

---

---

ON THE DESIGN OF LEGAL RULES:  
BALANCING VERSUS STRUCTURED  
DECISION PROCEDURES

*Louis Kaplow*

CONTENTS

I.	INTRODUCTION.....	993
II.	LEGAL RULE DESIGN .....	997
	<i>A. Set-up.....</i>	997
	<i>B. Final Decision Rule.....</i>	1000
	<i>C. Less Restrictive Alternatives.....</i>	1003
III.	INFORMATION GATHERING .....	1015
	<i>A. Optimal Information Collection .....</i>	1015
	<i>B. Structured Information Protocols.....</i>	1024
IV.	ADDITIONAL CONSIDERATIONS .....	1033
	<i>A. Conduct of Legal Proceedings.....</i>	1034
	<i>B. Screening.....</i>	1043
	<i>C. Reflections on Balancing.....</i>	1047
	1. <i>Queasiness About Balancing.....</i>	1047
	2. <i>Rules Designed to Constrain Balancing .....</i>	1055
V.	CONCLUSION.....	1061

---

---

## ON THE DESIGN OF LEGAL RULES: BALANCING VERSUS STRUCTURED DECISION PROCEDURES

*Louis Kaplow\**

*Important doctrines in diverse areas of law employ structured decision procedures requiring, in rough terms, that the plaintiff first make some demonstration of harm; if but only if that is done, the defendant must make some showing of benefit; and if but only if that occurs, balancing is performed. This Article compares such protocols to unconstrained balancing and finds them to be inferior with respect to the quality of final decisions: they sometimes fail to impose liability even though the harm is greater than the benefit, and they sometimes impose liability even though the benefit exceeds the harm. The Article also develops the principles of optimal information (evidence) collection and shows how structured decision procedures violate every core lesson and presuppose distinctions that often are incoherent or impractical to implement. The analysis addresses concerns about balancing that may motivate structured protocols, how less restrictive alternatives should be assessed, and the extent to which legal proceedings are conducted in conformity with either approach, as well as how they might be reformed.*

### I. INTRODUCTION

Balancing is a familiar mode of decisionmaking in the law and beyond. When one consideration favors a particular decision (say, liability) and another opposes it, it seems to be the essence of reason that the superior decision reflects the balance of the competing forces, taking into account the weight of the evidence and the importance of each factor. Many legal rules, such as the negligence test for tort liability, operate in this fashion.

Sometimes, however, structured decision procedures are used instead for these types of decisions. As a benchmark for comparison with balancing, this Article examines the following stylized version:

- (1) The plaintiff must show that the harm of the defendant's act exceeds some threshold. If not, there is no liability. If so:
- (2) The defendant must show that the benefit of its act exceeds some other threshold. If not, there is liability. If so:
- (3) The harm and benefit are balanced, and there is liability if and only if the harm is greater.

---

\* Finn M.W. Caspersen and Household International Professor of Law and Economics, Harvard Law School, and Research Associate, National Bureau of Economic Research. I thank Erwin Chemerinsky, Steve Churchill, John Donohue, Richard Fallon, James Greiner, Vicki Jackson, Benjamin Sachs, Holger Spamann, Matthew Stephenson, Adrian Vermeule, and workshop participants at Berkeley, Chicago, Harvard, Northwestern, and Stanford for helpful discussions and comments; Tracy Cui, Carys Johnson, Kevin Keller, Carsten Koenig, Iacopo Lash, Andrea Lowe, Asher Lowenstein, Clarissa Lu, Junesoo Seong, and Aluma Zernik for research assistance; and Harvard's John M. Olin Center for Law, Economics, and Business for financial support. Disclaimer: I occasionally consult on antitrust cases, and my spouse is in the legal department of a financial services firm.

---

---

Similar or related schemes are thought to characterize some existing legal decision procedures or have been proposed — notably, in antitrust law (rule of reason, mergers), discrimination law (Title VII disparate impact), and constitutional law (strict scrutiny, proportionality analysis).<sup>1</sup> Yet structured decision procedures of this sort are neither used nor advocated in most other areas of law.

It is natural to inquire when and why legal rules should employ structured decision procedures in lieu of balancing.<sup>2</sup> And, when they are used, it is necessary to set the two decision thresholds, for their height is critical to the procedure's bite and, in particular, its difference from unconstrained balancing. (If both thresholds are set to zero, the structured protocol dissolves into balancing.) This Article aims to answer these questions and others concerning the design of legal rules and procedures. It analyzes large and heretofore unrecognized differences between structured decision procedures and pure balancing with respect to case outcomes and prescriptions for information (evidence) collection. In both respects, structured protocols are demonstrated to be inferior.

Part II compares the two approaches as final decision rules when all of the information that will be considered is before the decisionmaker. In many cases, including all those that reach step 3 under the structured protocol, the decisions will be the same. In important settings, however, the outcomes will differ. Moreover, in all such cases the outcome under the structured rule is necessarily inferior in the sense that such cases involve either the assignment of liability when the benefit exceeds the harm or a failure to assign liability when the harm exceeds the benefit. In addition, the purported virtue of structured rules in avoiding difficult balancing turns out to be misleading. They avoid balancing in many easy cases (for example, the harm is huge and the benefit negligible, so we stop after step 2), but neither effort nor error is reduced. And they avoid balancing in some hard cases, but that is precisely when they stop short of the balancing performed in step 3 even though the resulting outcome from step 1 or step 2 may well be incorrect. Moreover, they sometimes require close comparisons with the thresholds even though balancing would have been easy. On reflection, it is remarkable that structured decision procedures are believed to prevail in important areas of law and are advocated as replacements for balancing, yet analysts have not even asked the basic question of how outcomes under the two methods differ.

---

<sup>1</sup> See, e.g., 1 ABA SECTION OF ANTITRUST LAW, ANTITRUST LAW DEVELOPMENTS 61–62 (8th ed. 2017) (describing lower courts' evolving articulation of antitrust's rule of reason); 42 U.S.C. § 2000e-2(k)(1)(A) (2012) (stating Title VII's disparate impact test); ERWIN CHEMERINSKY, CONSTITUTIONAL LAW: PRINCIPLES AND POLICIES 567 (5th ed. 2015) (presenting strict scrutiny). For extensive discussion of the literature, see Louis Kaplow, *Balancing Versus Structured Decision Procedures: Antitrust, Title VII Disparate Impact, and Constitutional Law Strict Scrutiny*, 167 U. PA. L. REV. (forthcoming 2019).

<sup>2</sup> Another important type of legal rule uses categorization. See *infra* subsection IV.C.2.

Part II also examines how the analysis of less restrictive alternatives (or alternative employment practices, narrow tailoring, and the like) fits in. It is interesting that this form of inquiry — which asks if we can obtain some or all of the defendant's benefit without causing as much harm — although generic, seems to take an independent doctrinal form mainly where structured decision procedures are operative. The explanation is that, on one hand, the consideration of alternatives is a natural, inseparable part of balancing. (What a defendant might have done to avoid an accident is at the essence of the negligence inquiry, not some appendage.) On the other hand, because structured decision procedures can readily lead us astray, adding this additional inquiry is akin to incorporating an epicycle in the Ptolemaic system: the machinery as a whole appears odd and thereby suggests that something is fundamentally amiss; but, taking the misguided core as given, the addition does improve outcomes (predictions). This perspective — which sees inquiries into less restrictive alternatives as an appendage to a flawed protocol — illuminates disputes about how such inquiries should be conducted and where in the multistep regimen they might best be attached.

Part III turns to a comparison of the two approaches as guides to information gathering. Although structured decision procedures are sometimes favored because of their supposed advantage in this domain (stopping early, at step 1, economizes on effort), this source of gain proves to be largely illusory. Optimal, unconstrained information gathering, which is what ideally would be done under a balancing approach, involves a number of principles that are sharply violated by the structured protocol, stemming from the latter's sequential separation of the investigation of the harm and the benefit of a challenged practice. First, much evidence is expressly comparative; indeed, characterization evidence is relevant precisely to the extent that it bears differentially on competing understandings of the defendant's alleged act. Attempting to separate the two is artificial and fraught, somewhat like using scissors, disjoined, one blade at a time. Second, evidence often naturally clumps by source rather than by subject: internal documents, witnesses, and experts; not harm and benefit. Third, even if all evidence bore only on harm or only on benefit and naturally clustered in single-issue bundles, it is a priori unlikely that the optimal order of gathering and assessing evidence would be to do first all of one type (harm), followed by all of the other (benefit). Instead, it is (roughly) sensible to collect first, second, and so forth whatever bundle has the highest diagnosticity to cost ratio; at any given point, the most promising bundle may just as plausibly involve benefit as harm. Indeed, since there tend to be diminishing returns with respect to the exploration of each issue, it is unlikely that all of the most promising avenues would concern only one and all of the least promising only the other.

Part IV addresses a number of additional considerations. First, the two approaches to legal rule design are assessed with respect to the conduct of legal proceedings. Enforcement agencies have substantial control over their internal operations and hence can, and probably do, operate to a significant degree in both information gathering and decisionmaking in the manner suggested by a balancing approach, even when the legal rule is a structured decision procedure. U.S. civil litigation, however, has an organization of its own, and one that for the most part does not comport with the dictates of either of the two approaches, except for decisionmaking at the conclusion of a trial. Indeed, this fact casts a curious light on many discussions of structured decision procedures with regard to information gathering and burden shifting. For example, unless litigation is fully bifurcated from the start (before discovery), how can it be suggested (say, in some areas of antitrust law) that, first, all of the information on harm (anticompetitive effects of a practice) is collected, and only then, and only if harm is proved to some significant degree, do we begin to gather information on benefit (pro-competitive effects)? In any event, this Article's analysis of optimal information gathering should prove useful in guiding case management under existing procedures and in constructing procedural reforms. Part IV also revisits step 1 of these structured decision procedures with an eye to the role it may play in screening out weak cases.

Finally, Part IV reflects on balancing itself, an idea that generates significant resistance in many legal settings. Some are queasy about balancing because of difficulties concerning quantification or commensurability. Although both can be serious practical challenges, they are not conceptual limitations. Neither justifies the avoidance of balancing through the use of structured decision procedures. Another suggestion is that, in certain realms (constitutional law being the most obvious), it may be optimal to employ rules designed to constrain balancing. This point is valid but does not fit well with the sort of constraint imposed by the structured decision procedures under consideration here.

This Article illuminates a number of broad questions regarding the legal system, including the relationship between the structure of legal rules and outcomes, information collection as a central aspect of adjudication, various dimensions of balancing (quantification, commensurability, and constraints), less restrictive alternatives, and overlooked mismatches between legal discourse and the actual operation of civil litigation. As sketched in the Conclusion, a sequel to this Article makes these general lessons more concrete by applying them to important doctrines in antitrust, Title VII, and constitutional law — in the process suggesting significant respects in which both positive and normative analysis in those fields misses the mark.<sup>3</sup>

---

<sup>3</sup> See Kaplow, *supra* note 1.

## II. LEGAL RULE DESIGN

### A. *Set-up*

This section presents the stylized setting that will be used as a benchmark for analysis in this Article. No claim is made that the bare-bones statement of structured decision procedures offered here depicts any particular existing legal doctrine; such structured decision procedures are in various respects murky and contested, differ from each other, and (depending on the interpretation) diverge in various ways from this baseline.<sup>4</sup> The proffered stark formulation is chosen for clarity, in order to sharpen discourse about legal rule design as well as our understanding of particular applications.

A legal decisionmaker — an agency, court, or other tribunal — is confronted by a case.<sup>5</sup> Its ultimate choice is whether or not to assign liability, which for ease of exposition will be taken to involve injunctive relief (the application to *ex ante* behavior is elaborated in the footnote).<sup>6</sup> The imposition of liability, relative to a finding of no liability, results in the avoidance of a harm of *H* and the loss of a benefit of *B*. Either value might be zero, each may be highly uncertain, and part of the task (elaborated in Part III) involves the gathering of information to sharpen these estimates.<sup>7</sup>

Both *H* and *B* are taken here to be denominated in relevant units from a social perspective, making this fairly simple representation rather general in this respect. Further elaboration regarding matters of quantification and commensurability is deferred to subsection IV.C.1. The core feature of this set-up is that there are (at least) two considerations

---

<sup>4</sup> *See id.*

<sup>5</sup> Most of the discussion abstracts from the fact that the flow of cases is itself endogenous to the decision procedure. One important dimension concerns screening, including discouraging the filing of frivolous cases and avoiding the suppression of valid, valuable cases. *See infra* section IV.B.

<sup>6</sup> In many legal settings, a significant, even primary, function of liability is to deter harmful conduct while avoiding the chilling of beneficial behavior. The *H* and *B* introduced below can be interpreted as stand-ins for deterrence benefits and chilling costs (or other, broader effects of the application of the law), although there are important (but subtle) differences between this function of liability and settings in which the decision concerns the prohibition or permission of an act going forward (which is nominally true for merger review, zoning decisions, drug approval, and injunctions more broadly). For a formal analysis of the differences, see Louis Kaplow, *On the Optimal Burden of Proof*, 119 J. POL. ECON. 1104 (2011) [hereinafter Kaplow, *Optimal Burden of Proof*]. Informal analysis and substantial elaboration appear in Louis Kaplow, *Burden of Proof*, 121 YALE L.J. 738 (2012); and Louis Kaplow, *Likelihood Ratio Tests and Legal Decision Rules*, 16 AM. L. & ECON. REV. 1 (2014) [hereinafter Kaplow, *Likelihood Ratio Tests*]. The analysis is extended to multistage decisionmaking in Louis Kaplow, *Multistage Adjudication*, 126 HARV. L. REV. 1179 (2013); and Louis Kaplow, *Optimal Multistage Adjudication*, 33 J.L. ECON. & ORG. 613 (2017).

