Cost-Benefit Analysis, Precautionary Principles, and Counterterrorism Surveillance
Katherine J. Strandburg
NYU School of Law

I. Introduction

This Article addresses one small piece of a difficult and important puzzle: How should government surveillance powers be constrained at a time when technological developments mean that surveillance poses an increasingly serious threat to privacy and civil liberties while the global political situation continues to raise the specter of international terrorism? While a rather vast literature addresses the constitutionality of expansive counterterrorism surveillance powers granted following the 9/11 attacks,1 a several commentators have begun to consider the question of counterterrorism surveillance from a regulatory perspective.2 This Article explores one of the many questions that arise out of the regulatory framing: What is the best decision-making framework for assessing the appropriate scope of counterterrorism surveillance powers?

In other regulatory contexts involving risk management, such as environmental regulation, the leading contenders for decision-making framework are a kind of souped-up cost-benefits analysis, which I’ll call CBA 2.0,3 and the so-called Precautionary Principle.4 Cost-benefit

---


3 Professor Cass R. Sunstein is one of the most prominent developers and advocates of CBA 2.0. His academic work on this subject includes Valuing Life: Humanizing the Regulatory State (2014); Worst-Case Scenarios (2007); The Real World of Cost-Benefit Analysis: Thirty-Six Questions (And Almost as Many Answers), 114 Colum. L. Rev. 167 (2014); The Limits of Quantification, 103 Cal. L. REV. 1369 (2014); The Catastrophic Harm Precautionary Principle, Issues in Legal Scholarship, Art. 3 (2007); Laws of Fear (2006). Sunstein also developed and implemented his views on regulatory decision-making during his tenure as Administrator of the Office of Information and Regulatory Affairs, Office of Management and Budget (OIRA) from 2009-12. For a sampling of other important literature on the topic, see, e.g., Richard Revesz and Michael Livermore, Retaking Rationality: How Cost-Benefit Analysis Can Better Protect the Environment and Our Health (Oxford University Press, 2008); Matthew Adler, Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis (Oxford 2012); Matthew Adler, Cost-Benefit Analysis and Distributional Weights (2013); Matthew Adler, Cost-Benefit Analysis, Static Efficiency
analysis is favored in the U.S. regulatory context. Indeed, its preferential status has been enshrined in executive orders since President Ronald Reagan issued Executive Order 12291 on February 17, 1981 mandating that the costs and benefits of regulation be considered. The government’s heavy reliance on cost-benefit analysis has been heavily criticized for a variety of reasons, including the difficulty of incorporating unquantifiable values into the analysis. CBA 2.0 responds to a number of the most common critiques, primarily by finding methods of quantitatively valuing seeming unquantifiable costs and benefits. The counterterrorism surveillance context poses particularly intractable problems for cost-benefit analysis, however, because of the uncertainty and normatively important values on both sides of the issue.

The Precautionary Principle, one version of which states: “Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation,” is intended to deal with regulatory contexts involving highly uncertain, but potentially severe, harms. It too, however, has been the subject of blistering criticism. Most significantly, it has been derided as fundamentally inconsistent because of the possibility that precautionary regulations may inflect harm equal in severity to the potential harms they are intended to prevent. As Sunstein puts it: “Risks are on all sides of social situations, and efforts to reduce risks can themselves create risks. For this reason, the Precautionary Principle forbids the very steps that it requires.” Recently, however, philosopher Daniel Steel has taken on the Precautionary Principle’s critics, setting out a more precise framework, which I’ll call PP 2.0, that avoids many of the Precautionary Principle’s pitfalls, including the inconsistency problem.

This Article compare CBA 2.0 and PP 2.0 with the primary goal of determining whether PP 2.0 has the potential to be useful in assessing the desirability of particular constraints on counterterrorism surveillance. The discussion here is quite general, and, as a result, the conclusion of the inquiry is a resounding “maybe”: PP 2.0 seems to have the potential to provide useful traction on the issues, in a way that CBA 2.0 has been unable to do. In part, this Article responds to Professor Cass Sunstein’s recent symposium article, Beyond Cheneyism and

---


5 See also EO 12866 (Sept. 30, 1993); EO 13563 (Jan. 18, 2011)

6 1992 Rio Declaration.

7 See, e.g. Sunstein, Laws of Fear.


9 Daniel Steel, Philosophy and the Precautionary Principle: Science, Evidence, and Environmental Policy (Cambridge 2014). See Ch. 2 for Steel’s response to the inconsistency critique.
Snowdenism, in which he decries the tendency to adopt opposing extreme views about counterterrorism surveillance, linking those views to the Precautionary Principle approach. Sunstein, who is probably the most prominent proponent of CBA 2.0, argues that the Precautionary Principle is unhelpful because of its fundamental inconsistency. He acknowledges, however, that CBA 2.0 is hard pressed to grapple with the uncertainty and normative tensions of the counterterrorism surveillance context. As a result, he proposes only four general ideas about how to approach the problem. The first, break-even analysis, is the most closely tied to CBA 2.0. The others -- avoiding gratuitous costs, avoiding illicit grounds for surveillance, and avoiding the “worst of the worst” cases -- are rather untethered to cost-benefit analysis. In any event, while none of these suggestions is objectionable, it seems doubtful that they will take us very far. In light of this situation, a more serious exploration of PP 2.0 seems in order.

Part II of this article describes CBA 2.0 and PP 2.0 in more detail and compares their features. Part III describes CBA 2.0’s difficulties in resolving the conundrums associated with counterterrorism surveillance, while Part IV applies PP 2.0 to the issue. Part IV demonstrates that it is possible to define sensible and consistent versions of the Precautionary Principle in the counterterrorism surveillance context. It also demonstrates that analogues of Sunstein’s “avoid gratuitous costs” and “avoid illicit grounds for surveillance” principles arise naturally from the PP 2.0 approach and explains how PP 2.0 relates to the break-even analysis and “worst of the worst” case principles. It also demonstrates how a specific PP 2.0 application might be used to make useful progress on normative questions regarding counterterrorism surveillance constraints, even though it does not definitively resolve them. In essence, PP 2.0 is like CBA 2.0 in that it structures the analysis in ways that help to avoid particular decision-making failures.

II. Cost-Benefit Analysis 2.0 and Precautionary Principle 2.0: Moving Toward the Middle?

Both cost-benefit analysis and the Precautionary Principle are mechanisms for systematizing decision-making so as to avoid predictable mistakes and failure. Cost-benefit analysis is concerned about the dangers of agency bias and neglect of regulatory costs, while the Precautionary Principle focuses on concerns about neglecting potentially severe, but uncertain harms, particularly if they are not easily monetizable. Each has its critics and each has responded to those critics to some extent. This Part explores and compares the two approaches in preparation for applying them to the counterterrorism surveillance problem. It argues that both CBA 2.0 and PP 2.0 are viable approaches to decision-making, though neither is workable, or normatively defensible, in all decision-making contexts. While CBA 2.0’s more inclusive approach to non-market values and PP 2.0’s consistency and efficiency requirements have brought the two analyses somewhat closer together, they are distinct and likely to remain so, in

---

10 Cass R. Sunstein, Beyond Cheneyism and Snowdenism, U. Chi. L. Rev. (forthcoming), available at http://ssrn.com/abstract=2589636. (Note that this Article responds to an early draft version of Sunstein’s paper and presumably will need to be updated when the final version appears.)
part because they are aimed correcting very different potential flaws in the decision-making process.

A. Cost-Benefit Analysis 2.0

Cost-benefit analysis is a means of systematizing and quantifying decision-making by “weighing the pros and cons.” Cost-benefit analysis is the subject of a rather vast literature debating its pros and cons as a theoretical and normative matter and in practical application. The requirement of cost-benefit analysis for federal regulations originally was imposed with the ideological goal of cracking down on excessive regulation, which was understood to impose unnecessary economic costs. The requirement has survived successive Democratic and Republican administrations and currently is enshrined in Executive Orders 12866, promulgated in 1993, and 13563, promulgated in 2011. Its form, however, has been adapted over the years to respond to some of its critics’ concerns.

In general, to weigh the pros and cons of a proposed regulation [R] of some activity A, one needs a means to determine whether future states of the world will be better according to some normative criterion if [R] is imposed. One ordinarily cannot predict the future effects of a regulation precisely in order to decide whether to impose it. For each alternative, some range of outcomes ordinarily is possible. To assess whether the future will be better if [R] is imposed, one thus needs i) some means for predicting the outcomes that are likely to occur if one imposes [R] (or not), ii) some means for valuing potential outcomes (“value metric”) and iii) some means for aggregating those valuations over potential outcomes to assess whether the state of the world is likely to be better or worse if one chooses to impose regulation [R].


14 See OIRA Primer at 14-15, instructing agencies to:

*Specify potential scenarios.* As a first step, the agency should specify a set of plausible, mutually exclusive scenarios for both the baseline and for each regulatory alternative. Each scenario represents a complete description of a state of the world, including its evolution through time, that could arise. The goal is to specify scenarios that cover the full range of how the benefits and costs of the rule might vary. Typically, this is done by specifying the set of factors that affect the benefits and costs of the regulatory alternatives.

*Calculate the benefits and costs associated with each scenario.* Once the set of plausible scenarios has been specified, the agency can calculate the benefits and costs associated with each scenario. …

This description comes from the part of the primer dealing with uncertainty analysis, but it is implicit in the discussion that the same process is involved in deciding whether the benefits of a rule outweigh its costs.
Traditional cost-benefit analysis compares the monetary costs and benefits of imposing regulation [R] to those associated with a baseline in which activity A continues unregulated. Traditional cost-benefit analysis employs a wealth maximization approach to individual utility, in the sense that it assumes that all “goods” and “bads” can be quantified along a single scale (usually taken to be monetary) and that utility is the sum of the goods minus the sum of the benefits. Overall costs and benefits for each possible outcome scenario are calculated according to a utilitarian framework, in which social value is assumed to be the sum of individual utilities. To aggregate the possible outcomes and produce an overall cost-benefit result, costs and benefits from all potential outcomes of a given decision are aggregated using a probability-weighted expectation value. Traditional cost-benefit analysis also assumes that the probabilities of the various potential outcomes of the decision can be predicted and quantified to make this possible. 

The use of traditional cost-benefit analysis (CBA) in government policymaking long has been subject to a variety of critiques. Some critics argue that its focus on quantification leads to the neglect of important values that are difficult or impossible to quantify. Others object to its utilitarian approach to social welfare on the grounds that it does not account for distributive impact and various sorts of incommensurabilities. Another possible critique, consistent with

---

15 There is no necessary logical connection between using a cost-benefit value metric for assessing states of the world and this particular aggregation approach. Nonetheless, when “cost-benefit analysis” is discussed or employed, this aggregation approach is presumed.


