these issues, see Jeremy Waldron, “Moments of Carelessness and Massive Loss.”

<sup>58</sup> For an example of an author who seems genuinely to mean it, see Michael Moore, *Placing Blame: A Theory of the Criminal Law*, ch. 5 (1997); *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics* (2009).

perspective concerns whether the driver, given the information that was available to him when deciding whether or not to drive and with what level of precautions, chose wisely. The judgment from an ex post perspective expresses something else: our emotional reactions (regret, remorse, anger, blame, etc.) to the unfortunate consequences of that decision, or our belief that the driver, notwithstanding that he acted as we would have had him act, nonetheless has an obligation to alleviate the bad consequences ex post, by apologizing, offering compensation, etc. If this is what is going on, there is no paradox to be resolved. The two judgments are answering different questions.

There are a host of interesting things to say about our ex post responses when things turn out badly through no 'fault' of the actors. Perhaps, as Strawsonians have argued, our reactive attitudes have a moral, and not just a psychological or emotional, basis. But if they do, the moral component has nothing to do with how the driver should have acted in the first instance—in particular whether his ex ante choice should have been governed by aggregative or nonaggregative reasoning.

The identical confusion between ex ante and ex post perspectives is built into the moral imperative at the heart of the nonconsequentialist literature on harm to others: Do not harm others/do not violate their rights. That imperative economically collapses the ex ante perspective (do not act in a manner that may possibly/is likely to/is certain to result in harm to others or to violate of their rights) and the ex post perspective (if what you did actually caused harm/violated someone's rights, you were wrong to do it).

In a world in which the consequences of our actions are known with certainty ex ante (the fantasy world of trolley problems), the two epistemological perspectives

converge, because they are responding to the same information. Ex ante, we knew that act X would produce Y consequences. Ex post, when it in fact produces Y consequences, we have learned nothing new. In a world of epistemological uncertainty, however, the two perspectives by definition come apart. If (as is the case in the world in which we live) all available actions carry some uncertain risk of harm to others, we cannot choose to act so as *not to harm others*. We can, of course, act so as to minimize the chances that we will harm others, given the other objectives we are trying to achieve. But doing so necessarily requires that we discount harms and benefits by their probability (frequency) of occurrence. That is to say, it requires that we let the numbers count.

Notwithstanding the logical impossibility of acting *so as not to harm others*, the moral imperative to do so is the usual starting point for working out the scope of our ‘duty not to harm’ in the Kantian and corrective justice literature.<sup>59</sup> In the moral luck literature, the conflict between ex ante and ex post perspectives is at least on the table for all to see and argue over. In contrast, when the same logical impossibility is housed in the imperative to “act so as not to harm others,” it slips by undetected.

Among non-libertarians, the imperative to “act so as not to harm others” has led to some very odd conclusions, including the second of deontologists’ unacceptable alternatives: we cannot say anything about the permissibility of the act until it actually results in harm to others. In John Goldberg and Benjamin Zipursky’s formulation, “The wronging [meaning harming] of one person by another is the very essence of the enterprise [of acting wrongfully], and until such an event happens, there is no occasion

---

<sup>59</sup> For further discussion, see B. Fried, “The Limits of a Nonconsequentialist Approach to Torts,” *Legal Theory* 18: 231-262 (2012).

to inquire whether an actor can or should be held to have acted wrongfully by violating a moral or legal obligation of conduct.”<sup>60</sup>

Among libertarians, the hindsight bias built into the “duty not to harm others” operates somewhat differently, because libertarians typically don’t wrestle with the permissibility of conduct from an ex ante epistemological perspective at all. They enter the (hypothetical) scene ex post-- that is, only after they conclude from the fact that someone’s ‘boundary’ was crossed that a ‘rights violation’ has occurred. Most never address the question of how an agent, operating in a world in which every action carries some risk of harm to others, should go about deciding ex ante how to act. As noted above, Nozick’s response in a moment of candor was that deontology could not answer that question. He is right. But if so, how does it come to be that rights theorists can offer a confident judgment on the same question ex post: what you did was wrong after all, on deontological grounds, because in the event it ended up violating another’s absolute right to their person or property? The explanation, once again, is simple hindsight bias.