<sup>7</sup> The formulation presented here allows both the interpretation that a single act may generate some *H* and some *B* and that there may be two different types of acts, each generating only *H* or only *B*, but due to uncertainty regarding characterization, it is possible that the expected values of both magnitudes are positive for the single act before the decisionmaker.

that are relevant to the decision that can point in different directions, and it is further assumed that there are situations in which each of these considerations might be more important than the other.<sup>8</sup>

Under balancing,<sup>9</sup> liability is assigned if and only if  $H > B$ . (The appropriate outcome when there is a tie will be ignored throughout, with conventional tiebreaking norms employed without further comment.<sup>10</sup>) Because there is often uncertainty, which may remain significant even after information gathering is complete,  $H$  and  $B$  are best interpreted as expected values.<sup>11</sup>

Under a structured decision procedure, it will be assumed that liability is determined by the following three-step protocol:

- (1) If  $H > H^*$ , proceed to step 2. Otherwise, assign no liability and stop.
- (2) If  $B > B^*$ , proceed to step 3. Otherwise, assign liability and stop.
- (3) If  $H > B$ , assign liability. Otherwise, assign no liability. And stop.

Step 1 asks if the estimated harm,  $H$ , exceeds a stated threshold,  $H^*$ .<sup>12</sup> If not, there is no liability. If it does, we proceed to step 2, which

<sup>8</sup> The present discussion abstracts from administrative costs in assessing legal decision rules. The analysis of information gathering in Part III and the treatment of screening in section IV.B devote substantial attention to the costs of processing cases.

<sup>9</sup> To avoid excessive verbiage, the term “balancing,” standing alone, refers to pure or unconstrained balancing, in contrast to a structured decision procedure, unless the context clearly indicates otherwise (notably, when reference is being made to the balancing that occurs in step 3 of such a structured decision procedure).

<sup>10</sup> On one hand, if the  $H$  and  $B$  associated with acts are each drawn from continuous distributions, the probability of an exact tie is zero, so tiebreaking is irrelevant. See Kaplow, *Burden of Proof*, *supra* note 6, at 758 n.34. On the other hand, disputes about tiebreaking (such as in debating who should bear the burden of persuasion under the preponderance rule, when interpreted as “more likely than not”) are taken to be important, presumably because it is assumed that actual decisionmakers may be inclined to treat some mass of cases that are near equipoise as if they were exact ties or that the tilt in the instruction may nontrivially shift their prior probabilities or process of inference. See Baruch Fischhoff & Wändi Bruine de Bruin, *Fifty-Fifty = 50%?*, 12 J. BEHAV. DECISION MAKING 149 (1999); Kaplow, *Burden of Proof*, *supra* note 6, at 758 n.34; Charles M. Yablon, *A Theory of Presumptions*, 2 LAW PROBABILITY & RISK 227 (2003). This issue will resurface in section IV.A’s discussion of the conduct of legal proceedings.

<sup>11</sup> This interpretation assumes that the decisionmaker is risk neutral. If not,  $H$  and  $B$  can be reinterpreted as risk-adjusted values; more precisely, we can interpret the quantity  $H - B$  as a risk-adjusted value and assign liability when it is positive.

<sup>12</sup> In light of the aforementioned importance of uncertainty, there are alternative ways that we can interpret the tests in steps 1 and 2 with regard to what it means for a value to be greater than the stated threshold. One (which will largely be followed here) is that the expected value of  $H$  must exceed a numerical threshold  $H^*$ . A variant, which we might imagine with separate application of a preponderance rule to each step, would be to require that there be a greater than 50% likelihood that the actual value of  $H$  exceeds  $H^*$ . Depending on the distribution of possible values of  $H$ , either could be more demanding. (For example, it may be most likely that  $H$  is very low, even zero, but if there is a 10% chance that  $H$  exceeds  $H^*$  and, when it does, it does so by more than tenfold, then the expectation version of the test is met but not the preponderance version. On the other hand,

then asks the analogous question with regard to the estimated benefit,  $B$ . Here, if it does not exceed the stated threshold,  $B^*$ , we assign liability (harm is taken to be established, and benefit is not). If it does, we proceed to step 3, where we weigh harm and benefit (both having been established) and assign liability if and only if the former is greater.

The decision thresholds,  $H^*$  and  $B^*$ , play a critical role in defining a structured decision procedure and, importantly for present purposes, in distinguishing it from balancing. If  $H^* = 0$  and  $B^* = 0$ , this procedure dissolves into unconstrained balancing.<sup>13</sup> Roughly speaking, the greater are  $H^*$  and  $B^*$ , the more the structured procedure differs from balancing as a decision rule.<sup>14</sup> (As we will see in Part III, however, even with low thresholds, it is possible for large differences to exist with respect to how information gathering is conducted.)

The height of each decision threshold is a major focus in the sequel article's analysis of applications to antitrust, Title VII, and constitutional law.<sup>15</sup> The point here is simply that these thresholds in an important sense define the substance of a structured decision rule.<sup>16</sup> In principle, we could contemplate setting  $H^*$  and  $B^*$  optimally, in light of

---

there could be a 60% chance that  $H$  barely exceeds  $H^*$  and a 40% chance that  $H$  equals 0, in which case the preponderance version is met but the expectation version is not.) Note further that essentially the same point applies to step 3's balancing.

Although not the focus of this investigation, an issue worth noting is the generic inferiority of the preponderance rule on the ground that, unlike when taking expected values, magnitudes are ignored. (Compare a doctor who undertakes an operation that has a 51% chance of generating a slight benefit and a 49% chance of killing the patient, and likewise fails to prescribe a medicine that has a 10% chance of saving the patient's life and a 90% chance of failing to do so while causing a rash.) For further analysis in the general legal domain, see Kaplow, *Burden of Proof*, *supra* note 6; Kaplow, *Optimal Burden of Proof*, *supra* note 6; and Kaplow, *Likelihood Ratio Tests*, *supra* note 6, at 13–26. See also C. Frederick Beckner III & Steven C. Salop, *Decision Theory and Antitrust Rules*, 67 ANTITRUST L.J. 41, 61–62 (1999) (advancing the criticism in the antitrust context); Michael L. Katz & Howard A. Shelanski, *Merger Analysis and the Treatment of Uncertainty: Should We Expect Better?*, 74 ANTITRUST L.J. 537, 543–46 (2007) (addressing the problem of using probability thresholds in merger decisionmaking).

<sup>13</sup> Step 1 fails (resulting in no liability) only if  $H = 0$ , but since we have taken  $B$  to be nonnegative, it is then impossible to have liability under balancing, which requires  $H > B$  ( $0 = H > B \geq 0$ , which implies  $0 > 0$ , is impossible). Step 2 fails (resulting in liability) only if  $B = 0$ . But since, to get to step 2, we had  $H > 0$ , we necessarily have  $H > B$ , which likewise results in liability under balancing. As will be discussed further, particularly in section IV.A on the conduct of legal proceedings and in section IV.B on screening, even a so-called zero threshold can matter in a nontrivial manner, depending on the interpretation of burdens of production and rules regarding motions to dismiss, summary judgment, and judgment as a matter of law. In light of the foregoing, this observation is particularly relevant to what might otherwise be depicted as pure balancing.

<sup>14</sup> The precise characterization will emerge in section B.

<sup>15</sup> See Kaplow, *supra* note 1.

<sup>16</sup> The analysis abstracts from the level of generality with which  $H^*$  and  $B^*$  may be set: for an area of law? a narrower class of cases? or case by case, using some initial indicator of a case's traits (for example, sales revenue or the number of employees)? In actual legal applications, this question seems largely to be unanswered even in statements of doctrine or proposed tests that imply that the thresholds are nontrivial.



how these magnitudes affect the quality of the resulting decisions. As section B now demonstrates, these optimal levels are indeed zero, which is to say that unconstrained balancing is optimal in this basic environment.<sup>17</sup>

### B. Final Decision Rule

This section compares this stylized structured decision procedure to balancing as a final decision rule, taking as given the set of information before the tribunal. As the introduction suggests, whenever the outcome under the former differs from that under the latter, we will have made a mistake in the sense that one of two things must be true: either no liability is assigned even though  $H$  is greater than  $B$  (so balancing would assign liability) or liability is assigned even though  $H$  is less than  $B$  (so balancing would not assign liability). The analysis here describes how these results can come about and relates them to the two decision thresholds,  $H^*$  and  $B^*$ .

Begin with cases in which the structured decision procedure results in no liability when balancing would assign liability. These arise as a direct consequence of step 1's decision threshold. No liability is assigned whenever  $H \leq H^*$ . This outcome, moreover, is determined without regard to the magnitude of  $B$ , so it is possible that  $H > B$ . This happens whenever  $B$  falls in the range from 0 to  $H$ , that is, when we have  $0 \leq B < H \leq H^*$ .<sup>18</sup> The only way to eliminate this possibility is to set  $H^*$  equal to zero, guaranteeing that the first step never matters.<sup>19</sup>

Cases in which the structured protocol results in liability when balancing would assign no liability are, correspondingly, a subset of outcomes in which step 2 is binding. There, liability is assigned when, having found that  $H > H^*$  in step 1 (which is required to reach step 2), we also have  $B \leq B^*$ . This outcome does depend on the magnitudes of both  $B$  and  $H$ , but it is not determined by a direct comparison of the two as it would be under balancing. A divergence in outcomes can arise when  $B^*$  is sufficiently large because then it is possible that  $H < B$  even though step 2 fails. This divergence happens whenever  $B$  falls in the range from  $H$  to  $B^*$ , that is, when we have  $H^* < H < B \leq B^*$ .

In reflecting on this result, it is useful to focus on the possible relationships between  $H^*$  and  $B^*$ . First, suppose that  $H^* < B^*$ . In this case, we can have the aforementioned error in which liability is assigned even though  $H < B$ . The failure of step 2 indicates only that  $B \leq B^*$ . But, because  $H^* < B^*$ , this means that an error *might* occur when  $B$

<sup>17</sup> On the possibility that  $H^*$  may optimally exceed zero to aid in screening, see section IV.B.

<sup>18</sup> Because ties are ignored, the reader should feel free to interpret all of the non-strict inequalities (here, the less-than-or-equals signs) as if they were strict inequalities (less than). The non-strict inequality when comparing  $B$  or  $H$  to 0 arises because  $B$  and  $H$  are not assumed to be positive, merely nonnegative. The other non-strict inequalities (for example,  $H \leq H^*$ ) are simply the negation of strict inequalities (here, the assumed failure of the condition that  $H > H^*$ ).

<sup>19</sup> Except in the case of ties, which, as mentioned, are ignored.

is in the range from  $H^*$  to  $B^*$ , and it *will* occur when  $B$  is in the range from  $H$  to  $B^*$ , because then we will have, as just stated,  $H^* < H < B \leq B^*$ . The first inequality indicates that step 1 passed, the last inequality tells us that step 2 failed (resulting in liability), and the middle inequality shows that liability is undesirable.

Second, suppose that  $H^* > B^*$ . Then this problem cannot arise, which is immediately apparent from the just-mentioned inequality sequence. Put more directly, in this case the combination of  $H > H^*$  (the first step passes) and  $B \leq B^*$  (the second step fails) obviously guarantees that  $H > B$ , so that liability is appropriate. We have  $B \leq B^* < H^* < H$ .

Observe that the impossibility of mistakenly assigning liability in this second case arises precisely because step 2 is rendered redundant in the following sense: If, after step 1, we had gone straight to balancing, we would have assigned liability only when  $H > B$ . But we already know from step 1 that  $H > H^*$ , and we are assuming that our protocol sets  $H^* > B^*$ , which implies that, as we leave step 1, we know that  $H > B^*$ . It is pointless to ask first whether  $B$  is at least as high as  $B^*$  when, if it is, we will then immediately ask whether it is at least as high as  $H$ , a more demanding test.

Whenever  $H \leq B$ , nothing is added by asking as well whether  $B$  meets what we already know, going into step 2, to be the weaker threshold,  $B^*$ . Moreover, whenever  $H > B$ , we will assign liability anyhow, so nothing is saved by first checking whether  $B > B^*$ . In various areas of law, it is sometimes suggested in particular legal contexts that structured decision procedures are appealing because they save effort by avoiding difficult balancing.<sup>20</sup> But the balancing of  $H$  and  $B$  is hardly difficult when  $B \leq B^*$  and we know as well that  $H > H^*$ . Instead, we may sometimes be making extra work for ourselves. Consider cases in which it is a close question whether  $B > B^*$ . Then, we need to struggle with step 2 even though the final outcome may well be immediately obvious if we peeked ahead to the balance required in step 3, which in this instance is not as close a call.<sup>21</sup>

Third, suppose that  $H^* = B^*$ . This case is essentially the same as the second. To enter step 2, we know that  $H > H^*$ . Hence, asking in step 2 whether  $B > B^*$ , which is equivalent in this case to asking whether  $B > H^*$ , is pointless: If it is not, step 3's balance would favor liability by a greater (and thus easier-to-determine) margin. And if it is true that  $B > H^*$ , we must also undertake step 3's more demanding balance in any event.

<sup>20</sup> See Kaplow, *supra* note 1 (discussing advocacy of a structured decision procedure in antitrust).

<sup>21</sup> Suppose, for example, that  $B^* = 50$ , and in step 1 it was determined that  $H = 100$ . Moreover, in step 2, imagine that there is significant uncertainty about where  $B$  falls in the range from 40 to 60. In principle, the structured decision procedure requires that we resolve that uncertainty about  $B$  in order to complete step 2, even though if we skipped that step and went straight to balancing, the outcome would be obvious.

Combining the three cases, we learn that step 2 matters to outcomes only when it dictates the wrong one, and in all other cases its only possible difference from balancing is that it may require greater effort. Note further that the point about possible added effort also applies to step 1, if we now think ahead two steps: Even when step 1 passes, that is,  $H > H^*$ , we then have to determine  $B$ , which we would have to do under balancing as well. Also, if step 1 was close, we may have had to undertake extra effort even if it would ultimately turn out that the  $H > B$  balancing in step 3 would have been easy because  $B$  was notably lower than  $H$ . Of course, step 1 sometimes does save work, namely, the need to examine  $B$  at all when step 1 fails — but this savings arises precisely in the cases in which we may be reaching the wrong outcome on account of step 1's decision threshold.<sup>22</sup>

It is also useful to reconsider step 2 from another perspective, asking what is the relationship between  $B^*$  and  $H$  rather than between  $B^*$  and  $H^*$ . Now,  $B^*$  is a component of our legal rule and  $H$  a conclusion regarding the facts of a particular case, so this is not a comparison of like types. Nevertheless, because we are interested in the relative merits of the stated structured decision procedure and balancing, this question is a natural one to consider, particularly since we had to assess  $H$  in step 1. Again, there are three cases. (Throughout, we will again suppose that we have reached step 2, so that  $H > H^*$ .)