17 There are various flavors of the incommensurability critique. Some critics reject a consequentialist social welfare approach entirely, taking a purely deontological view that certain actions should be prohibited or required without reference to their effects on “utility.” Others maintain a consequentialist view, but reject the wealth maximization assumption that individual utility can be computed by summing values along a single scale. Many would argue, for example, that individual welfare depends on “baskets” of different goods, in the sense that having two pears and a peach might be better than having three pears regardless of whether one prefers pears or peaches. Market prices reflect only the value to a marginal consumer of adding something to that consumer’s “basket of fruit” at a given time. The “basket” incommensurability critique can be applied generally to the way that individual utility depends on various aspects of life, such as health, work, family, friends, and material goods. Some have proposed alternative metrics for individual utility, such as happiness or quality adjusted life years (QALYs), which might overcome some of these incommensurability critiques by valuing an individual’s “basket” of life aspects directly. See, e.g., John Bronsteen et al, Well-Being Analysis v. Cost-Benefit Analysis, 62 Duke L. J. 1603 (2013), Matthew D. Adler et al., Would You Choose to be Happy? Tradeoffs between Happiness and the Other Dimensions of Life in a Large Population Survey, Duke Law School Public Law & Legal Theory Series No. 2015-35 (August 5, 2015), available at SSRN: http://ssrn.com/abstract=2640117; Carol Graham, An Economist’s Perspective on Well-Being Analysis and Cost-Benefit Analysis, 62 Duke L. J. 1691 (2013); Cass R. Sunstein, Cost-Benefit Analysis, Who’s Your Daddy?, J. Benefit-Cost Analysis (Forthcoming 2015); George E. Torrance, Measurement of health state utilities for economic appraisal: A review, 5 J. Health Econ. 1 (1986). Operationalizing such measures for policymaking is not straightforward, however, while they are designed to measure total individual utility, when used to assess proposed regulation, they are more likely to serve as alternatives to monetization for measuring the average marginal impact of some particular change along a single value scale. Finally, some incommensurability critiques argue that certain categories of values (often associated with rights) should be compared only with values of the same type. Thus, even if tradeoffs among rights are sometimes necessary, tradeoffs between rights and more mundane goods should never be made (or should be made only when the mundane harms are exceeding large). See, e.g., Amartya Sen, Rights and Agency, 11 Philosophy & Public Affairs, 3 (1982); Eyal Zamir and Barak Medina, Law, Morality,
advocacy of the Precautionary Principle, is that CBA’s use of probability-weighted expectation values inadequately accounts for possible outcomes in which harms are especially large or uncertain.

Proponents of cost-benefit analysis have responded to these critiques in three primary ways. First, they have developed and advocated methods for quantifying and monetizing costs and benefits that are not easily monetized because they are not traded on markets. The most well-known effort of this type is probably the attempt to determine a monetized “value of statistical life” (VSL). 18 Second, where costs, benefits and probabilities cannot be monetized precisely, they propose a series of increasingly approximate methods for quantifying all costs, benefits, and probabilities to the greatest possible extent. Third, they propose that when all attempts at quantification fail, decisionmakers may take the remaining unquantifiable values into consideration in their final decisions. For convenience, in what follows I will refer to this approach, much of which was put into practice by Sunstein while he was head of the Office of Information Management, Office of Management and Budget (OIRA), as Cost-Benefit Analysis 2.0. CBA 2.0 differs from traditional CBA in that it attempts to be as inclusive as possible of societal values and is willing to sacrifice some degree of quantitative precision to do so. It employs various monetization and quantification strategies to approximate a full cost-benefit analysis as closely as possible, even in situations in which costs, benefits, or probabilities are difficult to quantify. It is roughly reminiscent of a Taylor expansion around a traditional CBA. CBA 2.0 remains true to traditional CBA, however, in its commitments to quantification and to probability-weighted aggregation of costs and benefits over possible outcomes, its translation of all goods onto a single scale, and its utilitarianism.

These commitments are reflected in OIRA’s instructions to agencies. Thus, while CBA 2.0 allows in principle for the consideration of “values that are difficult or impossible to quantify, including equity, human dignity, fairness, and distributive impacts,”19 an agency wishing to take such values into account must first attempt to monetize or quantify them.20 CBA 2.0 does not use a social welfare function that incorporates distributive effects, but treats distributive impact as an “unquantified value,”21 which can be considered only after and in light of a monetized costs-benefit calculation.


19 EO 13563, Sec. 1(c)

20 See OIRA Primer at 3, 12.

21 OIRA Primer at 7. There are various proposals for how distributional effects might be incorporated into a social welfare analysis. See, e.g., Adler, Well-Being and Fairness, supra at ___; Adler, Cost-Benefit Analysis and Distributive Weights, supra at ___.

6
CBA 2.0’s primary approach to incorporating unquantified values into cost-benefit determinations is “breakeven analysis.” Breakeven analysis “answers the question, ‘How large would the value of the non-quantified benefits have to be for the rule to yield positive net benefits?’” It can be used if an agency’s quantified cost-benefit calculation comes up short in justifying a proposed regulation. In breakeven analysis, the agency usually considers whether the unquantified benefits are weighty enough to overcome the cost-benefit deficit. The intuition justifying breakeven analysis is that even if it is not possible to say “The benefits of this proposed regulation will be approximately $X,” it may be possible to say “The benefits of this proposed regulation will (or will not) be more than $Y,” where $Y is the approximate cost of the proposal.

While breakeven analysis does not purport to quantify the unquantifiable benefits it is using to overcome a quantitative cost-benefit deficit, break-even analysis makes use of a decisionmaker’s rough, implicit sense of the magnitude of those benefits. Breakeven analysis will be ineffective when such a rough comparison either is not decisive or cannot be made. Whether that is the case depends on why a given benefit or cost has been deemed “unquantifiable.” If the unquantifiability is due primarily to uncertainty about the “magnitude” or probability of a particular cost or benefit, break-even analysis can circumvent the need to establish a particular valuation as long as a rough comparison is enough to determine whether the unquantifiable factor makes up for the difference between quantified costs and benefits. A rough comparison will not be decisive when the uncertainty on either or both sides of the ledger is great enough that one cannot estimate whether the qualitative benefits exceed the cost-benefit deficit. That may happen either because costs on different sides of the ledger are close or because costs on one side or the other are highly uncertain, perhaps because the effects of regulating (or not) are unpredictable so that the range of possible outcomes is wide. How much uncertainty can be tolerated depends on how close a question the break-even analysis is. If the uncertainty is substantial enough, a rough comparison may not provide any answers. Break-even analysis is less likely to be successful if the unquantifiability stems from serious normative controversy about the importance of the value at issue or from disagreement about how (or whether) incommensurable values, such as distributive impact, should be incorporated into the analysis.

CBA 2.0’s use of unquantified and distributive values only as a means to make up deficits in a quantitative cost-benefit analysis almost inevitably will discourage agencies from proposing regulations whose justification is based solely or even strongly on unquantifiable values. If an agency wishes to rely on non-monetized values to justify a proposed regulation, it must demonstrate that it attempted to monetize or quantify them and “explain why” it could not do so. If it seeks to rely on distributive effects, it must first analyze them – quantitatively if possible – in a “separate description.” The agency thus bears a high burden of detailed justification for any reliance on unmonetized values. Given these procedural costs and the difficulty of assessing unquantified benefits and costs, agencies have incentives not to put

---

22 See Sunstein, Limits of Quantification, supra note __
23 Beyond Cheneyism and Snowdenism at ___; OIRA Primer at 13; Sunstein, Limits of Quantification, supra note __
24 OIRA Primer at 13.
25 OIRA Primer at 7.
themselves in the position of having to defend any significant deviation from the quantitative cost-benefit result.\textsuperscript{26}

The priority given to quantitative cost-benefit assessment also seems likely to affect the types of regulatory options that agencies consider.\textsuperscript{27} In a recently published overview, Sunstein observes that “the monetized benefits exceeded the monetized costs for nearly every recent economically significant rule for which agencies monetized both benefits and costs.”\textsuperscript{28} Perhaps this is an indication of CBA 2.0’s success – i.e. that most agencies have done an extremely efficient job of monetizing all relevant considerations. A less sanguine view, however, would be that the emphasis on quantification has led to a selective focus on regulations whose quantifiable benefits outweigh their quantifiable costs. Critics have contended that the emphasis on quantification distorts the analysis, particularly in cases where non-quantified impacts are central to the purpose of the regulation, and that agency discretion as to whether to invoke non-quantified impacts reduces accountability and leaves room for strategic behavior.\textsuperscript{29}

CBA 2.0 also provides few tools for handling situations in which the likelihoods of various potential outcomes are uncertain and their likelihoods not easily quantified as probabilities.\textsuperscript{30} The OIRA Primer directs agencies that are unable to estimate probabilities for potential outcomes to “develop a central scenario … that reflects the agency’s best estimate of the likely consequences of each regulatory alternative” and then “use the benefits and costs of these estimates to approximate the expected value of the benefits and costs of each regulatory alternative.”\textsuperscript{31} To obtain a sense of the uncertainty in the cost-benefit result obtained using the “best estimate” scenario, the agency is to use other “plausible” outcome scenarios to estimate “ranges of plausible benefits, costs, and net benefits of each regulatory alternative.”\textsuperscript{32} If an agency is lucky, it will be clear which scenarios are “plausible” outcomes for each regulatory alternative and the net benefits associated with all plausible scenarios will cluster tightly around the value for the “best estimate.” In that case, uncertainty about the relative likelihoods of possible scenarios is unimportant because it cannot lead to much uncertainty about the eventual CBA determination. If the agency is unlucky, however, the plausible outcomes of the regulatory alternatives may have widely divergent net benefit values. In that case, the cost-benefit result will depend very strongly on the choice of “best estimate.” If that choice is uncertain because the relative likelihoods of plausible outcomes are unknown, CBA 2.0 provides no clear direction to the decisionmaker.

Overall, then, CBA 2.0 has come far toward accomplishing its primary goal of accounting better for benefits and costs that are not traded on markets, as long as those costs and benefits can be

\textsuperscript{26} Sometimes, of course, an agency’s statutory basis for regulation mandates a regulation that deviates from the cost-benefit analysis result. Regulators also might have normative commitments that motivate them to incorporate distributive and unquantifiable considerations. CBA 2.0 allows them to do so if they are willing to make the effort.

\textsuperscript{27} See, e.g. Lisa Heinzerling, Quality Control: A Reply to Professor Sunstein, 102 Calif. L. Rev. 1457 (2014)

\textsuperscript{28} Sunstein, Real World Cost Benefit Analysis, 114 Colum. L. Rev. 167 (2014) at 179.

\textsuperscript{29} Lisa Heinzerling, Quality Control: A Reply to Professor Sunstein, 102 Calif. L. Rev. 1457 (2014)


\textsuperscript{31} OIRA Primer at 14-15.

\textsuperscript{32} Id.
approximately quantified. It does relatively little, however, to respond to criticisms of its underlying quantitative, utilitarian framework and seems unlikely to produce decisions that rely significantly on unquantified benefits or costs. Moreover, it provides virtually no guidance for dealing with uncertainty about the likelihood of potential outcomes when the net benefits of those outcomes vary significantly.

B. Precautionary Principle 2.0

The Precautionary Principle focuses on those cases in which there is an uncertain, but plausible, possibility of severe harm from some activity if regulation is not imposed. While Sunstein and others dismiss the Precautionary Principle approach as doomed to be either incoherent or trivial, its defenders have argued that these problems can be resolved if one moves away from the straw man “strong” Precautionary Principle. Here, I explore a sophisticated Precautionary Principle framework set out in recent work by philosopher Daniel Steel, which I refer to here as PP 2.0.

1. PP 2.0 Basics

Focusing on environmental regulation, where the Precautionary Principle has had its greatest prominence, Steel argues that Sunstein’s inconsistency objection to the Precautionary Principle, along with various other objections, can be overcome by requiring that any version of the Precautionary Principle applied to policymaking meet three requirements:

- The Meta-Precautionary Principle (MPP): The MPP asserts that uncertainty should not be a reason for inaction in the face of serious threats. This principle places a restriction on what sorts of rules should be used, namely decision rules that are susceptible to paralysis by scientific uncertainty should be avoided.

- The “Tripod”: The term “tripod” refers to the knowledge condition, harm condition, and recommended precaution involved in any application of a Precautionary Principle. There are multiple ways of specifying the knowledge condition, harm condition, and recommended precaution, and careful consideration of the particulars of each application are relevant to deciding how to fill in the blanks. For example, “If it is possible that an activity might lead to irreversible harm, then that activity should be banned” is one version of PP, while “If there is some scientific evidence that an activity will lead to irreversible harm, then an alternative should be substituted for that activity if feasible” is another. Which version of PP should be used in a given application is influenced by MPP – we should avoid versions that turn scientific uncertainty into paralysis – as well as by the next and final component.

- Proportionality: Roughly, proportionality is the idea that the aggressiveness of the precaution should correspond to the plausibility and severity of the threat. I propose that proportionality be defined more precisely in terms of two subsidiary principles that I call consistency and efficiency. Consistency requires that the precaution not be recommended against by the same version of PP that was used to justify it. Efficiency states that, among

See, articles cited supra note __.