#### **E. The Future of the Nonaggregationist Project.**

Granting everything I have argued above, there is still some domain in which nonaggregative decision rules, however morally appealing or unappealing one might find them, may present a viable (in the minimal sense of operationalizable) solution to hard interpersonal tradeoffs. Manifestly, we *can* spend \$10 million to rescue one person rather than spending the same sum to keep thousands or millions out of harm’s

---

<sup>60</sup> John Goldberg and Benjamin Zipursky, “Tort Law and Moral Luck,” at 1138. For further discussion of this issue, see B. Fried, “The Limits of Nonconsequentialist Approaches to Tort Law.”

way to begin with. We can, that is, provided that the occasions we do so are few and the costs not 'too great'.

These luxury indulgences of our nonaggregationist instincts dominate the landscape in moral philosophy. But put together they add up to an infinitesimal fraction of the tragic tradeoffs we face every day. They clearly do not add up to a viable, systematic alternative to aggregation. Whether they add up to enough to justify the extraordinary attention that has been devoted to them over the past fifty years I leave to nonconsequentialists to resolve for themselves.

But at least for Rawlsians and other left-liberal leaning Kantians, given their policy goals this surely cannot be the best instrumental use of their time. Those goals, in my view, are just a hair's breadth from consequentialists', properly understood.<sup>61</sup> Rather than making common cause with consequentialists, Kantians have spent much of the last fifty years obsessing over their differences, real and imagined. For Kantians, this seems to me a valuable opportunity squandered.

Libertarians are in a very different prudential position. As long as some form of instinctual 'don't tread on me' libertarianism remains a strong force in the American psyche and leads to policy outcomes that are congenial to them, libertarians have no instrumental motivation to question the capacity of rights theory to resolve the real-world problems we actually face.

## **F. Non-aggregative instincts at large in the real world.**

Moral philosophers are not the only ones who have refused to face up to the problem of scarcity and the ubiquitous, tragic tradeoffs it necessitates. The same

---

<sup>61</sup> See Fried, *Left-Libertarianism*; Fried, "The Unwritten Theory of Justice."

wistful picture of a world in which one individual's fundamental interests rarely if ever have to be traded off against another's is at the heart of the origin myth of America, and of our persistent, common-sense moral intuitions.

Many of the worst pathologies in contemporary American public policy can be traced, directly or indirectly, to our refusal to face up to scarcity and the tradeoffs it unavoidably entails. I include here our inability to think systematically about risk—that is, about potential problems *before* they become actual ones with actual victims; our indifference to costs (in lives or money) that are out of our immediate line of sight; our collective refusal to think or talk about the kinds of interpersonal tradeoffs that we will inevitably make, overtly, covertly or by default; our instinctual reaction whenever a choice turns out badly that the choice itself must have been wrong and the officials responsible for it are to blame; and our fatal attraction to government by anecdote, which holds our judgments hostage to the manipulations of politicians and others with a strong interest in hijacking them.

All of these lay reactions have close analogues in contemporary moral philosophy: the conviction that no action can be judged wrongful until it eventuates in harm; the conviction that certain harms and uncertain harms are different moral kinds, and that it is outside of—if not beneath—the job description of moral philosophers to think about the latter; the practice of honing moral instincts on one-off, bizarre hypotheticals in which available choices all have determinate consequences; and most of all the paradigmatic moral thought that “the grave objections of a single individual [must be allowed] to defeat a project or plan that would otherwise be expedient for many.”

If nonconsequentialist philosophers do not mean to aid and abet the real-world

instantiations of their moral intuitions, I think they have some soul-searching to do. Philosophers may have limited influence over commonsense morality. But at the very least, they ought to be pushing back against its catastrophic failures, rather than encouraging the habits of mind that produced them.