First, suppose that  $B^* > H$ . Here, we know that we can get the wrong outcome. The genesis of this sort of error is now even more apparent. After all, we are asking in step 2 whether  $B > B^*$ , knowing full well that  $B^* > H$ . That is, we are applying a patently more stringent test than the situation demands, so it is obvious that we might err, specifically, by assigning liability due to the failure of step 2 when  $B^* \geq B > H$ .

Second, suppose that  $B^* < H$ . In this case, we cannot err. Instead, what is apparent is that we are asking a pointless question because the assessment of whether  $B > B^*$  will immediately be followed (if it passes) by the more stringent test of whether  $B \geq H$ . And, when the step 2 test,  $B > B^*$ , fails, it would have been clearer that the more stringent test of whether  $B \geq H$  would have failed, so we are hardly easing the decision task.

Third, suppose that  $B^* = H$ . This configuration avoids the above problems. But it does so by converting the second step into step 3's balancing inquiry (except for ties). After all, asking whether  $B > B^*$  and whether  $B > H$  pose the same question when  $B^* = H$ .

---

<sup>22</sup> As elaborated in section IV.B on screening, setting  $H^*$  somewhat above 0 in the first step may be justified by cost savings with regard to determinations of  $B$  and by discouraging case filings (which, of course, is not an unmitigated benefit since some cases eliminated may involve  $H > B$ , but the frequency and extent of any such loss is modest if  $H^*$  is fairly low). Viewed as a whole, avoiding the effort of balancing itself would typically be a secondary factor in determining optimally how much to raise  $H^*$  above 0.

Taken together, we can see that, if we reach step 2, it makes the most sense — regardless of what our structured decision procedure might otherwise require — to simply skip the step. Or, equivalently, we can interpret  $B^*$  as if it equals the  $H$  we determined in step 1, so that step 2 is converted into a balancing test (leaving only the errors from step 1).

Summarizing this section, if in step 1 we always set  $H^* = 0$ , and then in step 2 we always set  $B^* = H$ , all the shortcomings of the structured decision procedure as a final decision rule are avoided — because we have converted it into an unconstrained balancing test.<sup>23</sup> The most significant point, however, is that whenever the two decision methods generate different results, the outcome under the structured decision procedure is always the one that is in error. Moreover, such errors can be made even when all the information required for balancing has already been processed (which occurs when errors are made at step 2).<sup>24</sup> Perhaps most remarkable, structured protocols are thought to prevail in important areas of law, and are sometimes proposed as replacements for balancing in others, without having even asked how the outcomes under the two approaches differ.

### C. *Less Restrictive Alternatives*

It is common under structured decision procedures — indeed, particularly under them<sup>25</sup> — to append in some formal fashion an inquiry into less restrictive alternatives. (This Article uses the terminology of antitrust law<sup>26</sup> for the sort of supplement that is referred to in Title VII disparate impact law as alternative employment practices and in constitutional law’s strict scrutiny as narrow tailoring — or, in proportionality analysis, as minimal impairment.) The central idea is that, when a defendant purports to justify an action by reference to its producing  $B$ , we should consider whether some or all of that  $B$  might be achieved through an alternative arrangement that causes less  $H$ .

This section’s inquiry is animated by the question of how less restrictive alternatives analysis relates to the three-step structured rule with

<sup>23</sup> As mentioned earlier, we can also convert the structured protocol into balancing by setting  $B^* = 0$  in step 2. The discussion in this section further implies that any  $B^* \leq H$  will suffice, except that we may waste effort determining whether  $B > B^*$  when it is obvious that  $B < H$ .

<sup>24</sup> If  $H$  was merely guesstimated in step 1, a softened version of the argument in the text would apply. For example, if all we know from step 1 is that  $H > H^*$  (we decided nothing more precise about our best estimate of  $H$ ), it would seem hard to rationalize setting  $B^* \neq H^*$ .

<sup>25</sup> As will emerge, this probably is not an accident. Although a reminder to consider alternatives is often useful, the analysis in this section suggests that the rigidity of structured decision procedures, with their suboptimal separation of the analysis of  $H$  and  $B$ , may be what motivates the creation of an explicit doctrinal requirement regarding less restrictive alternatives.

<sup>26</sup> This is the rubric under the rule of reason; for mergers, the test is phrased as a requirement that purported efficiencies ( $B$ ) be “merger-specific.” For discussions, with references, to all the applications noted in the text, see Kaplow, *supra* note 1.

which we began. Specifically, where does it fit in and how does it amend the conclusion regarding liability that otherwise would be reached? Less restrictive alternatives inquiries can better be understood against the background of our previous comparison of balancing to structured decision procedures. Accordingly, this section considers first how less restrictive alternatives would be considered under unconstrained balancing and then uses the results to address our questions about how they may operate under the three-step regimen.

*Balancing.* — Under balancing, we ask whether  $H > B$ . When that is so, liability is assigned. Whether the defendant might have achieved its  $B$  in some other fashion and hence its  $B$  should be discounted or ignored (treated as zero) is moot.<sup>27</sup> (Note that when liability is thereby assigned, the defendant may find it advantageous to pursue a less restrictive alternative, if there is one that would not itself give rise to liability.) Accordingly, we will focus on the case in which  $H \leq B$ .

Suppose that there is a single less restrictive alternative that seems worth considering.<sup>28</sup> It will generate its own levels of harm and benefit, which here will be denoted  $H'$  and  $B'$ , respectively. And, if the alternative is indeed less restrictive, we have that  $H' < H$ .<sup>29</sup> For the postulated less restrictive alternative to also be a more desirable one, it must be that the net social harm generated by the alternative is less than that generated by the defendant's original action. Using our notation, this requirement is that  $H' - B' < H - B$ .<sup>30</sup>

Because this inquiry into less restrictive alternatives only matters when  $H \leq B$ , we can further state that  $H' - B' < H - B \leq 0$ . We already

---

<sup>27</sup> Keep in mind that the present discussion concerns final decisions, not information gathering. Less restrictive alternatives analysis also has significant implications for the latter. For example (some others are discussed later in this section), if it is difficult to determine  $B$  but easy to see that there is a less restrictive alternative that greatly reduces  $H$  without much affecting  $B$  (whatever its magnitude may be), it would be optimal to examine that information early in the process, conclude that there should be liability, and thus be spared the effort of assessing  $B$  more closely. Also, if the consideration of a less restrictive alternative does involve additional information-gathering costs, it may not be optimal for trivial superiority to be sufficient for liability, although a competing consideration is that the prospect of such a consideration would deter prospective defendants from adopting inferior practices in the first place. When we add uncertainties in a decisionmaker's assessments to their unpredictability *ex ante*, a conjecture is that it may continue to be optimal to require a less restrictive alternative to be nontrivially superior. This point is related to the discussion in section IV.B of setting  $H^* > 0$  as a screening device.

<sup>28</sup> As suggested in the preceding footnote, the present discussion importantly abstracts from information gathering, which is a first-order concern in the possible search for and assessment of potential less restrictive alternatives. For further discussion of some aspects of this question, see the discussion later in this section and in note 45.

<sup>29</sup> As a matter of logic, the analysis that follows suggests that it could be optimal to assign liability, even though  $H \leq B$ , due to the existence of a *more* restrictive alternative because the higher  $H$  might be more than offset by an even greater  $B$ . This possibility will not be considered further, in part for the sorts of reasons suggested in note 45. See also *infra* note 40 and accompanying text.

<sup>30</sup> Just as elsewhere in this Article, ties are ignored. (In this instance, it would make little sense to condemn a practice because there exists an alternative of merely equal social impact: had the defendant instead adopted such an alternative, it would then have been found liable for not using the original version.)

knew that the defendant's original action was, standing alone, not socially undesirable (and, absent a tie, desirable), so this observation reaffirms that substituting the less restrictive alternative is even more desirable. Note that, in the process of considering this series of inequalities, we really are performing two balancing tests: the original one (which, if it had been  $H > B$ , would have resulted in an assignment of liability with no need to inquire into less restrictive alternatives) and a second one (comparing the alternative to the original practice).

It is also helpful to restate the less restrictive alternatives test — the second balance just mentioned — in what will be referred to here as a delta/delta test. Starting with  $H' - B' < H - B$  from just above, we can rearrange terms to express this equivalently as  $H - H' > B - B'$ . That is, a less restrictive alternative is superior to the original action if and only if it reduces the harm by more than it reduces the benefit. Introducing the further notation  $\Delta H = H - H'$  and  $\Delta B = B - B'$ , this rearranged version can also be written as the requirement that  $\Delta H > \Delta B$ . So, instead of performing a second balancing test, we can directly compute the two deltas and see which is greater. Note that, whichever formulation is employed, it is often necessary to quantify  $H$ ,  $B$ ,  $H'$ , and  $B'$  in order to assess whether the less restrictive alternative calls for changing the decision from no liability to liability (although this is not always so).<sup>31</sup>

The foregoing logic of how to analyze less restrictive alternatives under unconstrained balancing is straightforward. The conclusions follow directly from the concept of balancing itself. Indeed, in some settings in which balancing is employed, the consideration of what might be referred to as less restrictive alternatives is seamless.<sup>32</sup> In a simple negligence inquiry, for example, it might be said that the balancing test just *is* a less restrictive alternatives test. An allegation of negligence, after

---

<sup>31</sup> In a number of legal domains, it is sometimes suggested that less restrictive alternatives analysis is generically easier to perform than a pure balancing test regarding the original practice. For references, see Kaplow, *supra* note 1. As a broad claim, this view is implausible, for to determine  $\Delta H$  and  $\Delta B$  or to perform the second of the two balances, it is usually necessary to have determined both  $H$  and  $B$ . However, as considered briefly at the end of this section, there may be particular cases or classes of cases in which the outcome of the less restrictive alternatives inquiry is obvious, in favor of liability, even though the determination of whether  $H > B$  is more difficult. In addition to cases in which  $\Delta H$  is obviously positive and  $\Delta B$  is obviously zero, it may sometimes be easier to measure the two deltas directly — focusing on the differences between the challenged practice and the less restrictive alternative rather than on the differences between either and inaction. To believe that the assessment of less restrictive alternatives is generically rather than contingently easier, such cases would have to be typical whereas undertaking the primary balance ( $H > B$ ) would need to be obvious infrequently.

<sup>32</sup> This is not to deny that reminders can be helpful. For example, in performing cost-benefit analyses of environmental and safety regulations, a recurring practical issue is whether sufficient attention has been given to alternatives that would be less costly and still substantially effective.

all, asserts that there is something the defendant could have done differently that would have reduced the risk of harm sufficiently to justify the additional cost. Let  $H$  and  $B$  refer, respectively, to the (expected) harm and benefit (to the defendant<sup>33</sup>) of what the defendant actually did, and let  $H'$  and  $B'$  refer to the lower level of harm and benefit associated with what the plaintiff alleges the defendant should have done. Then, the plaintiff's claim of negligence amounts to arguing that  $\Delta H > \Delta B$  — that the reduction in expected harm exceeds the reduction in benefit (increase in the defendant's costs) — precisely our delta/delta test for less restrictive alternatives.

*Structured Decision Procedures.* — Let us now consider how an inquiry into less restrictive alternatives might fit into the three-step structured decision procedure set out in section A. Because the inquiry concerns, in essence, whether the defendant's  $B$  should be deemed to justify the resulting  $H$  (even though  $H \leq B$ ), the possible placements are: as part of step 2, as an intermediate step between step 2 and step 3, as a component of step 3's balancing, and as a substitute for step 3's balancing. Each will be considered in turn.

As we will see, in some cases the differences may be little more than semantic, but in others, not. And in some instances, we will see that the implications are fanciful, further revealing the incompleteness of prior analysis of how the assessment of less restrictive alternatives should be related to the underlying structured decision procedure to which it is attached. The most important question, however, will be how less restrictive alternatives, when incorporated into the structured decision procedure, affect the bottom-line liability decision. We can anticipate that the answers might have some interesting twists because we already know that incorrect outcomes readily arise in the basic setting in which no less restrictive alternative exists.

In all of the cases, we will assume that  $H > H^*$ , for otherwise we never reach step 2. As a consequence, we know from the outset that less restrictive alternatives analysis cannot save us from the problem that step 1 may well exonerate a defendant even though  $H > B$ .

The situation here is actually worse when we take into account the possibility of less restrictive alternatives, which will not be reached under a structured decision procedure when step 1 fails. First, there can exist cases in which: (1) we assign no liability due to the failure of step 1, and (2) this would not have involved a mistake in the basic case that we considered in section B because it happened to be true that  $H \leq B$  (even though the decisionmaker did not check this), but (3) had we engaged in unconstrained balancing, including the appropriate consideration of less restrictive alternatives described just above, liability would

---

<sup>33</sup> In an ordinary negligence setting, it is assumed that the benefit to the defendant (say, lower cost) is a social benefit. By contrast, this often is not true in some areas in which structured decision rules are employed, as will be elaborated later in this section.

have been appropriate because  $\Delta H > \Delta B$ . Second, in some of the cases in which no liability is erroneously assigned in the basic setting, the magnitude of the error resulting from this mistake will be larger: some of the  $B$  (even though less than  $H$ ) might have been retained had liability been assigned because the defendant would then have employed a less restrictive alternative. Put another way, the net benefit from correctly assigning liability, as the decisionmaker would have done under an unconstrained balancing test, might have been larger because the forgone  $B$  would have been smaller.

Let us now turn explicitly to the different possible locations for less restrictive alternatives analysis. Begin with the case in which it is located in step 2, which asks whether  $B > B^*$ . Suppose that we consider less restrictive alternatives only when this step would be passed. That is, if  $B$  is not in any event high enough to avoid the assignment of liability at step 2, there is no need to consider whether some of that  $B$  should not be counted against liability. An immediate implication is that less restrictive alternatives analysis fails to rescue us from the other major decisional defect of structured decision procedures. As section B explains, such a rule assigns liability at step 2 when  $H > H^*$  (required to reach step 2) and  $B \leq B^*$ ; yet, whenever  $B^* > H$ , there is a range for  $B$ , specifically  $H < B \leq B^*$ , such that we erroneously assign liability.