See, supra.
those precautions that can be consistently recommended, the less costly should be preferred. Consistency and efficiency place important constraints on what can be justified by PP in a given context. For example, if there is no version of PP that can consistently recommend an action (say, preemptive war) in a given context (say, a US invasion of Iraq in 2003), then PP cannot justify that action in those circumstances. Finally, MPP affects how proportionality is applied. For example, comparisons of the relative efficiency of policy options should not be done in a way that makes scientific uncertainty grounds for continual delay.35

The tripod of “knowledge condition,” “harm condition,” and “recommended precaution” specifies a particular version of the Precautionary Principle (“PP version”), which can be tested for consistency and efficiency under the proportionality principle. Note that specifying the tripod conditions requires that one assume a baseline36 state in which one determines the state of knowledge, from which one measures harm, and to which one might apply a regulation. To elaborate on what the tripod conditions mean, a few examples are useful. In the environmental context, Steel provides these three examples of PP versions, which I have annotated to indicate the knowledge condition, [K], harm condition [H], and regulation [R] associated with each version:

1. If there is [K] some scientific evidence that an activity, A, leads to [H] a significant and irreversible harm, then [R] an alternative should be substituted for activity A if feasible.

2. If [K] a scientifically plausible mechanism exists whereby an activity A can lead to [H] a catastrophe, then [R] that activity should be phased out or significantly restricted.

3. If it [K] is possible that an activity A will lead to [H] a catastrophe, then [R] that activity should be prohibited.37

The general form of PP that Steel discussed thus can be written as:

PP: If [K] implies that A will lead to [H], then impose [R], where A is the regulated activity and [K], [H], and [R] are the knowledge condition, harm condition, and proposed regulation, respectively.

While the possible combinations of knowledge condition, harm condition, and regulation are virtually endless, the proportionality principle constrains the possible PP versions by requiring consistency and efficiency. For any PP version, consistency requires that the regulation itself, when taken as an activity, must be permissible under that PP version’s knowledge and harm conditions. In other words, the precaution must survive the consistency test:

CR: If [K] implies that [R] will lead to [H], then do not impose [R].

35 STEEL, supra note __, at 9-10. (citations omitted).
36 See also, OIRA Primer at 4-5, discussing the need to define a baseline for cost-benefit analysis.
37 STEEL, supra note __.
If PP directs that [R] should be imposed and CR direct that it should not, then PP is not a consistent precautionary principle version. If there is more than one consistent PP version, the one that is the most “precautionary” should be selected.

Importantly, consistency requires that the same knowledge condition and harm condition be used to evaluate both the harm prevented by [R] and the harm caused by [R]. Only in this way can PP 2.0 rebut the incoherency accusation of its critics. This requirement says something about the generality of the harm condition. Though harm conditions can vary by severity, they must be stated generally enough to account for all of the harms that might flow from either the activity or the regulation. In this sense, PP 2.0 does not necessarily relieve us of the need to compare very different kinds of harm and sometimes may encounter difficulties similar to those that plague CBA 2.0 when unquantifiable harms are significant in a decision-making context.

PP 2.0 may be less likely than CBA 2.0 to be undermined by the difficulties of comparing unquantifiable harms, however, precisely because it focuses on substantial harms and does not insist on quantification. PP 2.0 shares some of the advantages of break-even analysis. Indeed, it is analogous to conducting a pair of qualitative break-even analyses, asking first whether the potential harms of activity A are “big enough” to be worth regulating and then whether the potential harms of regulation [R] are “too big” to incur. Even though the harm condition threshold is not a quantitative point of comparison, separating the analysis up in this way still may be useful in getting decisionmakers out of the quagmire of trying to compare two unquantifiable harms. The focus on substantial harms may also help to ease the difficulties in comparing very different harms, especially where there is a lot of uncertainty. It may simply be easier to conceptualize and compare the harms associated with plausible scenarios of a certain severity and uncertainty than to attempt to assess and compare the average harms over a whole range of more and less likely scenarios. In some situations, the distribution of potential harms from regulation may be tightly clustered and relatively certain, while the distribution of harms from the activity of concern may be broad, uncertain, or “fat-tailed.” If that is the case, it may be hard to compare average expected costs and benefits, but easy to see that the significant potential harms from the activity outweigh the significant potential harms from regulation. (Or vice versa, of course.) Thus, while PP 2.0 is not immune from problems of comparing unquantifiable, incommensurable, or distributive harms, it ameliorates those problems in some decision-making contexts.

We can rephrase Sunstein’s argument that the Precautionary Principle is incoherent (or trivial) as an assertion that there is no consistent PP version that does not reduce essentially to CBA (or, perhaps CBA 2.0). That clearly is not the case for PP 2.0. Of course, the consistency requirement limits the situations in which PP 2.0 can be applied, meaning that it, like cost-benefit analysis, leaves society vulnerable to low probability catastrophes that would be simply too costly to ameliorate. There also will be some circumstances in which PP 2.0 and CBA 2.0 will agree and others in which neither is useful in coping with the decision-making challenge at hand.

2. When Does PP 2.0 Favor Regulation?

Despite their differences, CBA 2.0 and PP 2.0 sometimes will lead to the same eventual conclusion about whether regulation is advisable. To determine whether PP 2.0 is a potentially
useful tool, we must investigate the circumstances under which PP 2.0 favors regulation, while CBA 2.0 does not. To begin, it is helpful to rewrite PP 2.0’s consistency requirement as:

PP: If \( C(\{A\},K) > [H] \), impose \([R]\)
CR: If \( P(R,K) > [H] \), do not impose \([R]\)

Here, \( C(\{A\},K) \), represents the maximum harm that we expect, with certainty \([K]\), activity \( A \) to cause if it continues unregulated, while \( P(R,K) \) represents the maximum harm that we predict, with certainty \([K]\), to result from imposing regulation \([R]\). We can then represent the consistency requirement for imposing \([R]\) as:

If \( P(R,K) < [H] < C(\{A\},K) \), impose \([R]\)

We also can write CBA 2.0’s condition for imposing \([R]\) as:

If \( EC(R) < EC(A) \), impose \([R]\),

where \( EC(A) \) and \( EC(R) \) are the probability-weighted expectation values of the net costs of activity \( A \) and regulation \([R]\), respectively.

There are three sources of disagreement between PP 2.0 and CBA 2.0. First, PP 2.0 and CBA 2.0 may value the harm associated with a given possible outcome differently. Such discrepancies are most likely in situations where unquantifiable, incommensurate, or distributive harms are likely to be significant. Second, even if one uses the same value metric, the PP 2.0 and CBA 2.0 criteria for imposing regulation use different ways of aggregating over possible outcomes. CBA 2.0 uses the probability-weighted expectation value, while PP 2.0 uses the maximum harm expected with a given degree of certainty. Third, PP 2.0 requires a minimum degree of harm before imposing regulation, while CBA 2.0 would impose it whenever the costs of activity \( A \) exceed its benefits.

To investigate how these sources of discrepancy affect the results of the analysis, first imagine using the CBA 2.0 methodology to value the harm associated with each potential outcome scenario in a PP 2.0 analysis. Using the CBA 2.0 metric, \( C(\{A\},K) \) is the maximum net cost that we expect, with certainty \([K]\), activity \( A \) to cause, while \( P(R,K) \) is the maximum net cost that we expect, with certainty \([K]\), regulation \([R]\) to impose. Maximum net costs will, of course, be greater than the probability weighted expectation value for net costs, so we can write:

\[
C(\{A\},K) = EC(A) + \Delta(A,K) \\
P(R,K) = EC(R) + \Delta(R,K)
\]

Here \( \Delta(A,K) \) is the amount by which the maximum net cost of activity \( A \) exceeds its average and \( \Delta(R,K) \) is the amount by which the maximum net costs of regulation \([R]\) exceeds its average.

Thus, when harms are valued according to the CBA 2.0 methodology, we can write the PP 2.0 consistency condition for imposing regulation \([R]\) as:
PP 2.0: If $EC(R) + \Delta(R,K) < EC(A) + \Delta(A,K)$ and $[H] < EC(A) + \Delta(A,K)$, impose $[R]$

There are now four possible circumstances: i) both approaches recommend in favor of regulation, ii) CBA 2.0 recommends against regulation and PP 2.0 is neutral, iii) CBA 2.0 favors regulation, while PP 2.0 is neutral, and iv) CBA 2.0 opposes regulation, while PP 2.0 favors it. (By its nature, PP 2.0 cannot be said to “oppose” regulation, but only to favor it or remain neutral.) Note that all four of these situations are possible (which may surprise those who assume that the Precautionary Principle is always more regulation-friendly than cost-benefit analysis). However, the only situation that properly can be characterized as “disagreement” is iv), when PP 2.0 favors regulation $[R]$, while CBA 2.0 opposes it. If we define $D = EC(R) - EC(A)$, then the condition for disagreement is:

$D + \Delta(A,K) > \Delta(R,K)$, where $D > 0$ and the harm condition is met.

Figure 1 illustrates this scenario, which occurs only when the spread in potential harms from activity A is sufficiently greater than the spread in potential harms from regulation $[R]$.

There are several factors that affect whether PP 2.0 matters, in that it favors regulation that would not be imposed under CBA 2.0:

*Harm condition severity.* PP 2.0’s harm condition puts a cutoff on the circumstances under which PP 2.0 favors regulation. The higher the harm condition, the less likely PP 2.0 will favor regulation.

*Size of cost-benefit difference between imposing and not imposing $[R]$.* Disagreement between CBA 2.0 and PP 2.0 becomes more likely as $D$, the cost-benefit difference between imposing $[R]$ and allowing activity A to continue, becomes smaller. Note that as $D$ becomes smaller,
insistence on sticking to the CBA 2.0 result when PP 2.0 favors regulation becomes less and less justifiable as long as there is some uncertainty in the cost-benefit calculation.

Permissiveness of knowledge condition. If the probability distribution of harms is known with reasonable certainty, the knowledge condition acts essentially as a cut-off on the probability distribution. Thus, the more rigorous the knowledge condition, the smaller both $\Delta(A,K)$ and $\Delta(R,K)$ will be. For very rigorous knowledge conditions, CBA 2.0 and PP 2.0 will never disagree (as long as harm is valued by the same metric). Disagreement is more likely for more permissive knowledge conditions. Thus, PP 2.0 is more likely to matter when the knowledge condition is more permissive. However, the overall effect of a more permissive knowledge condition is ambiguous, since $\Delta(A,K)$ and $\Delta(R,K)$ both grow as the knowledge condition becomes more permissive. The relative rates at which they increase depends on the shapes of the probability distributions.

Uncertainty about probability estimates. If our knowledge about the probability distribution of outcomes is also uncertain, there is more potential for disagreement because the knowledge condition and the probability distribution no longer measure the same thing. CBA 2.0’s probability-weighted expectation calculation is insensitive to uncertainty about whether the probability estimate is correct. The knowledge condition, however, incorporates both the probability that a harmful outcome will occur and our lack of certainty about the probability estimate. The result is that uncertainty about the probability distribution will increase the harm considered by PP 2.0 for any given knowledge condition. For example, if it is known that the worst plausible outcome of an activity will occur with .1% probability, a PP 2.0 analysis applying a knowledge condition of “small, but not remote possibility” might not take it into account. If, however, the uncertainty as to that probability estimate is great, ranging perhaps for 0 to 1%, then a PP 2.0 analysis applying the “small but not remote possibility” might very well take it into account. Uncertainty about the probability of harm resulting from activity A unambiguously makes it more likely that PP 2.0 will favor regulation. However, if the probability of harm resulting from regulation [R] is also uncertain, the net effect will depend on the relative degree of uncertainty and also on how close the maximum possible harms from A and from [R] are to the harm condition. Such sensitivity to the relative uncertainty of harms associated with regulation and the regulated activity is, of course, a feature of PP analysis, not a bug.

Method of valuing harm. So far, this section has assumed that “harm” in both CBA 2.0 and PP 2.0 is assessed using a cost-benefit value metric. The range of potential disagreement between the two methods widens if some other value metric is used to assess harm. Different methods for assessing harm can be an extremely significant source of disagreement between CBA 2.0 and PP 2.0, particularly in contexts in which unquantifiable, incommensurable, or distributive values play a significant role.

---

C. Different Cures for Different Ailments

Both cost-benefit analysis and the Precautionary Principle aim to protect against decision-making mistakes, but they focus on different sources of error. Cost-benefit analysis has its roots in a concern that agencies would fail to account properly for the economic costs of regulation. CBA 2.0 broadens cost-benefit analysis to incorporate a wider range of potentially neglected regulatory benefits and costs, in part to respond to the criticism that cost-benefit analysis neglects non-economic costs, but also to help regulators avoid cognitive and other biases. As Sunstein puts it, CBA “reduces (without eliminating) the twin dangers of selective attention and motivated reasoning.”

Cost-benefit analysis also attempts to provide democratic accountability (perhaps the central concern of administrative law). By setting out a step-by-step approach to accounting for costs and benefits, it promotes agency decision-making that is systematic, well-reasoned, and transparent to other government officials, to regulated parties, and to the public. Centralizing review of cost-benefit analyses in OIRA, thus giving more power to the politically accountable parts of the executive branch, is another means of promoting democratic accountability. For example, one of the most common value trade-offs confronting regulatory agencies is that between life and money. This trade-off has spawned a large literature dealing with the methodology for computing the “value of statistical life” (VSL). The trend also has been to harmonize the value of statistical life across agencies, thus removing this very common normative assessment from the purview of individual agencies.

Because of the particular decision-making flaws it is intended to address, cost-benefit analysis is centrally devoted to monetization and quantification and its methods flow from that core commitment. Thus, while the choice of a monetized value metric may originally have been driven by anti-regulatory ideology, more regulation-friendly CBA 2.0 proponents remain committed to it primarily because they are committed to the virtues of quantitative analysis. With its fundamental commitment to quantification, cost-benefit analysis is weakest when decisions are centrally dependent on values that cannot easily or convincingly be monetized or quantified or when faced with severe unpredictability and uncertainty. While CBA 2.0 permits such values to be “considered,” it retains these weaknesses.

Unfortunately, while quantification is a very successful way to mitigate some decision-making flaws, it is unhelpful for dealing with others, such as trade-offs between competing normative values, incommensurability of various sorts, and unpredictability. If these issues are centrally important to a decision, CBA 2.0’s dominant commitment to quantification may lead it astray. Precisely because of its commitment to quantification, CBA 2.0 also demands considerable, sometimes even heroic, technical expertise. For these reasons, it functions best in institutional settings in which trade-offs between competing values and issues of incommensurability have been settled by other means. Of course, some normative trade-offs are made in advance, by the legislature, and enshrined in the statutory authorizations for agency action. One might hope that most of the core normative issues would be handled in this way. Unfortunately, core normative trade-offs often cannot be made at a high level of generality. Indeed, it may be impossible to understand, or even pose, the normative questions without technical expertise or information that

39 Beyond Cheneyism and Snowdenism at 3.
40 See, e.g. OIRA Primer at 10 (“current agency practice provides a VSL ranging from roughly $5 million to $9 million per life”)
resides in the agencies. Moreover, while CBA 2.0’s commitment to systematization and transparency goes at least some distance toward putting normative trade-offs into the hands of democratic processes, it does much less to mitigate problems of unpredictability and uncertainty, some of which are inherent in the very goal of predicting long-term outcomes of complex systems.  

The Precautionary Principle’s core concern is not quantification, but the fear that decisionmakers will neglect uncertain, but potentially severe, harms. It responds to that concern with the relatively straightforward prescription that, when an activity may cause substantial harm, one should err on the side of caution by regulating it, even when the harm is uncertain to occur. A simple thought experiment demonstrates that uncertainty, rather than quantification, is at the heart of the Precautionary Principle. Imagine an activity that might cause severe harm X, where X is easily and accurately valued and monetized. Now suppose that the likelihood that the activity will cause X is highly uncertain. Surely, the Precautionary Principle could be applied to argue for regulation of activity A in this situation, even though X is easily quantified. Indeed, unlike cost-benefit analysis, which is wedded to quantification, wealth maximization, and utilitarianism because of the particular decision-making flaws it aims to address, the Precautionary Principle is not fundamentally committed to any particular metric for valuing the severity of harm. In fact, unquantifiability makes application of PP 2.0 consistency and efficiency requirements more difficult.

Nonetheless, proponents of the Precaution Principle often criticize CBA’s emphasis on quantification or its utilitarian approach to social welfare. The reason for this correlation of critiques is that the difficulty in predicting and quantifying harm related to certain kinds of values can be an important source of uncertainty about the likelihood that a given activity will result in severe harm. CBA’s inability to deal with uncertainty and its commitment to quantification thus combine to make it susceptible to the Precautionary Principle’s fundamental concerns about unpredictability and uncertainty.

The weakness of the Precautionary Principle approach, and the heart of Sunstein’s objection to it, is that regulation aiming to prevent such harmful, but uncertain, outcomes not only consumes resources that could be used to address more predictable harms, but may cause similarly severe harms. PP 2.0’s consistency requirement addresses this concern by ensuring that a regulatory measure that is equally or more certain than the regulated activity to cause a certain level of harm is rejected. The Precautionary Principle is of no use, however, if it is not consistent for some meaningful and practical range of knowledge and harm conditions.

Summing up, the choice between CBA 2.0 and PP 2.0 involves a mixture of normative and practical issues, including: i) the extent to which relevant costs and benefits can be quantified in a rational and normatively acceptable way, including whether the institutional context provides

the necessary technical expertise; ii) the importance of distributive impact, incommensurability, and unquantifiable values to the regulatory context; iii) the extent to which potential outcomes can be predicted and the degree of uncertainty as to their likelihood of occurring; iv) whether the regulated activity and/or the proposed regulation have the potential to cause serious harm; and v) which potential flaws in government decision-making are of predominant concern in the specific context.

III. CBA 2.0 in a Counterterrorism Surveillance Context

The counter-terrorism context poses daunting challenges to any cost-benefit comparison.\footnote{See, e.g. Kenneth Anderson, The Assumptions Behind the Assumptions in the War on Terror: Risk Assessment as an Example of Foundational Disagreement in Counterterrorism Policy, 54 Wayne L. Rev. 505 (2008); Beyond Cheneyism and Snowdenism at 12-13.} Monetizing the “cost” from a terrorist event (or the “benefit” of preventing one) involves grappling not only with complicated questions about how to value life and physical injury, but also assessing other possible impacts, from the economic to the psychological and social. Moreover, the connection between surveillance and terrorism prevention is attenuated and hard to predict. Monetizing costs arising from surveillance is, if anything, even more difficult, since the effects of surveillance range from intrusions into individual privacy to significant distributive impacts to society-wide effects on liberty, autonomy, and democratic participation. As if these difficulties were not enough, the outcomes like to result from undertaking (or declining to undertake) any particular surveillance power are highly unpredictable and uncertain, so that it would be difficult, if not impossible, to aggregate the possible costs and benefits to compute probability-weighted expectation values. Comparisons between surveillance harms and surveillance benefits also raise controversial issues the normative value of preventing terrorism and preserving privacy and civil liberty.

In Beyond Cheneyism and Snowdenism, Sunstein readily acknowledges that a full quantitative CBA 2.0 analysis is virtually impossible in light of these difficulties. His main contention is that the Precautionary Principle is the wrong way to cope with the difficulties because of its purported incoherence. As an alternative, to avoid the “pathologies” of the precautionary approach, he suggests four ideas for how to handle the counterterrorism surveillance issue: break-even analysis, avoiding gratuitous costs, avoiding illicit grounds for surveillance, and avoiding the “worst of the worst” cases:

Break-Even Analysis. As discussed above, break-even analysis is a part of CBA 2.0, to be used to determine whether the results of quantitative cost-benefit analysis should be over-ruled by unquantifiable benefits or costs. Sunstein is fairly pessimistic about the prospects for break-even analysis in the counterterrorism surveillance context, where “hard-to-quantify costs are on both sides of the ledger” under “the reasonable assumptions that a great deal of important information is missing, and that moral valuations will play an inescapable role.”\footnote{Beyond Cheneyism and Snowdenism at 12-13.}

As discussed above, PP 2.0 is somewhat akin to break-even analysis. That similarity is both good and bad for PP 2.0’s prospects in the counterterrorism surveillance context, since it means that PP 2.0 and break-even analysis suffer from some of the same difficulties in coping with unquantifiable harms. PP 2.0 sometimes can be more powerful than break-even analysis, in
coping with uncertainty, however, so it is worth considering as an alternative. (Of course, PP 2.0’s normative underpinnings also are somewhat different, given the priority it puts on avoiding substantial and uncertain harms, so one might object to it on those grounds.)

At a minimum, as we will see, PP 2.0 has some of the same advantages that Sunstein cites for breakeven analysis in promoting systematic analysis and transparency (at least to anyone who the opportunity to peruse the analysis).

Avoid Gratuitous Costs. By gratuitous costs, Sunstein means harms imposed by surveillance that produces no or de minimis benefits. The proposal to avoid gratuitous costs would seem almost to go without saying (whether one is engaging cost-benefit analysis or any other reasoned decision-making process). Understanding why this principle might have some bite depends on recognizing two key points. First, the principle can be applied granularly. The question is not only, for example, whether the total benefits of some general category of surveillance are de minimis. If modifying the activity would make it less costly with only a de minimis reduction in its benefits, the modification would “avoid gratuitous costs.” Sunstein exemplifies how the avoid gratuitous costs principle would justify an incremental reduction in surveillance powers by referring to the President’s Review Group’s proposal that bulk telephony metadata should be held by telecom providers, rather than by the NSA. The Group concluded that storing copies of the data on the NSA’s servers provided de minimis incremental terrorism prevention benefits over allowing the NSA to access data stored on telecom company servers. In light of the de minimis incremental benefits, the privacy and civil liberties risks associated with NSA storage were gratuitous costs.

Upon further reflection, however, the “avoid gratuitous costs” principle is more complicated than it first appears because any practical version of the principle relies on a comparison of costs and benefits, rather than a straightforward determination that benefits are de minimis on some absolute social value scale. To illustrate this point, let’s return to the telephony metadata example. According to the NSA, the benefits associated with its storage of telephony metadata arose from more convenient access to the data. NSA presumably believed that those benefits justified its internal costs of hosting the data. If those had been the only costs presumably no one would have questioned the NSA’s decision to host the data itself. What made the benefits of NSA hosting de minimis, was that they paled in light of the potential social harm from leaving so much personal data in the NSA’s hands. Surveillance costs are gratuitous when their associated benefits are de minimis in comparison to those costs.

But if gratuitous costs are simply those that far outweigh their net benefits, the “avoid gratuitous costs” principle reduces to breakeven analysis and determining whether costs are gratuitous would, in principle, require some assessment of all costs and benefits (including, presumably, the costs of the NSA servers, the benefits of convenience to NSA employees, and so forth). This may what Sunstein has in mind. There is another possible interpretation, however, that picks up the sense of “gratuitous” as “unjustified.” Under this alternative interpretation, the only benefits that count are improvements in terrorism prevention. We deem the surveillance costs of NSA holding telephony metadata for the entire population gratuitous not (or not only) because we believe that potential privacy and civil liberties harms should be traded for minor improvements in counterterrorism capability. As demonstrated in Part IV, this version of the “avoid gratuitous
costs” principle can be derived from a consistent PP version.