Note that, although less restrictive alternatives analysis does not literally avoid this mistake because we do not get that far,<sup>34</sup> the possible existence of such alternatives means that the magnitude of this mistake will sometimes be smaller because the assignment of liability will induce the defendant to switch to the alternative. Moreover, when  $\Delta H > \Delta B$  for the less restrictive alternative, assigning liability turns out not to be a mistake, even though the analysis to determine whether this is so was not undertaken.

Returning now to scenarios in which  $B > B^*$ , consider how less restrictive alternatives analysis might play out. The core oddity is that, as we have seen, the proper way to examine less restrictive alternatives involves quantifying  $H$ ,  $B$ ,  $H'$ , and  $B'$ , in order to perform the second balancing test or the delta/delta test. But we are, after all, only at step

---

<sup>34</sup> We could instead insist on inquiring into less restrictive alternatives — ordinarily understood to negate part of the defendant's claimed  $B$  — in aid of the defendant. As explained above, a full (proper) analysis may involve performing two balances, including the original one that is deferred to step 3 under the structured decision procedure. If, thereby, we can move forward the balancing test and exonerate the defendant who would erroneously have been found liable, then we could avoid this error. Essentially, we would be sneaking balancing in through the back door. (Although it is sometimes said that less restrictive alternatives analysis is indeed a way to engage in balancing amidst tests that purport not to, I am unaware of situations in which a defendant argues for the possibility of a less restrictive alternative with the intention of losing that argument but hoping to prevail in the end because, in performing such analysis, it will be discovered that, under the deferred or omitted balancing test, the defendant should win for other reasons.)



2, which has no balancing; the explicit comparison of  $H$  and  $B$  is deferred to step 3.<sup>35</sup> So the decisionmaker, under this location for less restrictive alternatives analysis, often<sup>36</sup> must do all that is required for full balancing, and more, even though the balancing step has yet to be reached.

A related observation is that step 2 purports to be only about  $B$ , yet placing less restrictive alternatives analysis in this step makes it about  $H$  as well. Indeed, how could it be otherwise? After all, it is called a less *restrictive* alternative, and the meaning of the emphasized term is that the alternative reduces  $H$ , so by definition such analysis is at least partly about  $H$ .

Another surprising feature is that the effort involved with less restrictive alternatives analysis seems to be mandated even if  $H > B$  to begin with, in which case it seems pointless. As explained in section B, depending on how  $B^*$  is set, there can exist cases that pass step 2 even though  $H > B$ . Indeed, a structured decision procedure's inclusion of step 3, with balancing, presumes such a possibility. (Why have a third step with balancing if it would always fail to result in liability?) Now, a challenger (whether the government or a private plaintiff) to a defendant's action need not raise the possibility of a less restrictive alternative and would have no interest in doing so if it was certain that the decisionmaker would find at step 3 that  $H > B$  and thus assign liability regardless. But often the challenger will not be so sure, at least at the outset, what the decisionmaker will conclude about the balancing of  $H$  and  $B$ . So, as the decisionmaker undertakes less restrictive alternatives analysis while performing step 2, it may well discover in the process that  $H > B$ . Indeed, if it undertakes the second balancing test for a less restrictive alternative, it must quantify  $H - B$  for the right side of that inequality, and it will be hard not to notice whether this is positive or negative. But the decisionmaker still has further work to do in order to determine the outcome of the second balancing (it must determine the other side of the inequality, when that may well be moot) — or, equivalently, to perform the delta/delta test.<sup>37</sup> A sensible decisionmaker who discovers along the way that  $H > B$  would skip step 2, go straight to step 3, and assign liability, without resolving all aspects of the less restrictive alternative, or even bothering to decide whether  $B > B^*$  for

---

<sup>35</sup> And under some variants of structured decision procedures, there is no step 3 balancing at all, in which event the predicament is even stranger.

<sup>36</sup> But not always. See *supra* note 31.

<sup>37</sup> In those cases in which it finds in undertaking this inquiry that  $H \leq B$ , less restrictive alternatives analysis could matter, changing what otherwise would be a decision of no liability into an assignment of liability. A further (purely semantic) curiosity is that, in this situation, if the less restrictive alternative fails, the decisionmaker already knows that  $H \leq B$ , so there is nothing left to do at step 3 except to announce that result, which it formally is not supposed to do at step 2, where it first learned that this was so.

---

---

that matter. (By similar logic, the decisionmaker might as well skip steps 1 and 2 in all cases, going straight to step 3's balancing, but that of course violates the structured decision procedure, as does the more moderate act of civil disobedience involved in short-circuiting only step 2.<sup>38</sup>)

Next, suppose that less restrictive alternatives analysis is located between what we have been calling step 2 and step 3. Then, this inquiry can only be reached in cases in which we have determined that  $H > H^*$  and  $B > B^*$ . A number of observations, flowing directly from the preceding discussion, are apparent. First, any decision errors produced by the structured decision protocol cannot be corrected. They involve mistaken assignments of no liability in some of the cases that fail at step 1 and mistaken assignments of liability in some of the cases that fail at step 2. Because this location of the less restrictive alternatives test (and those considered next) means that, when reaching it, we must have passed both of these prior tests, less restrictive alternatives analysis cannot fix the incorrect outcomes, all of which were produced by conclusions at steps 1 and 2 that involve a final decision on liability that terminates further inquiry.

Second, most of the aforementioned anomalies that arise when less restrictive alternatives analysis is located within step 2 remain: the decisionmaker is doing all that is necessary to resolve the case (because  $H$ ,  $B$ ,  $H'$ , and  $B'$  must all be quantified), and all the work required to perform less restrictive alternatives analysis is undertaken even when  $H > B$  in any event. The main difference is that the disharmonious optics of examining  $H$  (and  $H'$ ) at a step that purports to be only about  $B$  are avoided.

Now suppose that less restrictive alternatives analysis is located within step 3's balancing test. Then, at least in cases that get this far, we are back to the earlier depiction of how less restrictive alternatives inquiries are properly conducted as part of unconstrained balancing. The main difference from the prior locations is not in outcomes but in possible reductions in the requisite effort and degree of discord: it would be entirely appropriate, when the decisionmaker stumbles on the fact that  $H > B$  while conducting its less restrictive alternatives analysis, to assign liability without further ado. (In comparing the preceding locations, it should be noted that the dissonance that they can generate may not be cost-free because it may produce confusion or produce perhaps unconscious contortions of the analysis in attempts to reduce it.)

---

<sup>38</sup> If a decisionmaker announced, midstream in step 2, that it had concluded  $H > B$ , the parties might not object to the step being skipped. Or the decisionmaker might simply decide the less restrictive alternatives question in the defendant's favor. However, if a system of review reversed the step 3 balancing conclusion, which had led to the assignment of liability, but not the step 2 analysis of the less restrictive alternative, then further error might result.

Finally, there can exist structured decision procedures that have no step 3 and associated balancing. As mentioned previously, upon a finding in step 2 (if we get there) that  $B > B^*$ , such a rule may simply assign no liability and stop. In such a situation, appending less restrictive alternatives analysis as an additional, final step constitutes an improvement. Then, in some (but only some) of the cases in which the defendant would otherwise have been exonerated, there may be an assignment of liability when that would be optimal, notably, when the delta/delta test passes, that is, when  $\Delta H > \Delta B$ .

Notice, however, that a less restrictive alternatives test is not a full substitute for having a final balancing step, and if the less restrictive alternatives test replaces balancing rather than filling a void, the results could be worse overall. The reason is simple: even without regard to less restrictive alternatives, we know from section B that it is possible to have  $H > H^*$  and  $B > B^*$  (the latter implying that, under the contemplated rule that replaces step 3's balancing with an analysis of less restrictive alternatives, there would be no liability), even though  $H > B$ . Whenever  $B^* < H$ , we know that this is possible because then we may have  $B^* < B < H$ . This scenario is particularly strange. The decisionmaker is required to assign no liability if there is no less restrictive alternative even though the original action is undesirable: its  $B$  was not high enough to justify its larger  $H$ . The decisionmaker will, thankfully, assign liability if there is an effective less restrictive alternative. But when the less restrictive alternative fails, there is no liability even though the decisionmaker, in conducting its less restrictive alternatives analysis (or, actually, upon completing step 2), directly confronts the fact that  $H > B$ .

This final point combined with some of those above shows how less restrictive alternatives analysis does have a role and, when properly undertaken, has many features of balancing, but it does not come close to rectifying the errors from failing to employ unconstrained balancing in the first place. And it compounds them if the choice to assess less restrictive alternatives replaces balancing that otherwise would have been performed, even as a final step in a structured decision procedure.

*Equal Effectiveness.* — Turn now to another question that has arisen with respect to the analysis of less restrictive alternatives under structured decision procedures. It is sometimes suggested that only less restrictive alternatives that are equally effective can count in the sense of being permitted to change a decision from no liability to liability when we initially had  $H \leq B$ .<sup>39</sup> Note, however, that no special role was played

---

<sup>39</sup> See, e.g., C. Scott Hemphill, *Less Restrictive Alternatives in Antitrust Law*, 116 COLUM. L. REV. 927, 943–46 (2016) (citing cases and examining the question in the antitrust context). Returning to our discussion of how less restrictive alternatives analysis was, in a sense, the core of the negligence test, the analogue would be that negligence could arise only if a proposed precaution that was not taken would have been literally cost-free to implement.

in the foregoing discussion by whether  $\Delta B = 0$ , which is to say by whether the less restrictive alternative was equally as effective, with regard to  $B$ , as the original action was. Recall that what less restrictive alternatives do, by definition, is reduce  $H$ , which is to say, they involve  $\Delta H > 0$ . When shifting the focus to  $\Delta B$ , as we are now, the relevant inquiry is our delta/delta test, which asks whether  $\Delta H > \Delta B$ . If the alternative is less restrictive, so that  $\Delta H > 0$ , and it also happens to be true that  $\Delta B = 0$ , then this test is obviously satisfied. Although  $\Delta B = 0$  is a sufficient condition, it is just one special case.

To elaborate, suppose instead that we have  $\Delta B < 0$ : that is, our less restrictive alternative involves a negative reduction in  $B$ , which is more clearly expressed as an increase in the social benefit. In our negligence setting, in which  $B$  may refer to avoided costs (greater profits), we would not expect this possibility to be important because a defendant has every incentive to save costs even without the threat of liability (and all the more so if a cost savings might also help to avoid liability). But in many legal domains, there can be a greater social benefit even when the defendant's private benefit is lower; or, put the other way, a greater private benefit may be associated with a smaller (even negative) social benefit.<sup>40</sup>

In addition — and where actual contention about less restrictive alternatives sometimes arises — it is entirely possible that  $\Delta H > \Delta B$  (making the alternative socially superior, thereby justifying liability under a balancing test) even though  $\Delta B > 0$ . That is, an alternative may reduce the benefit  $B$  yet reduce  $H$  more, indeed, possibly much more. Taking the extreme case, consider an alternative that entirely eliminates a substantial  $H$  at slight inconvenience to the defendant, so that  $\Delta B$  exceeds zero by just a bit. It would be absurd to assign no liability, needlessly suffering great harm to avoid losing a negligible benefit. This point is all the more stark when we consider that, under a requirement that less restrictive alternatives be equally effective, we would assign liability if  $\Delta B$  were zero rather than tiny, even when the savings in harm ( $\Delta H$ ) was minuscule rather than massive.

The central justification for contemplating a requirement of equal effectiveness is pragmatic: sometimes determining whether  $\Delta B = 0$  will be easy, whereas determining whether  $\Delta H > \Delta B$  in cases in which

---

<sup>40</sup> In the contexts in which structured decision rules are employed, gains to a defendant are often due precisely to the presence of social costs. In antitrust, anticompetitive effects can be what makes an action profitable. In employment discrimination, a defendant that wants to discriminate prefers, by hypothesis, to do so even though the law regards its "benefit" to be a social cost. In constitutional law, a government may seek to stifle dissent to entrench itself, which is precisely what makes the action socially detrimental. This observation raises the additional question, which will not be considered further here, of whether the less restrictive alternative would be profitable or otherwise sufficiently appealing to the defendant. If it would not, assigning liability will indeed result in the loss of a socially relevant  $B$  that exceeds the  $H$  that is prevented.

$\Delta B > 0$  may be more difficult. Consider briefly the premise of this argument (setting to the side whether, if the premise holds, the conclusion follows, which obviously it may not<sup>41</sup>): Which is easier to determine, whether  $\Delta H > \Delta B$  or whether  $\Delta B = 0$ ? It seems clear that sometimes either may be less demanding.

In some cases, it may seem obvious that  $\Delta B = 0$ , namely, when a defendant's conduct involves the bundling together of two essentially distinct activities.<sup>42</sup> If one of them causes much (or even some) of the harm and *none* of the benefit, then it can be severed from the other part of the bundle. The less restrictive alternative in this case is just that other part. For example, in the antitrust setting, if two firms form a joint venture to develop a new way of producing a component of one of their products and also, as part of the formation agreement, agree to fix their prices on some other product, the less restrictive alternative consists of doing the former without the latter.<sup>43</sup> Or, with constitutional law, if a legislature enacts a single bill that combines funding to address a serious epidemic with an unrelated restriction on speech, it is easy enough to invalidate the latter, leaving the former — which, by itself, is the less restrictive alternative.

These examples involve flagrant violations. Outside such obvious situations, the requirement of equal effectiveness is strong: there really needs to be no synergy or other cost savings that is lost by the severance or other proposed substitution. Even pure price fixing, when practiced in certain ways, may save costs (say, of hiring separate price-setting employees at each firm), a benefit that would be forgone if the aforementioned alternative were substituted for the defendants' plan. And requiring an employer in a Title VII disparate impact case to substitute a newly developed and equally effective test that predicts job performance would involve some transition costs, so  $B$  would be reduced somewhat. That is, suggestions that many proffered less restrictive alternatives involve  $\Delta B = 0$  rather than a  $\Delta B$  that is low, and surely lower than  $\Delta H$ , are often false and perhaps disingenuous.