_Avoid Illicit Grounds for Surveillance._ According to this principle, certain surveillance activities are to simply off-limits and should not “count in the balance at all.” For example, “surveillance that is designed to reduce risks to national security should not be designed to protect against criminal activity that raises no national security issue” and “surveillance should not be designed to give a commercial advantage to American firms.” Also, “surveillance cannot legitimately be used to punish people because of their political views or their religious convictions.” This principle directly implements moral prohibitions and thus has no particular connection to CBA 2.0. It is invoked as a kind of common ground to which both “Cheneyists” and “Snowdenists” should be able to agree. But its scope is unclear. There may well be disagreement about what grounds for surveillance are illicit from a moral perspective. In particular, if avoiding illicit grounds means only that surveillance cannot be undertaken if it is based solely on illicit grounds, the principle will be of little, if any, use where motivations are mixed, effects are unintended, or information is obtained as a byproduct of properly motivated surveillance activities. While Sunstein suggests that we can “imagine cases that might test the clarity of the line,” my own intuition is that much of the real world controversy about surveillance surrounds exactly such cases. As I discuss in Part IV, PP 2.0 produces a version of the “avoid illicit grounds” principle that provides some insights into how its scope should be determined.

_Avoid the Worst of the Worst Cases._ The concern about the “worst of the worst” cases is one of the motivations for the Precautionary Principle. Indeed, PP 2.0 provides one mechanism for aggregating over potential outcomes in a way that prioritizes the prevention of unlikely, but substantial, harms. Sunstein considers another approach, which is to apply the “maximin principle,” which directs decisionmakers to “choose the policy with the best worst-case outcome.” The maximin principle has well-known problems, however. It is not always easy to decide which outcome is the “worst.” Sunstein’s discussion illustrates this problem. He begins by asserting “the worst case associated with a successful terrorist attack is so much worse than the worst case associated with a breach of personal privacy” that maximin provides a good case for Cheneyism, but later observes that pervasive surveillance can threaten “both liberty and self-government themselves,” leading to a worst case that is “very bad indeed.” Maximin’s problems in the counterterrorism surveillance context are particular severe – and not only because of the difficulties of comparing worst cases. To see why, consider a situation in which the policy options under consideration are i) carry out surveillance program A or ii) do not carry out surveillance program A. Unless surveillance program A is completely comprehensive, or is sure to eliminate the possibility of some particularly horrible type of attack, the worst-case possible outcomes for scenarios i) and ii) are the same – a severe terrorist attack! Presumably not

---

44 Beyond Cheneyism and Snowdenism at __.
46 Beyond Cheneyism and Snowdenism at __.
47 See STEEL, supra, Ch. 3.3 for further discussion of the relationship between maximin and the Precautionary Principle.
48 This is a rather breathtaking assertion, the validity of which must, at the least, depend on what specific surveillance power is under consideration, what kind of surveillance costs it imposes, and what kind of terrorist attack might occur as a result of modifying or suspending that particular surveillance power.
even the most die-hard Cheneyist would want to impose the level of surveillance needed to “prevent” the worst case scenario. And, of course, doing so would usher in the worst case scenario collapse of liberty and self-governance.\textsuperscript{49}

Not surprisingly, in light of its many problems, Sunstein does not endorse the maximin approach. Instead, he concludes that “the best that can be done is to attempt to identify safeguards, with respect to privacy, that plausibly reduce the risks associated with worst-case scenarios, while also allowing officials to do what must be done with respect to protecting national security.” While Sunstein is hopeful that this goal might be easier to achieve in practice than it seems in the abstract, I fear it is question-begging in many situations that count and provides no principle by which to determine how much “must be done” to protect national security or how important it is to “reduce the risks” with respect to privacy. While PP 2.0 is not a magic bullet that does away with hard normative issues, it is a systematic approach that may take us further toward acceptable solutions.\textsuperscript{50} In any event, that is the possibility that Part IV begins to explore.

\textbf{IV. PP 2.0 in a Counterterrorism Surveillance Context}

In \textit{Beyond Cheneyism and Snowdenism}, Sunstein sets out two extreme perspectives that he associates with the Precautionary Principle. The first, which he dubs “Cheneyism”, holds that:

The world has become an unprecedentedly dangerous place. Terrorist threats are omnipresent. As the 9/11 attacks display, numerous people are prepared to engage in terrorism, and they sometimes succeed. In particular, they want to kill Americans. The first obligation of public officials is to keep the citizenry safe. To do that, the best methods may well involve widespread surveillance both domestically and abroad. If the result is to save lives, it is worth it. Even when the probability of harm is low, and even if government is operating in the midst of grave uncertainty, it is appropriate to do whatever must be done, and whatever technology allows, to prevent deaths and to protect the nation, even or perhaps especially from worst-case scenarios.\textsuperscript{51}

The second, which he dubs “Snowdenism,” holds that:

Americans face unprecedented threats from their own government. In the aftermath of the 9/11 attacks, the United States has seen the rise of a massive and (at least until recently) mostly secret security apparatus, involving the collection of vast quantities of data involving the communications of ordinary people. Personal privacy is now at serious risk, and the same is true of free speech. “Trust us” is never an adequate response to citizens’ legitimate concerns. We need to create aggressive safeguards to protect civil liberties not only now but also for periods in which government is in especially bad hands -- and to create precautions against the evident dangers, including worse-case scenarios.\textsuperscript{52}

\textsuperscript{49} This is not to suggest that there is some kind of inevitable or monotonic trade-off between liberty and security. Maximin, however, applies at the extreme.
\textsuperscript{50} See \textit{Steel}, supra note __, at __ for a more general discussion of how PP 2.0 avoids falling into the maximin trap.
\textsuperscript{51} \textit{Beyond Cheneyism and Snowdenism} at 2.
\textsuperscript{52} Id. at 2.
These extreme perspectives are a useful starting point for this Part’s exploration of how PP 2.0 avoids the incoherence trap associated with similarly extreme versions of the Precautionary Principle.

A. General Form of a Precautionary Principle for Counterterrorism Surveillance

To apply Steel’s PP 2.0 framework to the counter-terrorism/surveillance context, I begin by considering the choice of Meta-Precautionary Principle. In the environmental context, the MPP asserts that “uncertainty should not be a reason for inaction in the face of serious environmental threats.” The role of the MPP is to “place a restriction on what sorts of rules should be used for [making public policy], namely decision rules that are susceptible to paralysis by scientific uncertainty should be avoided.” In the environmental context, the relevant uncertainty is scientific uncertainty about the harms that may result from some private economic activity.

What is the appropriate MPP for counterterrorism surveillance? Sunstein suggests that both Cheneyism and Snowdenism are Precautionary Principles and, at first glance, it seems that we could frame an MPP emphasizing either threats from terrorism or threats from over-reaching surveillance, i.e.:

Cheneyist MPP: Uncertainty should not be a reason for constraining counterterrorism surveillance in the face of serious terrorist threats.

Snowdenist MPP: Uncertainty should not be a reason for permitting counterterrorism surveillance activity in the face of serious threats to privacy, civil liberties, or other important values.

On further reflection, however, it becomes evident that these framings are not equally valid bases for a precautionary principle. The “inaction” in the general version of the MPP is failure to regulate a potentially harmful activity. The regulation under consideration is some constraint on that activity. Here, the regulations at issue are constraints on surveillance, not constraints on terrorism. Hence, the regulated “activity” is surveillance, not terrorism. Indeed, it would not be sensible to think of surveillance as a “regulation” of terrorism. Terrorism is not an activity to be regulated; it is a harm to be prevented. In fact, it is precisely the harm that may result from over-regulation of surveillance. Consistently with this analysis, the uncertainty that makes it difficult to decide how much to regulate surveillance concerns the effects – both beneficial and harmful – of surveillance.

There are also normative reasons to adopting a “Snowdenist” (in perspective, at least) Meta-Precautionary Principle. Precautionary principles are normatively justified (if at all) when we anticipate that government decision-making is likely to be flawed because officials may neglect to pay sufficient attention to serious harms that might result from a particular activity. There is every reason to think that this is the case for the social harms of counterterrorism surveillance. The historical record (including our recent post-9/11 experience) teaches that, in the counterterrorism context, we are far more likely to neglect the harms and risks associated with over-reaching surveillance than we are to fail to engage in sufficient surveillance to address the
threat. Surveillance tends to be under-constrained for several reasons analogous to the usual rationales for regulation and the usual reasons for government failure. The “regulated entities” in this context are counterterrorism officials whose primary goal is to prevent terrorism, making them likely to focus myopically on that task. They directly internalize the benefits when surveillance succeeds in preventing terrorism events, but, like private sector regulated entities, are unlikely to account adequately for surveillance’s harms. Democratic oversight also is unlikely to reflect the full scope of those harms. Unlike the dramatic and salient harms of terrorism, surveillance’s harms are diffuse and accumulate slowly and incrementally as the scope of surveillance grows. The secrecy surrounding counterterrorism surveillance also makes it very difficult for outsiders to assess officials’ claims about the need for surveillance authority. Moreover, the worst effects of over-reaching surveillance tend to be visited on ethnic, religious, or political minorities, while the benefits are enjoyed by the majority. This is especially the case when surveillance constraints are loosely targeted, leaving considerable discretion in the hands of counterterrorism officials.

Surveillance harms also are uncertain, but potentially large. Its effects are felt, and linked, in complex ways, at individual, group, and societal levels. Thus, the potential for abuse or misuse of information grows non-linearly because information’s meaning depends on context, and hence on the baseline of information that already is available. Many of the most serious potential surveillance harms, such as chilling effects, destruction of community trust, and changes in social norms, are socially produced and potentially subject to tipping points and cascades. These uncertainties feed into the tendency for decisionmakers to account inadequately for surveillance harms.

Uncertainty about whether constraining surveillance powers will lead to terrorism events certainly exists, but it does not have a tendency to lead to under-investment in surveillance. In fact, uncertainty about surveillance’s effectiveness in preventing terrorism may lead (and often has led in the past) to overly expansive surveillance or, in the precautionary principle parlance, to insufficient regulation.

For all of the above reasons, in the discussion that follows, I treat surveillance as the regulated “activity” and consider a Meta-Precautionary Principle of the form:

MPP: Uncertainty about surveillance harm should not be a reason for granting counterterrorism surveillance powers in the face of serious threats to privacy, civil liberties, or other important values.

Given this version of the Meta-Precautionary Principle, the most general form of PP 2.0’s consistency requirement for counterterrorism surveillance would be:

\[ \text{MPP: } \text{Uncertainty about surveillance harm should not be a reason for granting counterterrorism surveillance powers in the face of serious threats to privacy, civil liberties, or other important values.} \]

\[ \text{Given this version of the Meta-Precautionary Principle, the most general form of PP 2.0’s consistency requirement for counterterrorism surveillance would be:} \]

---

53 Note, for example, that analysis of why the U.S. failed to stop the 9/11 attacks does not suggest that the problem was insufficient surveillance. See, e.g. 9/11 Commission Report at 254-63.

54 For discussion of this point, see, e.g., Anya Bernstein, The Hidden Costs of Terrorists Watch Lists, 61 Buffalo L. Rev. 461 (2013);

55 See STEEL, supra note__, Ch. 4 (making a similar argument in the context of environmental regulation).
PP: If there is [K] that counterterrorism surveillance activity A will cause at least [H], [R] should be imposed.

CR: If there is [K] that regulation [R] will cause at least [H], then regulation [R] should not be imposed.

In the surveillance context, the regulations available to us are not direct constraints on the amount or type of surveillance conducted by counterterrorism officials. Rather, the regulations [R] at our disposal are constraints on the suite of surveillance powers or authorities available to government officials. Thus, we might rewrite the precautionary principle and consistency requirement as:

PP: If there is [K] that the exercise of surveillance powers {A} will cause at least [H], constraint [R] should be imposed on those powers.

CR: If there is [K] that constraint [R] on the exercise of surveillance powers {A} will cause at least [H], constraint [R] should not be imposed.

Very generally, we can describe a constraint on surveillance powers [R] as not granting a certain subset of surveillance powers {A} on top of a baseline set {A′ - A}. (Any modification of a surveillance power can be described as the elimination of an incremental surveillance power.) Because surveillance harm is cumulative (and accumulates non-linearly), the additional harm associated with the government’s exercise of the subset of surveillance powers {A} depends on what baseline surveillance powers are available. In assessing the harm potentially caused or alleviated by eliminating {A}, we thus need to keep the baseline in mind.

In the counterterrorism context, the potential harm caused by not allowing the government to exercise surveillance powers {A} is failure to prevent terrorism events. We thus can describe the effect of insufficient surveillance authorities more intuitively (and equivalently) in terms of failing to prevent terrorism harm. (Again, the relevant terrorism harm arises only from terrorism events that would not be prevented by the exercise of baseline surveillance powers.)

Putting these points together, we have:

MPP: Uncertainty about surveillance harm should not be a reason for granting counterterrorism surveillance powers in the face of serious threats to privacy, civil liberties, or other important values.

PP: If there is [K] that the exercise of surveillance powers {A} over and above baseline surveillance powers {A′-A} will cause at least [H] additional harm, prohibit the exercise of {A}.

CR: If there is [K] that surveillance powers {A} will prevent at least [H] additional harm not prevented by baseline surveillance powers {A′ - A}, permit the exercise of surveillance powers {A}.

56 In making this statement, I assume, for the reasons discussed in Part IV.B, that neither collateral prevention of non-terrorism-related harm nor prevention of supporting activities are directly cognizable as harm in this analysis.
ER: Any surveillance power permitted under MPP and PP (subject to CR) should be deployed as effectively as possible to prevent unnecessary costs.

Using the notation established in Part II.B, we can also rewrite the consistency requirement as:

PP: If \( C(\{A\}, K) > [H] \), prohibit \( \{A\} \)
CR: If \( P(\{A\}, K) > [H] \), permit \( \{A\} \)

Here \( C(\{A\}, K) \) is the additional surveillance harm, assessed under knowledge condition \([K]\), that would be caused by the exercise of surveillance powers \( \{A\} \), while \( P(\{A\}, K) \) is the additional terrorism harm, assessed under knowledge condition \([K]\), that would be prevented by the exercise of surveillance powers \( \{A\} \).

If the predicates to both PP and CR are true, then PP is inconsistent. To consistently prohibit the exercise of surveillance powers \( \{A\} \), we therefore must have

\[ C(\{A\}, K) > [H] \] (predicate to PP is true) and \[ P(\{A\}, K) < [H] \] (predicate to CR is false).

B. Some Definitions and Other Preliminaries

1. Causation and Prevention

To proceed further, we need a handle on what it means for the exercise of a set of surveillance powers to “cause” or “prevent” harm. Here, we include all “but-for” causation and prevention of all cognizable harms that are possible in light of current scientific (including social scientific) understanding and available information,\(^\text{57}\) whether direct or indirect, intended or unintended. Note that the concepts of causation and prevention here do not incorporate any assessment of likelihood. Assessment of likelihood is incorporated into the knowledge condition.

2. Cognizable Harm

The meaning of “harm” also warrants some discussion. Cognizable “harm” encompasses total social harm, with certain exceptions discussed below. Thus “harms” caused by surveillance might include feelings of “creepiness,” identity theft resulting from data breaches, stalking or other malicious uses of personal information, chilling effects, breakdown of community trust, deleterious changes in social norms regarding speech, association, religious practice, or any other legitimate personal or social activities, harms to dignity, autonomy or liberty, negative effects on democracy, unfair distributive effects, and so on. Surveillance “harms” also include consequences of mistaken or over-inclusive counterterrorism actions undertaken based on

\(^{\text{57}}\), supra note __ at __. Note that cost-benefit analysis also requires the construction of a set of potential outcomes, to which costs, benefits, and probabilities can be assigned. In theory, CBA also might consider all possible outcomes, since implausible outcomes would be assigned extremely low probabilities. For practicability CBA is limited to some set of plausible outcome scenarios. See, e.g., OIRA Primer at 14. Indeed, the difficulty of calculating costs and benefits may mean that CBA accounts only for some even more limited set of reasonably likely outcomes.
information obtained from the surveillance power, such as harms related to arrests, detention, and torture or other mistreatment. “Harms” that may be prevented by surveillance include deaths, physical or emotional injuries, and property damage directly caused by terrorism events, along with the indirect effects of such events, such as economic consequences and fear engendered in those not directly affected.\(^58\)

We may want to exclude some types of harm from the analysis. A few issues about cognizable harm that are relevant to this discussion include:

- **Surveillance harms incurred by those engaged in plotting or carrying out terrorist attacks.** It would seem silly to count typical injuries associated with the acquisition of private information, such as injuries to reputation, feelings of “creepiness,” potential for identify theft, interference with social relationships and the like, as cognizable “harms” when they are incurred by terrorists as a result of legally permissible surveillance. However, harms resulting from morally unacceptable investigative techniques, generally as defined by constitutional or international humanitarian law, should be cognizable even if incurred by the guilty.\(^59\) Of course both kinds of harms are cognizable when they are incurred by those who are innocent of terrorist activity.

- **Non-terrorism-related criminal, tortious or other harms prevented or deterred as a side effect of surveillance.** Whether or not counterterrorism surveillance is designed or intended to prevent non-terrorism harms, it sometimes may have the felicitous side effect of doing so, raising the question of whether such harms should be cognizable in the consistency analysis. Here, I do not take such incidental harms into account for the following reasons:\(^60\) The constraints imposed on counterterrorism surveillance are distinct from the constraints imposed on surveillance for criminal investigations. Counterterrorism surveillance is given more leeway in some respects, presumably because of society’s assessment of the special importance of preventing terrorism events. Each set of constraints should be adjusted in light of the sorts of harmful activities at which it is directed. For purposes of the present discussion, law enforcement surveillance powers are part of the baseline state of the world, to which the harms prevented or caused by constraints on counterterrorism surveillance powers should be compared. If, on the one hand, permissible law enforcement surveillance also could have prevented a particular criminal or other non-terrorism harm, counterterrorism surveillance powers are not a “but-for” cause of its prevention. If, on the other hand, permissible law enforcement surveillance would not have prevented that non-terrorism harm, its failure to do so presumably was taken into the balance in determining the permissible bounds of law enforcement surveillance powers. Thus, even though the “accidental” prevention of non-terrorism harms by counterterrorism surveillance is a social benefit, taking the

\(^{58}\) See, e.g. Gerd Gigerenzer, Dread Risk, September 11, and Fatal Traffic Accidents, 15 Psych. Sci. 286 (2004) (discussing some of the hidden indirect costs of terrorism events);


\(^{60}\) Excluding non-terrorism-related harms from consideration in the PP 2.0 analysis, is akin to Sunstein’s proposed principle of avoiding surveillance designed to investigate non-terrorism crime, but is justified on somewhat different grounds. Beyond Cheneyism and Snowdenism at __.
prevention of non-terrorism harm into account as a reason for looser constraints on counterterrorism surveillance would undermine the integrity of the determination of how to set law enforcement constraints.\textsuperscript{61}

- \textit{Actions taken in aid of terrorist organizations}. Counterterrorism surveillance is intended to prevent terrorism events, but it does so via an indirect chain. The question here is whether to consider each link in the chain as a separately cognizable harm prevention. For example, surveillance may contribute to preventing terrorism events by uncovering or deterring actions (such as financial donations) that support terrorist organizations, thus undermining those organizations capacity for terrorist attacks.\textsuperscript{62} The question arises whether such supporting actions themselves should be treated as cognizable harm in the consistency analysis. Eliminating support for terrorist activity decreases the likelihood of a terrorism event. The likelihood that a surveillance power will prevent a terrorism event is taken into account by the knowledge condition. Counting such support activities themselves as cognizable harms prevented by the surveillance would amount to double counting. One could argue that such support activities produce collateral harms, such as disaffection or unrest among particular social groups or increases in organized crime, that should be independently cognizable in the analysis. As discussed above, however, counterterrorism surveillance is subject to different constraints (and generally given more latitude) because of a social consensus that terrorism events are especially heinous (in comparison to criminal acts). That distinction begins to erode when one moves away from harms caused by terrorism events to more nebulous harms, such as social disaffection and unrest, that may have multiple causes. To the extent we believe that there is a normative reason to apply different rules to counterterrorism surveillance and law enforcement surveillance, incorporating such “side effect harms” into the analysis begins to undermine that distinction.\textsuperscript{63} In what follows, I therefore treat only harms caused (directly or indirectly) by terrorism events as cognizable terrorism harms.

- \textit{Surveillance harms to foreigners}. There is an ongoing debate about whether surveillance harms to non-U.S. persons should be cognizable in determining the appropriateness of counterterrorism surveillance powers.\textsuperscript{64} Most would agree that certain kinds of harms to foreigners that might result indirectly from surveillance, such as loss of life, physical injury, and physical, mental and dignitary harms associated with unacceptable interrogation and detention practices should count in the balance. There is considerably

\textsuperscript{61} Note that this discussion bears only an oblique relationship to the debate about whether and when information uncovered from counterterrorism surveillance should be made available to law enforcement officials investigating non-terrorism crimes. That debate revolves around the potential for counterterrorism surveillance powers to be used improperly for non-terrorism-related investigations. The current analysis concerns the proper scope of legitimate counterterrorism surveillance.

\textsuperscript{62} Many such actions are criminalized as “material support”. 18 U.S.C. §§ 2339A, 2339B.

\textsuperscript{63} Again, I must emphasize that this is not to disagree that society may benefit if such harms are prevented or deterred as a side effect of legitimate counterterrorism surveillance.

less agreement, however, about whether more traditional privacy harms or economic harms imposed on foreigners by U.S. counterterrorism surveillance should be considered. Though this is a very important debate, I leave it to one side here.

Finally, there is a quirk relating to the relationship between the consistency requirement and the harm condition in this context. If there is no terrorism event that would occur “but for” the exercise of a set of surveillance powers, then we cannot say that the exercise of those surveillance powers “prevents” any harm. Thus, the minimum level of harm that any set of surveillance powers can be credited with preventing is the harm caused by a “minor” (only in the relative sense, of course) terrorism event, $H_m$. ($H_m$ is, of course, a significant degree of harm.) Thus, $P(\{A\},K)$ either is greater than $H_m$, meaning that the surveillance powers prevent (with the degree of certainty required by $[K]$) at least some terrorism event, or it is “none,” meaning that the surveillance powers do not prevent any terrorism event with the degree of certainty required by $[K]$. In other words, either $P(\{A\},K) > H_m$ or $P(\{A\},K) = 0$. In light of this relationship, a PP version is consistent in one of two circumstances:

1) $P(\{A\},K) < H_m$ (thus $= 0$) and $C(\{A\},K) > [H]$, for some $[H]$ or
2) $H_m < P(\{A\},K) < [H] < C(\{A\},K)$

C. When Precautionary Principles Go Wrong

With the above preliminaries out of the way, this Section explores two extreme versions of the knowledge condition and demonstrates why they are undesirable before considering a potentially sensible version of PP 2.0 in the next section.

1. Snowdenism, Cheneyism, and Very Permissive Knowledge Conditions

The endemic inconsistency that critics ascribe to the Precautionary Principle results in part from the assumption of extremely permissive knowledge conditions. Consider, for example, “Snowdenism,” which Sunstein describes as holding that “[w]e need to create aggressive safeguards to protect civil liberties not only now but also for periods in which government is in especially bad hands -- and to create precautions against the evident dangers, including worse-case scenarios.” To define a PP version encapsulating “Snowdenism,” we would need to be more specific, identifying particular surveillance powers to be eliminated, a harm condition, and a knowledge condition. (PP 2.0, like cost-benefit analysis, has the virtue of forcing us to think specifically.) “Snowdenism” presumably incorporates a very permissive knowledge condition, such as “remote possibility.” Assuming that knowledge condition, we have:

PP1: If there is a remote possibility that the exercise of surveillance powers $\{A\}$ over and above baseline surveillance powers $\{A'\}$ will cause at least $[H]$ additional harm, prohibit the exercise of $\{A\}$.