More broadly, determining whether we are in the knife-edge case in which  $\Delta B$  precisely equals zero will often be difficult (that is, if we do not simply rule it out altogether in light of the foregoing sorts of considerations). If an inquiry into whether  $\Delta B = 0$  is really an inquiry into whether  $\Delta B$  is near zero, the domain is larger. In many cases, however, we may have to engage in substantial effort to quantify both  $B$  and  $B'$  fairly accurately in order to determine this, in light of whatever "near" zero is taken to mean.

---

<sup>41</sup> It often is not optimal to substitute a substantively inferior rule just because it is easier to apply.

<sup>42</sup> And sometimes it will be obvious that  $\Delta B < 0$  (which, as explained above, can arise when the defendant's private benefit is a social cost).

<sup>43</sup> In antitrust law, this sort of inquiry is associated with the ancillary restraint test of *United States v. Addyston Pipe & Steel Co.*, 85 F. 271, 281–91 (6th Cir. 1898), *aff'd*, 175 U.S. 211 (1899).

By contrast, it will sometimes be fairly obvious that  $\Delta H > \Delta B$  even if it is hard to pin down precisely how close  $\Delta B$  is to zero. For example, an alternative may essentially eliminate a large  $H$  at a disputed but modest loss of  $B$ ; elaborate investigation to determine just how small  $\Delta B$  really is and arguing over whether that rather small amount is “small enough” hardly seems worthwhile. Indeed, even if it is clear that  $\Delta B$ , although modest, is not close enough for the alternative to be deemed equally effective, it may nevertheless be clear that it is less than  $\Delta H$ . That is, the delta/delta analysis and the equal effectiveness inquiries may both be clear — the former favoring liability and the latter opposing liability. In such cases, choosing the latter means choosing to be clearly wrong when the decisionmaker could instead have assigned liability and been just as clearly right.

Reflecting on some of the analysis in section B on final decision rules, and on that forthcoming in Part III on information gathering, it seems that it will vary — by the area of law, the type of case, and indeed the particulars of a given case — which inquiries will be easier than others. It is, of course, possible that in some applications certain generalizations may be possible, which might favor particular simplifications, shortcuts, or presumptions regarding one or another aspect of how best to conduct inquiries into less restrictive alternatives, just as is true regarding any other legal inquiry.

*Remarks on Information Gathering.* — This section focuses on how the consideration of less restrictive alternatives affects final decisions regarding liability under balancing and under structured decision procedures. The discussion also relates to some of the lessons that will be presented in Part III’s analysis of information gathering and, in particular, the many ways that structured information protocols depart from optimal information collection. For the present, two observations are offered.

First, regarding the placement of less restrictive alternatives inquiries in the three-step structured decision rule, it would seem that sooner is better as far as information gathering is concerned. Just as the sequential separation of  $H$  and  $B$  will be seen to be detrimental in numerous ways, so too would the additional sequential separation of  $H'$  and  $B'$  from the inquiries into  $H$  and  $B$ . Ideally, all would be mixed optimally.<sup>44</sup> If not, at least  $H'$  and  $B'$  might be investigated alongside  $B$  rather than afterwards.

Second, the analysis of less restrictive alternatives reinforces the point developed below that much evidence concerns both  $H$  and  $B$ , rendering separation incoherent to begin with. When discussing less re-

---

<sup>44</sup> See *supra* note 27 (offering an illustration of how the order of considering issues may be adjusted in light of what information is easiest to gather).

restrictive alternatives inquiries as part of step 2, we immediately confronted the fact that, even though step 2 purports to be only about *B*, less restrictive alternatives analysis by definition is about *H* as well.

*Reflections.* — Folding the consideration of potentially less restrictive alternatives into an unconstrained balancing test is conceptually straightforward. As with the basic balancing test itself, quantification and comparison may be highly challenging, and these difficulties pertain as well to the analysis of less restrictive alternatives.<sup>45</sup> But the basic logic is fairly simple.

By contrast, incorporating the analysis of less restrictive alternatives into a structured decision procedure seems to be a convoluted enterprise. It does not generally make matters worse, but it likewise fails to remedy the core shortcomings of these rules and, when less restrictive alternatives are important, will sometimes raise the rules' cost (by leaving more potential social gains on the table). When less restrictive alternatives are reached under a structured decision procedure, their consideration is beneficial compared to ignoring them.

Section B indicates that structured decision procedures are, upon examination, strange beasts. Their poor normative prescriptions are reminiscent of the inaccurate positive predictions from the geocentric model of the solar system (which erred as well in assuming that planetary motion was circular rather than elliptical). When a core model is fundamentally flawed in ways that are not apparent but some of the mismatch is noticed, there is a tendency to make ad hoc adjustments, such as the introduction of epicycles. This approach is often ridiculed, although it should be noted that performance is improved relative to the mediocre baseline. The real objection is that the core defects of the underlying model remain unaddressed. Observe further that the accuracy of the Ptolemaic system, with its ever-growing additions of epicycles, was fairly high. Such, unfortunately, is not the case with structured decision

---

<sup>45</sup> In an important sense, these difficulties can be greater with respect to the analysis of less restrictive alternatives because many might be considered and there is no obvious way to limit the scope of the inquiry. See, e.g., Alan Devlin, *Antitrust as Regulation*, 49 SAN DIEGO L. REV. 823, 828–29, 831–41 (2012). Moreover, such analysis will often involve hypothetical alternatives rather than actual ones. Agencies and courts are reluctant to force defendants to consider every imaginable option and, in essence, to force them to choose whatever is determined to be socially optimal, without limitation. Indeed, these concerns involving institutional competence — and, relatedly, information costs, uncertainty, and error — importantly shape underlying legal prohibitions and the design of remedies even without the expansion of the universe of possibilities involved with less restrictive alternatives analysis. (These ideas are also related to some of the motivations for the use of rules to constrain balancing that are considered in subsection IV.C.2.) For example, antitrust enforcers and courts are reluctant to second-guess firms' product designs and refusals to deal with competitors, Title VII is not broadly seen as justifying micromanagement of firms' employment decisions, and constitutional courts are disinclined to rethink a government's national security apparatus or the manner in which it designs the tax collection process. Consideration of these sorts of matters is beyond the scope of this investigation but obviously is intertwined with many aspects of legal rule design, including those considered here.

procedures, even with the useful appendage of inquiries into less restrictive alternatives.

### III. INFORMATION GATHERING

Part II analyzed final decisionmaking, taking as given the information (evidence) that is before the decisionmaker. This Part focuses on information gathering, a critical but underanalyzed feature of legal decision processes. The quality and relevance of the available information have an important influence on how well decisions will be made, and the cost of information gathering is a first-order consideration in legal system design. Moreover, structured decision procedures are sometimes favored on the ground that they economize on information costs because of the manner in which they sequence and truncate information collection, making this subject particularly pertinent to the present investigation.

This Part's treatment of information gathering will be simplified in a number of ways, most notably by examining the task as one guided in a centralized fashion by the decisionmaker. This portrayal is most apt (although still incomplete) regarding investigations and decisionmaking by specialized agencies and, to a degree, efforts in civil adjudication when significantly directed by the tribunal.<sup>46</sup> Discussion of actual practice is deferred to section IV.A, where we will see that U.S. civil litigation deviates in fundamental ways from the prescriptions of optimal information collection, which would be appropriate under balancing, and from those of structured decision procedures.

#### A. *Optimal Information Collection*

Because information collection is costly and potentially available information is vast, no one ever gathers literally all information that might possibly bear on a decision. Accordingly, it is necessary to determine how much information to collect.<sup>47</sup> The order in which information is

---

<sup>46</sup> In litigation, the two parties' incentives differ not only from each other but also from what a centralized optimization would dictate. See Louis Kaplow, *Information and the Aim of Adjudication: Truth or Consequences?*, 67 STAN. L. REV. 1303, 1352–55 (2015) [hereinafter Kaplow, *Information and Adjudication*]; Louis Kaplow, *The Value of Accuracy in Adjudication: An Economic Analysis*, 23 J. LEGAL STUD. 307, 338–45 (1994) [hereinafter Kaplow, *Value of Accuracy*]; Louis Kaplow & Steven Shavell, *Accuracy in the Assessment of Damages*, 39 J.L. & ECON. 191, 195–98, 206–09 (1996). Relatedly, it is possible that otherwise suboptimal decision or information-gathering protocols could be beneficial because they help correct parties' incentives or those of an agency or adjudicator, a subject not pursued here.

<sup>47</sup> Although this Part abstracts from civil litigants' incentives, it is worth noting that the potentially significant impositional costs — that is, costs imposed on the other party — are additional important concerns when we broaden the perspective in this way. As is familiar, both waste and the chilling of beneficial conduct can result if prospective plaintiffs can too readily initiate and



gathered also matters because not all information is equally valuable or equally costly to collect. When we stop at some point, the order of collection will determine the quality of the information that has been obtained and how much has been spent in the process.

Moreover, an optimal information-gathering process is not distinct from decisionmaking itself. After all, a decision to stop collecting information is associated with making a decision on liability, and the expected payoff of the particular decision that would be made at a given point in time, if stopping, is what should be compared to the expected payoff from further information gathering in making the decision whether to stop or to proceed.

The fundamentals of optimal information collection were developed a half century ago in connection with the emergence of the field of decision analysis.<sup>48</sup> Inquiries were motivated by problems like medical decisionmaking (order of testing and treatment decisions), industrial troubleshooting (order of testing to minimize downtime), and business decisionmaking (collection of production and marketing information to guide product launch decisions). In this field, the problem is often described as determining the value of (additional) information. That value justifies further effort if it exceeds the cost of collecting the information. Because these principles are unfamiliar for many legal analysts — or at least not crisp and front of mind — it is worth presenting central elements with some care because they illuminate the merits of different legal decision procedures.<sup>49</sup>

To begin, suppose that our decisionmaker at some, let us suppose early, stage had to make a decision based on the information in hand. It would then be necessary to estimate — perhaps guesstimate would be more apt — the magnitudes of  $H$  and  $B$ . Liability would be imposed under balancing (the focus in this section) if and only if, given these estimates,  $H > B$ .

---

conduct litigation that is very costly to defend even when there is little merit. Likewise, defendants' abilities to impose costs on plaintiffs can discourage valuable litigation.

<sup>48</sup> See, e.g., HOWARD RAIFFA & ROBERT SCHLAIFER, APPLIED STATISTICAL DECISION THEORY (1961); HOWARD RAIFFA, DECISION ANALYSIS: INTRODUCTORY LECTURES ON CHOICES UNDER UNCERTAINTY (1968); Ronald A. Howard, *Information Value Theory*, SSC-2 IEEE TRANSACTIONS ON SYS. SCI. & CYBERNETICS 22 (1966).

<sup>49</sup> For prior treatments that are suggestive of some of the points developed more explicitly here, see, for example, Beckner & Salop, *supra* note 12 (discussing aspects of the sequencing of information in the choice among certain antitrust rules that categorize conduct in different ways); Kaplow, *Information and Adjudication*, *supra* note 46, at 1332–44; and Kaplow, *Value of Accuracy*, *supra* note 46. As foreshadowed in note 6, the discussion in this Part adopts the more traditional and simpler framing of decision analysis — in which the import of a decision derives from its direct consequences — rather than attending to how the prospect of final decisions influences ex ante incentives.

*One Clump of Information.* — Next, suppose that there is a single clump of additional information that may be collected at some cost.<sup>50</sup> Should the decisionmaker stick with the aforementioned decision or make the expenditure to collect the additional information? The answer is determined using backward induction, which is a way of making precise what it means to look before we leap.

Suppose that the information was collected. What would the decision be then? When information is at all useful, we do not know for sure beforehand.<sup>51</sup> Hence, it is necessary to make further estimates about the likelihood that we would learn various things and the corresponding consequences associated with each decision (liability and no liability) in light of what we might learn.<sup>52</sup> Presuming that we would make the best decision (still under uncertainty) contingent on what we learn, we can then compute an expected value of the decisions we would make when better informed. That valuation, in turn, can be compared to the value associated with the decision we would make now, without the additional information. The difference in these two expected decisional values is referred to as the value of information. It is optimal to collect the clump of information if this value exceeds the cost.<sup>53</sup>

Reflecting on this methodology, some simple intuitions emerge.<sup>54</sup> Obviously, the lower the cost of collecting the clump of information, the wider the range of settings in which it is worthwhile to do so. Regarding

---

<sup>50</sup> It is conventional to think of the cost as monetary, but we should also include effort and the possible cost of delaying a decision, which in some contexts may be significant. There are also the costs of analyzing and otherwise processing the additional information, which is to say that, when the text refers to “gathering” information and the like, this is a shorthand for the full gamut of concomitant costs. Note that processing costs — although often small relative to gathering costs, narrowly construed — sometimes are large (and sometimes are not readily separable, because information is often processed as it is collected, such as when reading this sentence).

An interesting legal application concerns the role of bench trials. Consider why it might be that, if the judge will be the factfinder at trial, the judge should be constrained from deciding the case, one way or the other, at summary judgment. That is, why disallow prediction of the factfinder’s decision when the judge hearing the motion is the factfinder? One answer is that a trial — with all of its costs — is a means of taking the parties’ now-collected information and presenting it in a form that enables it to be processed in order to reach a decision. See *infra* note 93.

<sup>51</sup> If we did know, the decision would necessarily be the same as what we would have made without the information, in which case the information would have no value and accordingly could not justify any positive cost.

<sup>52</sup> A moment’s reflection on this statement will reveal that coming up with these estimates can be a daunting task, particularly when the decision problem is outside the domain of the information collector’s expertise. It is often recognized that specialized agencies may have advantages in analyzing information in order to make decisions, but it is insufficiently appreciated that they may have larger advantages in making choices about how optimally to gather information in the first place. As we move later in this section from there being a single clump of information to many, this point will be reinforced.

<sup>53</sup> For a simple exposition and illustration aimed at a legal audience, see HOWELL E. JACKSON ET AL., *ANALYTICAL METHODS FOR LAWYERS* 13–19 (3d ed. 2017). See also *infra* note 55 (offering a simple numerical example).