CR1: If there is a remote possibility that surveillance powers $\{A\}$ will prevent at least $[H]$ additional harm not prevented by baseline surveillance powers $\{A'\}$, permit the exercise of surveillance powers $\{A\}$. 
Note that the consistency requirement, CR1, looks much like “Cheneyism,” which holds that “Even when the probability of harm is low, and even if government is operating in the midst of grave uncertainty, it is appropriate to do whatever must be done, and whatever technology allows, to prevent deaths and to protect the nation, even or perhaps especially from worst-case scenarios.” The PP 2.0 framework thus immediately highlights the potential inconsistency between these two positions.

Recall that the consistency requirement is satisfied if either:

1) \( P(\{A\},K) < H_m = 0 \) and \( C(\{A\},K) > [H] \), for some \([H]\) or
2) \( H_m < P(\{A\},K) < [H] < C(\{A\},K) \)

For PP1, \([K]\) is “remote possibility.” PP1 satisfies the first prong of the consistency requirement if there is not even a remote possibility that the exercise of surveillance powers \(\{A\}\) will prevent a terrorism event, as long as the exercise of those surveillance powers causes some harm. Surveillance powers \(\{A\}\) satisfying the first version of PP1’s consistency requirement are thus gratuitous, in the sense that they cause surveillance harm without eliminating any terrorism harm. PP1 consistently prohibits them. PP1 thus is analogous to Sunstein’s “avoid gratuitous costs” principle. Like that principle, PP1 may have some nontrivial applications to situations such as NSA storage of telephony metadata or perhaps when surveillance is employed for non-terrorism-related purposes.

Apart from gratuitous surveillance activities, however, PP1 is unlikely to be consistent. For the alternative consistency condition, \(H_m < P(\{A\},K) < [H]\), to hold, \(\{A\}\) must have the potential to prevent a single terrorism event, but not even a remote possibility of preventing terrorism harm > [H], where [H] is set at some harm level > Hm. That scenario is quite unlikely. Most surveillance powers that have a remote possibility of preventing a single attack of harm Hm, however, also have a remote possibility of preventing much greater total harm, such as multiple attacks or more harmful attacks.

Overall, then, a PP version with a very permissive knowledge condition, such as “remote possibility,” consistently prohibits gratuitous exercises of surveillance powers, but is likely to be inconsistent for non-gratuitous exercises of surveillance powers. As anticipated, “Snowdenist” and “Cheneyist” Precautionary Principles are (mostly) inconsistent.

2. Highly Certain Knowledge Conditions and Minor Harm Conditions

At the other end of the spectrum, a PP version employing a very stringent knowledge condition is normatively undesirable because it could consistently prohibit the exercise of surveillance powers that would inflict only minor harm, yet have a relatively high chance of preventing a terrorism event. Consider, for example:

PP2: If there is high certainty that the exercise of surveillance powers \(\{A\}\) in addition to baseline surveillance powers \(\{b\}\) will cause at least [H] additional harm, prohibit the exercise of \(\{A\}\).

Beyond Snowdenism and Cheneyism at __.
CR2: If there is high certainty that surveillance powers \{A\} will prevent at least \[H\] additional harm not prevented by baseline surveillance powers \{b\}, permit the exercise of surveillance powers \{A\}.

Terrorism events, thankfully, are rare, especially in the United States. Thus, one can rarely be “highly certain” that a set of surveillance powers will prevent a terrorism event. In other words, when \[K\] is very stringent, it will often be the case that \(P(\{A\},K) = 0\), the first version of the consistency requirement applies, and the predicate of CR2 is false. Most surveillance powers are highly certain to cause at least some degree of harm, however, meaning \(C(\{A\},K) > [H]\), and the predicate for PP2 is true for some \[H\]. As a result, PP2 often will be consistent, even if the harm condition \[H\] is quite minor. PP2 thus consistently prohibit the exercise of surveillance powers that are likely, but not highly certain, to prevent terrorism, simply because they are highly certain to cause minor harm. Clearly this is not a normatively desirable result.

Even with a more stringent harm condition, PP2 would not be normatively satisfying. Consistency is a necessary condition for applying a PP version, not a sufficient condition. The purpose of the Precautionary Principle approach is to deal with uncertain harm. Setting the knowledge condition at too stringent a level undermines that purpose.

D. Toward a Sensible, Consistent, and Non-Trivial Precautionary Principle for Counterterrorism Surveillance

As illustrated by the previous section, if a precautionary principle for counterterrorism is to be both consistent and normatively desirable, its knowledge condition can be neither too permissive nor too stringent and its harm condition must be sufficiently high. Consider, for example:

PP3: If there is some likelihood that the exercise of surveillance powers \{A\} in addition to baseline surveillance powers \{A'\} will cause substantial additional harm, prohibit the exercise of \{A\}.

CR3: If there is some likelihood that surveillance powers \{A\} will prevent substantial additional harm not prevented by baseline surveillance powers \{A'\}, permit the exercise of surveillance powers \{A\}.

1. Illicit Grounds for Surveillance

We can use PP3 to derive something akin to Sunstein’s principle of avoiding surveillance conducted on illicit grounds. Certain exercises of surveillance powers, such as singling out a group that is disfavored or disadvantaged or stalking an individual for nefarious purposes, can always be deemed to have some likelihood of leading to substantial harm. The same is true of exercises of intrusive surveillance powers, such as bugging a home, taking a blood sample, or wiretapping a phone, that affect sufficient numbers of innocent people. The predicate for PP3 thus is true for these activities. Whether these surveillance powers are also somewhat likely to prevent substantial harm depends on how they are targeted. If the targeting bears no relation to reasonable suspicions of terrorism activity, these exercises of surveillance power will not be even
somewhat likely to prevent substantial terrorism harm. The predicate to CR3 is false in such cases. PP3 thus consistently prohibits exercises of surveillance powers that are targeted “illicitly,” in the sense that the surveillance is likely to be harmful, but unlikely to prevent terrorism events. Note that the surveillance prohibited by PP3 need not be entirely gratuitous. The fact that it is likely to cause substantial harm, but not likely to prevent terrorism events is sufficient.

2. An Intrusiveness and Targeting Model

To gain further insight into whether PP 2.0 will give us any traction in resolving more difficult cases, it is useful to consider a somewhat more specific model in which surveillance powers, \( A(I,T) \), are characterized by their intrusiveness \([I]\) and targeting constraint \([T]\). Intrusiveness \([I]\), connotes the impact a particular form of surveillance has on an individual subjected to it. Targeting constraint \([T]\), connotes the criteria used to determine which individuals can be subjected to a particular form of surveillance. In reality, surveillance powers vary qualitatively and contextually in ways that can affect both the harms they may cause and their usefulness in preventing terrorism events. Thus, neither intrusiveness nor targeting constraints can be aligned precisely along a single scale. Though one can say roughly that some forms of surveillance are more intrusive than others, (e.g. bugging a person’s apartment is more intrusive than listening to her conversations in a public park), there is no easy way to determine, without knowing the context,\(^66\) whether forcing an individual to have a blood test is “more” or “less” intrusive than wiretapping her phone. Similarly, it is difficult to say whether surveilling someone because she belongs to a radical political party is “more” or “less” targeted than surveilling her on the grounds that the surveillance may produce information that is “relevant” to an investigation. Nonetheless, we can learn something by considering a model in which a surveillance power \( A(I,T) \) is characterized simply by its degree of intrusiveness and the strictness of its targeting constraint.

In particular, consider targeting constraints whose strictness is defined by thresholds requiring a particular degree of evidence to suspect that surveilling a particular individual will uncover information useful for preventing a terrorism event. The higher the evidentiary threshold, the tighter the targeting constraint. Targeting constraints based on evidentiary thresholds, such as "probable cause" and "reasonable suspicion," are ubiquitous in the laws and regulations governing law enforcement and counterterrorism surveillance. They generally involve a determination by a specified government official (a judge or law enforcement officer) that the evidence supports a particular level of certainty that the information to be obtained will bear a particular kind of relationship to a particular kind of illicit activity.

In practice, targeting constraints usually are stricter for more intrusive forms of surveillance. Electronic surveillance of communications content is considered significantly intrusive. Under the Foreign Intelligence Surveillance Act, an electronic surveillance order for the interception of communications content may be issued only if an appropriate law enforcement officer submits facts providing probable cause to believe that the target is (among other possibilities) involved in activities in preparation for terrorism, and that specific procedures will be used to minimize the acquisition and retention of information concerning US persons, consistent with the need for the

information for counterterrorism purpose. By requiring a court order, employing a probable cause threshold, and mandating minimization procedures, this provision is intended to ensure that this type of surveillance is quite tightly targeted. A similar standard is employed for issuing warrants for physical searches under FISA. “Non-content” surveillance traditionally has been considered less intrusive. So-called “pen register” orders for the interception of “dialing, routing, addressing or signaling information” that does not include communications content are available under a much more lenient targeting constraint requiring only a “certification” by an appropriate law enforcement officer that the information likely to be obtained is “relevant” to an ongoing counterterrorism investigation and, if targeted at a U.S., “not conducted solely upon the basis of” First Amendment-protected activities. Some types of surveillance, such as listening to a conversation taking place in a public park, are considered only minimally intrusive and subject to no targeting constraint whatsoever.

In what follows, I will assume a similar correlation between targeting constraint and intrusiveness, such a surveillance power can be denoted as A(T), and intrusiveness is assumed to fall of as T is loosened.

3. How Targeting Affects Surveillance Harm

There are several mechanisms by which targeting affects surveillance harm. Most obviously, tighter targeting reduces the total number of individuals subject to surveillance. It also preferentially selects individuals for whom there is more evidence that surveillance will be useful in preventing terrorism events, thus reducing the chance that innocent individuals will be surveilled. Finally, tighter targeting reduces chilling effects and a whole host of other indirect social harms, such as breakdown of trust, decline in democratic participation, and the like, that arise when surveillance arbitrarily to innocent individuals. Because these indirect social harms arise from interactions between individuals, they are subject to nonlinear feedback and can grow more than proportionally as the number of innocent surveillance targets grows. So-called “tipping points” and “cascades” of surveillance harm thus become increasingly likely as targeting constraints are loosened. Taken together, these mechanisms suggest that total surveillance harm will increase in severity, both in total and at the margin, as targeting constraints become looser. The tendency for marginal surveillance harm to increase will be offset to some degree by the decreasing intrusiveness of more loosely targeted surveillance, though it is possible that the number of permissible targets may outweigh the decreasing intrusiveness, so that marginal surveillance harm will continue to increase. Eventually, however, when the targeting constraint becomes so loose that nearly the entire population is permissibly surveillable, the marginal surveillance harm will begin to drop.

In addition to the harms already discussed, surveillance causes distributive harms if it focuses on

---

67 50 U.S.C. §1801 et seq. There are many more details in FISA relating to the definition of terrorism, the prerequisites for obtaining the orders, and so forth. Here, I mention only the aspects needed to illustrate this discussion.

68 50 U.S.C. §1821 et seq.

69 Whether, and to what extent, that continues to be the case when large amounts of non-content data are aggregated is an issue of current debate. See, e.g. U.S. v. Jones, 565 U.S. __ (2012) (Sotomayor, J., concurring).

70 50 U.S.C. §1841 et seq.
disfavored or disadvantaged ethnic, political, or religious groups. Members of disfavored or disadvantaged groups may be singled out for surveillance in at least three different ways: illicit targeting, disproportionate impact, and discretion bias. The previous section discussed illicit targeting, concluding that it is consistently prohibited by PP3. Here I consider distributive harms that can arise from disproportionate impact and discretion bias even when targeting constraints use evidence-based thresholds.

Members of a disfavored group will be disproportionately impacted by evidence-based targeting whenever there is a known or suspected correlation between terrorism events and membership in a particular group. Because terrorism is, by definition, associated with political ends, that is the usual scenario. Perpetrators often come disproportionately from a minority religious, ethnic, or political group. The vast majority of members of such a group are entirely uninvolved in terrorist activity, yet will be disproportionately subject to surveillance. While the very tightest targeting criteria might require significant evidence of an individual’s own involvement in suspicious activities, looser targeting criteria give significant weight to social connections. If someone suspected of terrorist activity belongs to a particular social group, innocent members of that group are more likely than innocent members of the population at large to have some social connection to the activity. They are more likely to know the same people, eat at the same restaurants, shop at the same stores, live in the same neighborhoods, attend the same meetings, and so on. As a result, whatever the evidentiary threshold for a given targeting constraint, innocent group members are more likely than innocent members of the public at large to meet the threshold and be subject to surveillance. Distributive harm results directly from the disproportionate likelihood that innocent group members will incur surveillance harm and may be heightened by the group’s awareness that it is being singled out. Disproportionate surveillance also heightens indirect social harms and makes them more likely to be severe at a tighter level of targeting constraint. These distributive surveillance harms are real, and may be substantial, even if they arise entirely from unbiased application of neutral targeting conditions.

Distributive harm may be exacerbated if, for one of several possible reasons (including conscious and unconscious bias against the group, as well as various cognitive biases), counterterrorism officials exercise surveillance targeting discretion in a biased manner, such that group members are even more disproportionately subjected to surveillance. The looser the targeting criterion, the more the potential for biased exercise of discretion.

The overall impact of distributive harm will be to heighten the overall severity of the marginal surveillance harm. Distributive harm and indirect social effects may or may not fall off as quickly as individual surveillance harm as intrusiveness decreases.

Of course, this discussion is extremely general and speculative, but the pattern of marginal surveillance harm increasing as targeting constraints are loosened, and falling off after constraints become very loose seems likely to be fairly general. Of course, while marginal surveillance harm rises and falls, total surveillance harm always increases as the targeting constraint is loosened.
4. How Targeting Affects Terrorism Prevention

In general, one would assume that the looser the targeting constraint, the more likely it is that counterterrorism surveillance will prevent a terrorism event. Beginning from the tightest constraint, every incremental expansion of surveillance powers will include the power to conduct all the surveillance permitted by the tighter constraint, plus more. When the targeting constraint is relatively tight, loosening the targeting constraint increases the surveillance’s potential effectiveness in preventing terrorism events. Eventually, however, the advantage gained by loosening the targeting constraint will begin to decrease. As the targeting constraint is relaxed, there is less and less evidence that the newly permissible surveillance targets have any connection to terrorist activity and it becomes less and less likely that surveilling them will uncover new and useful information. Moreover, because terrorist activity is rare, it becomes less and less likely as the targeting constraint is loosened that there are any new surveillance targets out there that would provide useful, non-duplicative information. Preparing for a terrorist attack requires activity that usually creates at least some evidence. For example, plotters are often affiliated with terrorist or violent extremist organizations. They undertake other suspicious activities, such as purchasing weapons or explosives. They frequent websites in which violent extremist actions are discussed or promoted. As the targeting constraint becomes looser, it becomes less and less likely that anyone could plot a terrorism event without creating enough evidence to authorize surveillance under the existing constraint. This is especially the case because surveillance itself uncovers additional evidence that can be used to meet the evidence threshold, bringing additional individuals within the existing targeting constraint.

Even if a looser targeting constraint brings new and potentially useful targets within the purview of the government’s surveillance powers, as the targeting constraint becomes looser and looser it becomes less and less likely that data about those particular individuals will be obtained, identified, and analyzed so that it can actually help to prevent a terrorism event. The looser the targeting constraint, the more the search for useful information resembles a hunt for a needle in a large, and growing, haystack. Moreover, as the evidentiary threshold lowers and the permissible intrusiveness decreases, useful targets about whom there has not been sufficient evidence to meet the threshold are likely to be more and more difficult to distinguish from the rest of the hay. As the targeting criterion loosens, the task becomes more and more like looking for a particular blade of hay in a haystack.

Of course, as individuals leave more and more detailed “data trails” of their activities, it is becoming cheaper and cheaper to collect massive amounts of data about large number of individuals. Resource constraints on collection of intrusive information are much less significant than they were in the past; it is now possible to conduct mass surveillance “on the cheap.” The problem of searching through the haystack remains, however. Unless there is some “quick and dirty” way to comb through the data to find relevant information (the big data equivalent of the proverbial “dog sniff”), the signal-to-noise ratio of untargeted data analysis is likely to be such that its marginal impact on terrorism prevention is unlikely to be significant.71 The bulk telephony metadata program exemplifies these points. According to reports by the President’s Review Group72 and the PCLOB,73 the program never was used to obtain information that was

---

71 Indeed, this seems to have been one of the lessons of the recent NSA disclosures.
72 RICHARD A. CLARKE ET AL., LIBERTY AND SECURITY IN A CHANGING WORLD: REPORT AND
not duplicative or could not have been obtained under more tightly targeted surveillance authorities. Whatever information useful for preventing terrorism may have been buried within the haystack of everyone’s data, the NSA had no method for uncovering it.

The bottom line is that, while loosening targeting constraints initially may increase the likelihood of preventing terrorism harm, eventually the marginal terrorism prevention value of looser surveillance will decrease -- and decrease quite quickly once the number of permissible targets becomes significant.

5. A Precautionary Principle Based on the Targeting Constraint Model

Bearing in mind the general discussion about how surveillance harm and terrorism prevention are affected by targeting constraint stringency, it is now possible to attempt to construct a sensible and consistent PP version based on PP3. Consider a baseline set of surveillance powers \{A(T)\} and a proposed regulation that would permit surveillance activities for which \( T < T_R \) (where \(<\) is taken to mean “less loose”) and prohibit surveillance \( T > T_R \). Denote the proposed set of reduced permissible surveillance powers as \{A(T): T < T_R\} and the surveillance powers to be eliminated as \{A(T): T > T_R\}. PP3 can then be written as:

**PP3:** If there is some likelihood that the exercise of surveillance powers \{A(T): T > T_R\} in addition to baseline surveillance powers \{A(T): T < T_R\} will cause substantial additional harm, prohibit the exercise of \{A(T): T > T_R\}.

**CR3:** If there is some likelihood that surveillance powers \{A(T): T > T_R\} will prevent substantial additional harm not prevented by baseline surveillance powers \{A(T): T < T_R\}, permit the exercise of surveillance powers \{A(T): T > T_R\}.

Consider first how the predicate for PP3 depends on the choice of \( T_R \). If \( T_R \) is quite loose, surveillance powers \{A(T): T > T_R\} that we are considering eliminating will be responsible for only a relatively small contribution to the total surveillance harm. The harm they add is unlikely to be “substantial.” As \( T_R \) is tightened, the harm associated with \{A(T): T > T_R\} will increase until at some point, \( T_{PP} \), the harm contributed by the exercise of the additional surveillance powers \{A(T): T > T_R\} will be “substantial.” Since any further tightening simply increases the surveillance harm contributed by \{A(T): T > T_R\}, the predicate for PP3 will be true whenever \( T_R < T_{PP} \). The line below illustrates the situation.

```
<table>
<thead>
<tr>
<th>PP3 true</th>
<th>PP3 false</th>
</tr>
</thead>
<tbody>
<tr>
<td>[\text{T}_{PP}]</td>
<td></td>
</tr>
</tbody>
</table>
```

---

**RECOMMENDATIONS OF THE PRESIDENT’S REVIEW GROUP ON INTELLIGENCE AND COMMUNICATIONS TECHNOLOGIES (Dec. 12, 2013) (President’s Review Group Report).**

Now consider how the predicate for CR3 depends on the choice of $T_R$. To be somewhat likely to prevent substantial additional harm, exercising the potentially prohibited surveillance powers $\{A(T): T > T_R\}$ must at a minimum be somewhat likely to prevent a “small” terrorism event, $H_m$. When $T_R$ is very loose, surveillance powers $\{A(T): T > T_R\}$ that are eliminated will be very unlikely to be effective in preventing a terrorism event. Thus, the predicate for CR3 will be false. As $T_R$ is tightened, the elimination of $\{A(T): T > T_R\}$ becomes more likely fail to prevent a terrorism event. Eventually, there will be a point, $T_{CR}$, at which the additional surveillance powers provided by $\{A(T): T > T_R\}$ are “somewhat likely” to prevent substantial harm. For $T_R > T_{CR}$, the predicate for CR3 will be false.

Combining these two analyses, we can expect something like this:

\[
\begin{array}{ccc}
\text{PP3 true} & \text{PP3 true} & \text{PP3 false} \\
\hline
T_{CR} & T_{PP} \\
\hline
\text{CR3 true} & \text{CR3 false} & \text{CR3 false}
\end{array}
\]

$T_{CR}$, $T_{PP}$, and their relative positions depend on two factors: 1) the relative magnitudes of the total harms “somewhat likely” to be caused and prevented, respectively, by the exercise of surveillance powers $\{A(T)\}$ and 2) the points at which the total harms caused and prevented by the exercise of the surveillance become “substantial” as the targeting constraint is tightened (which in turn depend on the behavior of the marginal surveillance harm). For illustration, and in accordance with my intuitions, I have represented $T_{CR} < T_{PP}$. In principle, though, either $T_{CR}$ or $T_{PP}$ could occur at a looser constraint.

For PP3 to be consistent we need only find some $T_R$ for which the predicate to PP3 is true and the predicate to CR3 is false. Here there is a range of targeting constraints between $T_{CR}(I_R)$ and $T_{PP}(I_R)$, for which PP3 would be consistent. To minimize unnecessary surveillance harm, in conformity with the MPP, we should choose $T_R$ to be the more restrictive of $T_{CR}$ and $T_{PP}$.

The analysis in this section is necessarily sketchy, but it is enough to illustrate that it is possible, at least in principle, to use PP 2.0 to construct a sensible, consistent, and non-trivial Precautionary Principle for counterterrorism surveillance. It suggests that focusing on specifics, and on targeting constraints in particular, is a path toward escaping the inconsistency corner that we paint ourselves into with more broad brush approaches.
V. Conclusions

There is a tendency towards contentiousness in the debate about cost-benefit analysis and the Precautionary Principle. This tendency stems partly from normative disagreements, including differences of opinion about which threats to sound government decision-making loom largest: cost-benefit analysis focuses on mistakes caused by cognitive biases and the neglect of economic costs, while the Precautionary Principle focuses on mistakes caused by the neglect of uncertain, substantial harms. The contentious tenor of the debate seems also to have stemmed, however, from some lack of precision in defining the meaning and scope of the Precautionary Principle. Daniel Steel’s recent work adds significant clarity to the discussion. Using his Precautionary Principle framework, one can make meaningful comparisons between the two approaches and draw some general inferences about their relationships. Here, for example, I have explored the way in which factors such as harm condition severity, knowledge condition permissiveness, uncertainty about probability estimates, and valuation metrics affect the likelihood that the Precautionary Principle will recommend regulation when cost-benefit analysis recommends against it. I also have argued that there are important underlying similarities between the Precautionary Principle, as expounded by Steel, and the modern cost-benefit analysis technique of breakeven analysis.

In the counterterrorism surveillance context, the Article has responded to Professor Sunstein’s criticism of the precautionary principle approach by demonstrating that it is possible, at least in principle to construct sensible, consistent, and non-trivial version of the Precautionary Principle. It also shows that both the “avoid gratuitous costs” and “avoid illicitly grounds for surveillance” principles suggested by Sunstein are derivable from Steel’s Precautionary Principle Framework. I also argue (building on Steel’s more general argument) that the Precautionary Principle framework provides a means for handling “worst cases” that avoids particularly severe pathologies of the maximin principle in the counterterrorism surveillance context. Finally, the Article uses a simplified model to explore the way in which surveillance targeting constraints can be used to escape the apparent bind posed by the conflict between “Cheneyist” and “Snowdenist” approaches to surveillance regulation.

---