<sup>54</sup> See, e.g., Ronald W. Hilton, *The Determinants of Information Value: Synthesizing Some General Results*, 27 *MGMT. SCI.* 57 (1981).

the value of information, we can see that more information tends to be worth collecting when our uninformed decision is a close call than when it seems to be a slam dunk. If the expected harm is very high and the likely benefit is very low, it is less likely that additional information would change our preliminary decision to assign liability than when the expected harm barely exceeds the likely benefit. Similarly, when the expected harm is negligible and the likely benefit huge, additional information is unlikely to change our preliminary decision of no liability. It follows that the information-collection decision can be formulated as one involving three ranges: when  $H - B$  is very high, it will be best to assign liability; when it is in an intermediate range (straddling zero), to collect information and decide in light of what we learn; and when it is very low (in a negative range), to assign no liability.

A final consideration concerns how much we expect to learn. That, in turn, is a function of two features of the situation: How much uncertainty is there to begin with? And to what extent will the clump of information under consideration resolve that uncertainty? The first question is more obvious, but the second is equally important: even if we are highly uncertain, if the information we can collect, at a cost, is unlikely to refine our knowledge very much, it tends not to be worth obtaining. This lesson sharpens the preceding characterization of when it is optimal to collect information in that it determines the width of the intermediate range (if any) in which additional information is sufficiently valuable to justify the cost of collecting it.

As a final observation on this initial setting, note that information will matter only when it leads to a decision different from the one that would have been made without it. This point implies that the value of information relates importantly to how often we expect what we learn to change the decision we otherwise would have made. Moreover, unless the information is perfect, or at least sufficiently good that the changed decision is certain to be correct, it will sometimes be true that our subsequent decision, although based on superior information, will turn out to generate a worse outcome. The value of information, properly determined, reflects this feature as well.<sup>55</sup>

---

<sup>55</sup> To illustrate some of the foregoing points, consider a simple example involving perfect information. Suppose that either  $H = 30$  and  $B = 10$ , or  $H = 10$  and  $B = 30$ . In the former case, liability is the right decision and yields a net benefit of 20, and in the latter case no liability is right, also yielding a benefit of 20. Suppose next that, initially, we are 90% sure that the first scenario prevails. Then, if we had to decide now, we would impose liability. But 10% of the time we would be wrong, sacrificing 20. Finally, suppose that the information we could collect is perfect, indicating the true situation for sure. 90% of the time we will learn that the first scenario prevails and assign liability; 10% of the time we will learn that the second prevails and assign no liability. Because of the latter — the 10% of cases in which we change our decision — we have a gain. Moreover, in this simple case, we are never wrong because the information is perfect. So we gain 20 in that 10% of

---

---

*Two Clumps of Information.* — With appropriate iteration and permutation of this basic method, we can, in principle, determine the optimal sequencing of information collection and when it is optimal to stop collecting information and make a decision in settings with multiple clumps of information.<sup>56</sup> Most of the insights in this more general setting can be gleaned from introducing just a second clump of information that we may collect.

This slight modification significantly expands the range of possibilities: we can make a decision without collecting any information (assigning liability or not); collect the first clump and then decide; collect the first clump, then the second clump, and then decide; collect the second clump and then decide; collect the second clump, then the first clump, and then decide; or collect both clumps simultaneously and then decide. This final possibility is significant because of potential synergies: it may be cheaper to collect both simultaneously rather than to collect them sequentially.<sup>57</sup> As will be explained, however, simultaneous collection forgoes option value associated with the possibility of not collecting whichever clump would otherwise have come second.

Although there are many combinations to be considered, we can proceed in much the same way as we did with one clump of information. Suppose that we do collect the first clump. Then we are faced with the choice between deciding on liability at that point, with our given set of information, or instead collecting the second clump and then deciding. This choice is precisely analogous to the one we just analyzed. Notice, importantly, that the choice is made using what we learned from the first clump of information, which was unavailable at the outset. Hence, the best decision regarding whether to collect the second clump of information may well differ from what it would have been if we had been required to decide the matter at the outset, before learning the information in the first clump.

Consider next the path in which our first step involved the collection of the second clump of information rather than the first. The choice of

---

the cases, for an expected gain of 2. Hence, the value of information is 2, and it should be collected if and only if the cost of the information is less than 2.

<sup>56</sup> The central reason for this relates to the method of backward induction itself. If we generate a decision tree with every possible permutation, the very last decision on any branch will always be a decision with the information then in hand, which is where we started. Moving one step back, those decisions will involve whether to acquire a single (final) clump of information, which we have now analyzed as well. That analysis (whatever the outcome) generates a value from being at that point in the decision tree. Hence, when we move one further step back, we have a single question, looking forward, of whether to collect the now-next clump of information. And so on.

<sup>57</sup> One of the many reasons this may be so, as mentioned in note 50, is due to the fact that sequencing will often take more time than doing both together or, even if otherwise independent processes, in parallel. Medical testing readily brings to mind examples of all of these possibilities.

---

---

whether to decide on liability at that point or instead to proceed to collect the first clump of information is obviously the same (qualitatively) as the one we just considered when we had started in the other order.

If we solve each of these second-stage problems involving sequential information collection, we can identify the optimal choice midway on each path as a function of what is learned after the initial clump of information has been analyzed. Using those results, we can determine the value that arises if we indeed, at the outset, made the choice to collect the first clump of information or to collect the second clump of information. Then, we can compare each of those two values (subtracting the information cost pertaining to whichever clump is to be collected at the outset), both to each other and to the choice of deciding on liability up front without collecting any additional information.<sup>58</sup>

It remains to analyze the path of collecting both clumps simultaneously, which as mentioned may have appeal when there are synergies that are lost by sequencing.<sup>59</sup> Viewing these combined clumps as a single clump — which for all practical purposes they are if we are going to collect them simultaneously — we again have our original value of information problem in choosing whether to decide on liability up front or to collect a single (now aggregated) clump of information. This value of information, net of its cost, can be compared to the net valuation of the two sequential options and to making the liability decision without collecting any further information.

We have seen that the basic value of information technique developed for when there is only a single clump of information that we might collect (at a cost) enables us to tackle this more complex version of the problem with two clumps of information. Reflection on this expanded version nevertheless yields a number of additional insights.

First, comparing a one-clump-at-a-time sequential path — say, clump one followed by either a liability determination or a collection of clump two (followed by a liability decision) — to the both-clumps-at-once path illuminates the basic tradeoff between synergies and option value. When proceeding one clump at a time, if it turns out to be optimal to stop in light of what we learn from the first clump, we do not have to incur the cost of collecting the second clump. The full option value is net of forgone synergies as well as any information loss. If we find that it is not then optimal to stop, we incur the information cost for

---

<sup>58</sup> So far, we are comparing four choices at the outset: assign liability, assign no liability, collect only the first clump of information (deciding in light of what we learn whether to collect the second), and collect only the second clump of information (deciding in light of what we learn whether to collect the first).

<sup>59</sup> If there is literally no cost — involving expenditures, effort, or delay — in sequencing, then it is always better to proceed sequentially as long as there is any possibility that we might wish to stop after receiving and analyzing one of the clumps of information.

both clumps, which involves losing the synergies. To illustrate this tradeoff, suppose that it costs 10 to collect each clump on its own and 18 to collect both together.<sup>60</sup> Clearly, if we almost surely would continue to collect the second clump once we have the first, it would be better to collect both simultaneously, to save 2. But if there is a significant likelihood that we will stop and decide after the first clump, we will save 8 (and this savings, in turn, may outweigh the information loss from not having the second clump of information when we make our decision).

Second, it is important to drive home a point about option value: this benefit from sequential information gathering can be obtained only if the later decision whether to collect the second-sequenced clump of information is not foreordained but rather depends on what we learn from the first-sequenced clump of information. A central feature of the analysis is that, although it is optimal in principle to specify a complete plan of action up front, an optimal plan is a contingent one. Whenever there is sequential gathering of information, the decision regarding any subsequent step depends on what we learned from the preceding steps, which in general differs from what we knew at the outset.<sup>61</sup>

Indeed, a significant part of the value of information at various stages can lie in how it informs the choice of what information, if any, to collect next.<sup>62</sup> For example, if the first clump collected yields definitive evidence, one way or the other, we will stop, whereas if we learn little, we will proceed. Moreover, what we learn from the first clump of information in a setting with many clumps — considered just below — may be very probative regarding which subsequent step makes the most sense. Note further that, any time it is foreordained that we will collect a particular subsequent clump of information after a prior one, it will

---

<sup>60</sup> Note that there can also be some sequential synergies, which are omitted in the example in the text. That is, what we do while collecting or what we learn from the first clump may facilitate collecting the second clump. So, in our example, it might be that, whichever is collected first will cost 10, as stated, but then the second one collected will cost only 9.

<sup>61</sup> Cf. Susan Griffin, Nicky J. Welton & Karl Claxton, *Exploring the Research Decision Space: The Expected Value of Information for Sequential Research Designs*, 30 MED. DECISION MAKING 155, 161 (2010) (“However, [decision analysis] does not determine the order of research beyond the first study. Once the results of the first study or set of studies are available, the previous [expected value of information] calculations become obsolete and a new set incorporating the additional information is required.”); *id.* (“However, the value of a sequential design derives from the ability to avoid the cost of research on the second parameter in sequence on the basis of additional information collected on the first parameter in sequence. This suggests that sequential designs may be most valuable when the costs of investigating the second parameter in the sequence are significant and the benefits of investigating the second parameter in the sequence are sensitive to realizations of the first parameter.”).

<sup>62</sup> See, e.g., Allen C. Miller, *The Value of Sequential Information*, 22 MGMT. SCI. 1 (1975). An implication is that it may be optimal to collect, next, a clump of information that we are certain would not itself change our decision on liability — which may not be a close enough call to be reversed by what we contemplate collecting next — because the entirety of the value of what we learn will be in guiding what further information, if any, to collect. This point reinforces the lesson that optimal information gathering is by nature flexible, adapting to what is learned along the way.

---

---

be optimal to combine those clumps whenever there is any synergy gain from doing so.

Third, suppose that we are comparing the two sequential paths: clump one followed by a decide/clump two choice versus clump two followed by a decide/clump one choice. Which clump should be collected first? This question may be answered by reference to a heuristic that may be thought of as the diagnosticity/cost ratio. The guiding principle is to collect the higher-ratio clump first (if we are going to pursue a sequential strategy, rather than one of deciding as currently informed or collecting both together).

To elaborate, suppose first that both clumps are equally likely to be informative, considered alone. Then we should collect the cheapest one first as long as there is any chance that we will make a decision on liability without bothering to collect the other clump. The reason is that the likelihood of doing so is the same for both, but the savings is greater if it is the costlier one that we sequence second.

Suppose next that both clumps are equally costly, considered alone. Then we should collect the most diagnostic (informative) one first, as long as there is any chance we will make a decision on liability rather than collecting the other clump. If we do decide after only collecting a single clump, we will be making (on an expected basis) a better decision if we have more rather than less information. Moreover, collecting the more informative clump first makes it more likely that the option value will be realized, which is to say that it will be unnecessary to collect the second clump.

Consider how these lessons about the choice of ordering feed back on the decision whether to proceed sequentially rather than simultaneously. When there is a large difference in the diagnosticity/cost ratios of the two clumps, a sequential strategy tends to be superior. To illustrate, take a simple case in which one clump is both much more informative and much less costly.<sup>63</sup> Then, when it is sequenced first, there is a significant likelihood that it will no longer be sensible to collect the second, saving much of the cost and losing little information. If instead we were to collect both simultaneously, this large option value would be forgone.

Observe that all of the analysis pertaining to the choice between sequential and simultaneous strategies answers a further question implicit in the initial set-up: When it was first stated that there was a clump of information that might be collected, just what did “clump” mean? On reflection, it should be clear that the answer is not somehow predetermined but rather should be generated by engaging in the sort of analysis

---

<sup>63</sup> A particular case of note is where the information learned from examining each clump is highly correlated. In that case, whichever comes second may be fairly uninformative, in which event it would tend to be optimal to sequence and to examine the cheapest clump first. As this example illustrates, the diagnosticity of a clump of information depends on the order in which it is collected (as well as on what we have learned in particular from previously collected information).

---

---

just presented. For any two (or more) clumps, we could always consider whether it is best to combine them into a single clump, which is (with respect to them) to proceed simultaneously rather than sequentially. For any single clump, we could consider whether it is best to divide it into two (or more) clumps, in order to proceed sequentially. Applying the above lessons, high synergy tends to favor clumping whereas high option value (which typically exists when there are large differences in the diagnosticity/cost ratios) favors separation.

To conclude the list of lessons, one previous point bears elaboration. In an overall optimal scheme, information gathering and decisionmaking regarding liability are interdependent tasks. As explained throughout, a choice to collect a further clump of information is a choice to defer the liability decision. Conversely, a choice not to collect any further information is a choice to make a decision on liability. Moreover, it is a choice to make a particular decision about liability. That is, when choosing whether to gather further information, we are comparing the value along a particular path (whichever way forward is best) to the value of making the highest-value decision as currently informed. The values of the liability and no liability decisions differ,<sup>64</sup> and the correct comparison regarding whether to collect more information is made by reference to the more valuable of those two current decisions regarding liability. Obviously, to know the value of the better current decision requires knowing what that liability decision would be if the additional information was not collected. In other words, the choice to abstain, when done optimally, is necessarily made with a particular liability decision in mind.<sup>65</sup>

*Many Clumps of Information.* — To close this section, consider briefly how to generalize the analysis of the optimal collection of information when there are two clumps to the broader setting in which there are many clumps of information. Having seen how the analysis of optimal information collection can be extended from the case with only one clump of information to that in which there are two, we can readily see how, in principle, the analysis can be extended to three clumps, then four, and so on. Reflecting on how much the complexity of the problem grows in moving from one clump to two, however, we can imagine how rapidly complexity increases as the number of possible clumps expands.

---

<sup>64</sup> Except in the case of ties, which are ignored.

<sup>65</sup> More nuanced versions of this lesson could be offered by considering separately the costs of processing the information we obtain (which here have been treated as part of information-gathering costs, *see supra* note 50), recognizing that some processing of the information is required to make the choice at each point as we proceed, but if the choice was to decide on liability now, we might engage in further processing of the information in hand to confirm the correct decision (which may sometimes lead us to reopen the question whether to collect further information).



This challenge is well recognized in the literature on decision analysis.<sup>66</sup> The commonly suggested (and practiced) method does not involve the decisionmaker attempting to determine the full solution — particularly when the number of clumps is large — but rather proceeding somewhat myopically. For example, at the initial stage, we might look at the diagnosticity/cost ratio of each clump standing alone, focus on those clumps with the highest ratios, and consider how some of them might optimally be assembled, leading to either an immediate liability decision or a choice to collect one clump of information (or a handful). If we did collect some information, we would at that point reassess. After seeing what we had learned, we would reconsider the diagnosticity/cost ratio of the remaining clumps, focus on the most promising, consider which of them might best be combined, and decide what to do next. Throughout the process, each of the foregoing lessons derived from the cases with only one or two clumps of information would serve as a guide. Note that the additional complexity reinforces some of the key takeaways, particularly the importance of being guided by what we learn along the way. When it is too complicated to assess all of the possible permutations in advance, there is even greater reason to assess the situation after each step in light of what has been learned. After all, when we never even did a full preliminary assessment, we are calculating rather than recalculating our route. Rigid adherence to any initial guesses as to what may have made sense at later steps would often be far from optimal.

### *B. Structured Information Protocols*

As mentioned, structured decision procedures are often advanced because their prescription for decisionmaking is thought to have attractive implications for information gathering. Specifically, when the decision at the first step is to assign no liability and stop — that is, when  $H \leq H^*$  — the cost of collecting information on  $B$  is saved.<sup>67</sup>

In light of section A's analysis, it would be surprising if any particular, prespecified sequencing of information gathering turned out to be optimal across a broad range of cases. The optimal decision at any point along the way is highly contingent on what we already know, what we

---

<sup>66</sup> It is common to discuss this problem by reference to the curse of dimensionality, a term coined by Richard Bellman. See RICHARD BELLMAN, DYNAMIC PROGRAMMING ix (1957); RICHARD BELLMAN, ADAPTIVE CONTROL PROCESSES 94–95 (1961).

<sup>67</sup> This point about potential cost savings does raise the question of why we would ever wish to decide a case at step 2 rather than skipping the step and moving straight to balancing, as discussed in section II.B. That section was also critical of the final decisions made at step 1, but at least we might think that the social cost of those errors was outweighed by information cost savings, whereas mistaken outcomes that result from decisions at step 2 do not raise even the possibility of such an offset. (A more complete view would separate decisionmaking effort from information-collection effort, but the discussion in section II.B makes clear that savings on this dimension at step 2 are substantially illusory; indeed, greater effort may be required.)

are likely to learn, and the cost of doing so. Once some information is collected, what is optimal to do next depends on what was just discovered. These basic lessons contradict the plausibility of any a priori sequencing, including one that is entirely case specific, customized for a particular matter at the outset of the proceeding. Even routine medical diagnostic protocols, refined from the treatment of millions of patients who present with similar symptoms, are typically contingent on what we learn from initial and subsequent tests, including ones as simple as measuring blood pressure or listening with a stethoscope.

We will now see that the mode of information gathering that seems to be prescribed by the basic three-step structured decision procedure not only conflicts with all the teachings of section A but also suffers from additional defects. The core purported benefit — that sometimes we can limit information gathering to all of the evidence bearing on  $H$ , omitting all of that bearing on  $B$ , when our step 1 decision finds that  $H \leq H^*$  — rests on a number of critical assumptions that often fail miserably.<sup>68</sup>

Using the language of section A, it is imagined first that evidence naturally falls into  $H$  clumps and  $B$  clumps. This assumption, in turn, has two distinct aspects: that information bears either on  $H$  or on  $B$ , but not both (no information overlap); and that bits of information naturally cluster, in terms of synergies, into groups that are relevant entirely to  $H$  or to  $B$ , and not both (clumping by issue rather than by source). The other key assumption is that, even if both of these conditions held, the

---

<sup>68</sup> For present purposes, the stylized three-step structured decision procedure outlined in section II.A is taken to have these implications regarding information gathering. A natural variant to consider is that, rather than collecting all of the information on  $H$  at step 1, we might collect only enough information for  $H$  to reach  $H^*$  — deferring any remaining information on  $H$  to step 3, if we get that far — or only enough information to be sufficiently confident that  $H \leq H^*$ . Under this sort of guideline, cost savings would tend to be greater, but in any event the criticisms presented below would apply to all of the action before reaching the final, balancing step. Also, it is unclear what it would mean, for example, for  $H$  to “reach”  $H^*$ . (For example, if the first tea leaf suggested that  $H$  was very large, would we deem the step 1 test to be satisfied and proceed? Perhaps more plausibly, we might take a Bayesian prior regarding the distribution of possible values of  $H$ , evaluated at the outset, and, after each clump of evidence on  $H$  was gathered and the Bayesian prior was updated, ask at that point whether the mean of that interim distribution was greater than  $H^*$ . If it was, we might deem step 1 satisfied and proceed to step 2, commencing information gathering on  $B$ . Examining such possibilities in the context of how U.S. civil litigation is actually conducted, *see infra* section IV.A, we would also wish to consider competing evidence regarding  $H$  and how this sort of approach could possibly be implemented amidst the discovery process as ordinarily conducted.) Taking the suggestion further, one might view the hurdles at step 1 and/or step 2 as production burdens, meaning that the proponent on the issue would have to proffer sufficient evidence from which one might conclude that the test was met, which could be taken to be fairly minimal. Under that interpretation, the structured decision procedure would not be a significant determinant of either decisionmaking or information gathering, although (particularly at step 1) it might help in screening out weak cases. *See infra* section IV.B; *see also infra* section IV.A (calling into question the extent to which step 2 can serve more than a minimal function with regard to  $B$ ).

optimal sequencing would put all of the  $H$  clumps first, followed immediately (and only then) by a step 1 decision on whether  $H > H^*$ , and all of the  $B$  clumps second (with no further decision until that process was completed). Although we can imagine particular cases in which various of these assumptions might hold, at least to a good approximation, we will now see that in a substantial range of legal settings (including ones in which such structured procedures are employed or advocated), this is not the case.

*Information Overlap.* — Consider the implication of the assumption that information bears either on  $H$  or on  $B$ , but not on both. This supposition is plausible in certain contexts. For example, in considering whether to launch a fully specified new product, a firm's assessment of the likely market for the product and of the costs of producing it may be largely separate, with little information bearing on both.<sup>69</sup>

By contrast, in many (although not all) legal decision contexts, a central question bearing on whether to assign liability involves characterization: Is the act before the tribunal of the harmful or the beneficial type? In those cases, the very definition of relevant evidence is that which bears *differentially* on the competing characterizations. In such settings, we cannot coherently define what it means for information to bear on  $H$  but not on  $B$ , or vice versa.<sup>70</sup> As mentioned in the introduction, the attempt to proceed in this fashion is somewhat like trying to use scissors, disjointed, one blade at a time. This point is obvious and most powerful when there are two mutually exclusive characterizations, for then it is impossible to raise (or lower) the probability of one without lowering (or raising) the probability of the other — indeed, by precisely the same amount. Even when the competing explanations are not mutually exclusive, much information will be relevant to both  $H$  and  $B$ .

Consider some examples.<sup>71</sup> In many legal settings, a party's intent is thought to be relevant: in and of itself, or because understanding

---

<sup>69</sup> This illustration contains some favorable simplifications. Quite often in new product design, there is a significant back and forth between the two considerations, with market assessments informing product-design decisions that have cost implications, and information on costs suggesting design modifications that require reassessment of the potential market.

<sup>70</sup> A formal statement of the problem is that the strength of classification evidence, say, for Bayesian updating, is captured by what is referred to as the likelihood ratio — the ratio of the probability that the evidence would have been generated by the harmful sort of act to the probability that the evidence would have been generated by the benign sort — and it is impossible to state a ratio without regard to its denominator. See Kaplow, *Likelihood Ratio Tests*, *supra* note 6, at 5–13.

<sup>71</sup> An important concrete setting that is common in antitrust analysis fits these examples and the broader point under examination: the very definition of a harmful (anticompetitive) practice is one that hurts rivals *other than* by “competition on the merits,” the essence of a beneficial (procompetitive) practice. See, e.g., 3B PHILLIP E. AREEDA & HERBERT HOVENKAMP, *ANTITRUST LAW* 423 (4th ed. 2015).

intent illuminates the character of an act, including its likely effects.<sup>72</sup> In antitrust, if we know that a firm was trying to raise its rival's costs (boosting our estimate of  $H$ ), we are less likely to believe that its actions generate procompetitive benefits ( $B$ ), even if these are not mutually exclusive possibilities.<sup>73</sup> If an employer convinces us that it selected or refined an employment test to address a recognized shortfall of particular skills in its workforce ( $B$ ), we are less inclined to believe that the test is discriminatory ( $H$ ), although again both are possible. The cross-issue relevance may be even more apparent if we consider evidence negating each point: if we become convinced that a firm was not trying to raise its rival's costs, we are more likely to believe that its actions were procompetitive, and if we become convinced that no business justification underlies the introduction of a new test, we are more likely to believe that the test is discriminatory.

To take another situation, some information in legal disputes consists of a firm's internal documents. In such cases, there may be significant disagreements about interpretation. Often, a given statement will be ambiguous. The very definition of ambiguity is that there are multiple possible interpretations. One interpretation will boost our estimate of  $H$ , the other of  $B$ . In attempting to determine the correct interpretation, disputing parties will offer evidence of the context, with one side's profers supporting  $H$  and the other side's supporting  $B$ . It is incoherent to contemplate resolving an ambiguity by reference to only one possible meaning.

Of course, not all evidence that bears on  $H$ , one way or the other, necessarily bears on  $B$ .<sup>74</sup> But much will. We could, under a structured decision procedure, nevertheless attempt to stick to the three-step decision rule and follow the implied information-gathering protocol as best we can: First, collect all information that bears on  $H$ , even if much of it also bears on  $B$ . Next, make the step 1 decision, setting to the side any of the information's implications for  $B$ . But we already saw that much

---

<sup>72</sup> See, e.g., *Chi. Bd. of Trade v. United States*, 246 U.S. 231, 238 (1918) ("The history of the restraint, the evil believed to exist, the reason for adopting the particular remedy, the purpose or end sought to be attained, are all relevant facts. This is not because a good intention will save an otherwise objectionable regulation or the reverse; but because knowledge of intent may help the court to interpret facts and to predict consequences."); Richard H. Fallon, Jr., *Constitutionally Forbidden Legislative Intent*, 130 HARV. L. REV. 523 (2016).

<sup>73</sup> Consider another example: if a defendant engaged in an action that costs it some amount, showing that the profit potential of its procompetitive explanation is less than that cost rules out the procompetitive explanation standing alone and thereby rules in the anticompetitive explanation. This is so even if the information pertained only to the magnitude of the purported procompetitive benefit.

<sup>74</sup> Notably, much evidence will bear on the magnitude of  $H$  or the magnitude of  $B$ , given a particular characterization. These magnitudes will be relevant in determining whether  $H > B$  or in assessing, under the structured decision procedure, whether  $H > H^*$  and then whether  $B > B^*$ . However, as mentioned in the preceding footnote, even evidence, say, that bears directly only on the magnitude of  $B$  may thereby illuminate  $H$  in some settings.

---

---

of the jointly relevant information is relevant precisely because it bears on  $H$  compared to  $B$ , so this enterprise is dubious.

Moreover, even if the approach were coherent, a moment's reflection suggests that it is unwise. Section II.B already explained how step 1 of a structured decision procedure can generate the wrong outcome, notably when  $H \leq H^*$  even though it would turn out that  $H > B$ . In light of the foregoing, we can see that, when making a decision about  $H$  at step 1, we will often have in hand — even when attempting to follow the structured information protocol — significant information about  $B$ . Therefore, the structured decision procedure may direct us to make an inferior decision in many cases in which we are aware, or have a significant indication, that we are doing so.

Finally, think of the cost savings that the structured information protocol is supposed to generate. The suggestion that we do not need to incur the expense of gathering information on  $B$  in those cases in which we stop and assign no liability at step 1 is partly mistaken. As just explained, we will have already collected some, and perhaps much, of the information on  $B$ , whether we wanted to or not. Ignoring information in hand and thereby making an erroneous decision is hardly a sound way to economize on information-gathering costs.

*Clumping Information by Issue Versus by Source.* — Next, consider the implicit assumption that bits of information naturally cluster (in terms of synergies) into groups that are entirely relevant to  $H$  or to  $B$ , but not both. We know that this is not so when the bits of information, even at a fairly elemental level, individually pertain to both  $H$  and  $B$ , as just discussed. But what about when they do not?

Suppose now that we are considering information that bears only on  $H$  or only on  $B$ . Even in such cases, it will often be true that valuable synergies would be forgone if we first collected just the information on  $H$ , and then (if step 1 is passed) just that on  $B$ . The central reason is that much (again, not all) evidence naturally clusters by source, not by issue. Hence, section A's prescription about when clumps should be combined and when they should be divided is violated under the structured protocol.<sup>75</sup>

To illustrate, consider again the review of documents, understood broadly to include emails and any other form in which information may have been recorded. Doing comprehensive document identification, transfer to the other party or to an agency or a tribunal, and analysis, solely for purposes of assessing  $H$ , and then, later, undertaking a second full pass regarding  $B$ , would often be highly inefficient. Similarly, when

---

<sup>75</sup> This point illustrates how it is a mistake, when contemplating how actual litigation might be restructured, to limit attention to the possibility of bifurcation by issue rather than by source. See also *infra* section IV.A (addressing how the conduct of U.S. civil litigation deviates from section A's prescriptions on optimal information collection).

interviewing or deposing witnesses, it would often be more expeditious to undertake the process only once for each individual, covering all the issues. This point sometimes applies to the use of expert witnesses as well.<sup>76</sup>

We can see that the potential synergy loss from attempting to cluster evidence into  $H$  clumps and  $B$  clumps (if this were possible), when such clumps contain many of the same sources, can be large. If most cases would indeed terminate at step 1, using this technique would still economize on effort, on average, but much less so than we might have thought once it is recognized that the second wave, by source, may have been much cheaper had it been combined with the first. And for cases that do not terminate at step 1, costs will be substantially higher due to sequencing.

We could imagine instead combining the collection of information, at least as to those clusters where synergies are particularly high, in an attempt to reduce the potential for later wasteful duplication. In that event, when it comes time to ask whether  $H > H^*$ , we again would often be in a situation in which substantial information on  $B$  was in hand. Then, at little cost, we could significantly improve the quality of final decisions by considering directly whether it seems likely that  $H > B$ , based on the information then available, rather than adhering to the structured decision procedure, which would assign no liability if  $H \leq H^*$  even when we know that  $H > B$  is probably true. Moreover, once again the potential savings from stopping at step 1 are less than meets the eye because much information on  $B$  has already been collected.

*Ordering and Decision Points.* — Now consider how structured information protocols function with regard to the optimal sequencing of information collection (which, recall, is intimately intertwined with the timing of decisions on liability), the focus of much of the analysis in section A. Here, we will see that essentially every lesson is sharply violated. To consider the matter further, suppose that both of the aforementioned problems are nonexistent. That is, assume (favorably, for the structured protocol) that all information pertains only to  $H$  or only to  $B$  and, moreover, that when we form plausible clumps in light of cost synergies, every clump pertains only to  $H$  or only to  $B$ .

For concreteness, imagine that we have ten clumps of information that bear (only) on  $H$  and ten that bear (only) on  $B$ .<sup>77</sup> The structured protocol tells us to begin by collecting all ten clumps on  $H$ ; then to ask whether  $H > H^*$ ; and, if it is, to collect all ten clumps on  $B$  (to feed into

---

<sup>76</sup> Consider, for example, the question of what variables to include in a regression equation for purposes of assessing step 1 in Title VII disparate impact cases. See Kaplow, *supra* note 1.

<sup>77</sup> A clump might, for example, be a witness who (contrary to the above discussion) can illuminate only one or the other issue, or a particular set of documents.

the step 2 decision and, if that test passes, to undertake step 3's balancing test).

The shortcomings compared to optimal information collection are many and significant. First, we know that the optimal decision can come at any point, not just after clump 10 or clump 20. Second, the optimal decision, when made, can be in either direction, not keeping an assignment of liability off the table until all possible information is in.<sup>78</sup>

Third, regarding the ordering of the 20 clumps, consider whether we should even start with a clump on *H* rather than one on *B*. Suppose, for example, that each of the 20 clumps was just as likely as any other to have any particular diagnosticity/cost ratio. How probable is it that the best clump to collect first would be one of those on *H*? Obviously, 50%. That is, under these assumptions, half the time it is optimal to begin with a *B* clump rather than an *H* clump.

Fourth, consider the extreme front-loading with respect to *H*. How probable is it that not merely one but all ten of the clumps that bear on *H* will be those with the highest diagnosticity/cost ratios? The answer (assuming independence) is under 1 in 50,000.<sup>79</sup> Needless to say, when there are more than a few clumps of information, it is far-fetched to suppose that ordering all of the *H* clumps first might even approximately make sense.<sup>80</sup>

Note that this point holds even if we greatly relax the assumptions of this example in a manner that is favorable to the contemplated sequencing. We might imagine that, in some definable class of cases, *H* clumps often have higher diagnosticity/cost ratios than *B* clumps do. Even so, we would have to imagine that this was true for all of the *H*

---

<sup>78</sup> Although these two points are basic and powerful, they may seem strange in a legal setting, particularly with regard to early truncation followed by a decision to impose liability. There are, however, important rules that operate in this fashion, such as the rule of per se illegality for price fixing in antitrust, *see, e.g.*, *United States v. Socony-Vacuum Oil Co.*, 310 U.S. 150, 223 (1940), and, relatedly, the early truncation of some rule of reason analyses in a manner that imposes liability, *see, e.g.*, *FTC v. Ind. Fed'n of Dentists*, 476 U.S. 447, 457-64 (1986); *NCAA v. Bd. of Regents of Univ. of Okla.*, 468 U.S. 85 (1984). Similarly, countless medical treatment decisions — overwhelmingly, decisions not to treat or proceed further — are routinely made even though it is conceivable that exhaustive, costly, and possibly painful further testing would reveal an ailment that, ex ante, had a minuscule probability. In the other direction, sometimes highly consequential treatments are initiated despite significant uncertainty, and not only because time may be of the essence or no further testing is possible. Even when lives are at stake, decisionmaking systems do not and should not operate otherwise.

<sup>79</sup> From combinatorics, the value is given by  $(10!)^2/20! \approx 0.000018$ .

<sup>80</sup> A related point concerns the fact that the structured decision procedure not only puts all of the *H* clumps first but also (recalling the first and second objections, just above) places a single interim decision point after *all* of the *H* clumps have been gathered and before *any* of the *B* clumps have been gathered. Even if we believed that all of the *H* clumps had higher diagnosticity/cost ratios than any of the *B* clumps, we would have to believe that the first, substantial drop in the magnitude of this ratio occurred right at the divide between the lowest-ratio *H* clump and the highest-ratio *B* clump.

clumps and *B* clumps. A particular reason to doubt that this would be so with much generality is the phenomenon of diminishing returns. It makes sense to collect first those clumps that seem most promising, and as we investigate *H* further and further, remaining information will tend to become ever less promising. Well before the last *H*-type stone is turned over, we typically expect that there would be one or a few *B* clumps that have more favorable diagnosticity/cost ratios than those remaining in the *H* heap.<sup>81</sup> We usually would not want to complete an exhaustive investigation of *H* before having even a peek at *B*.

Fifth, the foregoing way of putting the question is too simple, and in a manner that is too favorable to the structured protocol. One of the central lessons of optimal information collection is that the ordering should be contingent. Questions like “what clump of information should come second?” can only be answered in light of what was learned from the first clump, and so forth. As explained in section A, a large part of the potential gain from sequential information collection and decisionmaking lies in taking advantage of what is learned along the way. Even in simpler cases, each decision after the first depends on what has been learned up until that point. Indeed, if it is ever the case that the information collected cannot affect the subsequent decision, then it was a mistake to incur the expense of collecting that information.<sup>82</sup>

Sixth, even granting our extreme, generous assumptions about separate, natural clumping that gives us pure *H* clumps and pure *B* clumps, it is possible that there would be additional synergy gains by combining some of those clumps, which may in turn involve combining an *H* clump and a *B* clump. Even if two clumps should not be combined at the outset, in choosing what to gather first, it may become optimal to combine them at a later point in light of what we then have learned (because the option value of separation has fallen).

*Synthesis.* — These six points on ordering and decisionmaking indicate that the core prescriptions of the structured information protocol deviate radically from the basic teachings of decision analysis.<sup>83</sup> Matters

---

<sup>81</sup> The exception would be where *B* was essentially inscrutable. This possibility is explored in the sequel article’s treatment of merger review, where it is explained how optimally to proceed in that setting and whether the factual predicate is likely to be true in most cases. See Kaplow, *supra* note 1.

<sup>82</sup> It may be that the information would prove valuable later. In that event, it generally should have been collected later (possibly in combination with other clumps of information, to achieve synergy gains).

<sup>83</sup> Considering interactions among these points gives us additional pause, particularly in light of the lesson in section A that optimal information gathering and optimal final decisions are inextricably intertwined. Notably, when contemplating the thresholds  $H^*$  and  $B^*$ , they are in principle taken as given, which is the very source of the distinction from balancing as a final decision rule. But if we, say, learn from some high diagnosticity/cost ratio clump of information that *B* is probably quite high, we would want a higher threshold for *H* (which may entail stopping, to assign no liability, if based on existing information *H* seems quite unlikely to be as high as *B*), whereas if we learn



---

---

are worse when we add back into the mix the fact that information often bears simultaneously on both *H* and *B* and that synergy-based clusters are often by source rather than by issue — that is, removing the two favorable and often unrealistic assumptions employed in making those six points. Structured decision procedures, in addition to their tendency to produce poor decisions with a given set of information, the subject of section II.B, are also extremely poor guides regarding how best to gather information in a class of cases or in a particular dispute.

If we examine other consequential decision contexts that have similar characteristics, this strong conclusion seems obvious. Consider the problem of differential medical diagnosis and treatment.<sup>84</sup> Specifically, take the simple case of a patient who presents with some initial symptoms that give rise to only two possible diagnoses, each implying a different course of treatment. If we took the patient's blood pressure or examined a blood test, would we consider the implications for one diagnosis but not the other? Would that even be meaningful in most cases? If some blood counts particularly illuminate one diagnosis and other counts illuminate another, would we first take a small sample of blood to have only the first set measured, await the results, see if that diagnosis might make sense, and only then proceed to perform another blood test, now measuring the other relevant set of variables? How about a biopsy that requires surgery? Would we perform it twice if two different things needed to be sampled? Would we undertake every possibly relevant assessment — from asking a patient to breathe deeply, to blood tests, scans, and biopsies, to getting a second opinion — that pertained to one diagnosis, making no decision along the way, and then, when all those results were in, make a decision to go one way with treatment or, in the alternative, to undertake the entire other course of assessments, beginning perhaps with another simple blood test, to consider the other diagnosis?

A moment's reflection reveals the absurdity of this sort of structured decision procedure. Of course, doctors do benefit from certain structured protocols that encapsulate the wisdom of research and experience, but these protocols look nothing like the structured information protocol we have been considering. Instead, they dictate a course of diagnosis and treatment that looks like the optimal information collection process described in section A. Even if some test always comes first, it does not

---

from that clump that *B* is probably very low, we would want a lower threshold for *H* (which may entail stopping, to assign liability).

<sup>84</sup> Or consider our prior discussion of the plausibly much more separable problem of marketing and cost analysis in a firm's product launch decision (that is, abstracting from the important complication of product design). How often would it make sense for a company, say, to do its entire analysis on the cost side (research, development, prototypes, and so forth) before undertaking any analysis of whether there is a market for the product? Or to undertake massive marketing studies before expending any effort to determine whether it is feasible to make the product and even roughly how much it might cost to do so?

---

---

always pertain to the course of treatment associated with action rather than inaction. A first or subsequent step may pertain to one, the other, or often both possibilities. And, most importantly, what information to collect next — or, indeed, whether to collect any more information rather than making a treatment decision (including perhaps sending the patient home) — depends on what was learned in the previous steps. Note also that the appropriate protocol tends to be customized, depending on the initial symptoms that the patient presents.

The point of this illustration is not that the particulars of optimal information gathering in certain medical contexts closely resemble the best procedures for various legal disputes. Rather, this sketch makes vivid how information is gathered and decisions are made in a highly consequential context in which much thought has been devoted to the challenge of collecting information and making decisions under initial, and sometimes not fully resolvable, uncertainty. Similar thoughtfulness should guide information gathering and decisionmaking in the legal system.<sup>85</sup>

#### IV. ADDITIONAL CONSIDERATIONS

Parts II and III systematically compare structured decision procedures to unconstrained balancing with respect to the final decisions they generate and the information collection protocols they recommend. This Part extends the discussion in a number of directions. Section A examines the actual conduct of legal proceedings in federal courts, showing that it conforms to neither decision method's information-gathering protocols — although agencies probably tend to collect information in a manner that resembles the recommendations under balancing. Section B addresses whether step 1 of a structured decision rule, perhaps employing a rather low threshold  $H^*$ , may have some appeal as a means of screening out weak cases. Section C steps back from the analytics that have been the focus of this Article to reflect on concerns that many have about balancing, which also may help to explain the allure of structured decision procedures.

---

<sup>85</sup> In the medical context, it is understood that many existing protocols are deficient, and, perhaps more problematically, some of the most valuable protocols are often not followed in practice. Moreover, it is sometimes thought best to guide or even constrain discretion in treating individual patients because the benefits of categorical rules that are usually correct may exceed their costs. *See also infra* note 144 (commenting on such benefits of rules in the legal setting). Of course, those protocols vary greatly by the initial symptoms a given patient presents and, as already suggested, have many of the features described in section A concerning the ordering of tests by the diagnosticity/cost ratio, adjusting the path in light of what initial tests reveal, and frequently terminating information collection to initiate treatment or to conclude that no treatment is required. And, as in legal settings, sometimes the information is fairly definitive, but often important decisions are made under great uncertainty.

*A. Conduct of Legal Proceedings*

This section juxtaposes Part III's discussion of information gathering against the manner in which legal proceedings are conducted. Two principal lessons emerge: First, although agencies may well collect information in a (roughly) optimal manner, the conduct of U.S. civil litigation does not ordinarily follow, even approximately, either the prescriptions of optimal information collection or structured information protocols (where they exist). Second, Part III's analysis can serve as a guide to reform and also to the exercise of discretion available to judges, arbitrators, and others who manage dispute resolution.

Much investigation and decisionmaking is done within specialized agencies. The governing framework often leaves much latitude regarding how this may be undertaken, and many internal regulations and standard operating procedures are chosen by the agencies themselves. Accordingly, they are free, within broad limits, to gather information in whatever order and to whatever extent they wish. This is particularly true when their ultimate decision is not to proceed, which often amounts in practice to assigning no liability. For example, a decision not to challenge a merger constitutes *de facto* permission.

Moreover, we hope and suspect that agencies roughly follow many of the prescriptions developed in section III.A regarding optimal information collection. Indeed, they may be inclined to do so even when a structured decision rule applies. However, in cases in which they seek to assign liability, they typically need to craft a justificatory opinion or go to court for an enforcement order, in which event they would have to rearrange the information in a fashion that fits the contours of the three-step regimen. Anticipation of this need may generate feedback effects because the structured decision procedure may call for incorrect decisions whereas optimal information collection is guided by a background test of unconstrained balancing. For example, an agency may be able to establish that  $H > B$ , but if it also needs to show that  $H > H^*$ , this may call for greater effort in gathering information on  $H$  or for termination short of a challenge if it emerges that the step 1 test fails even though  $H > B$ .<sup>86</sup>

It is also useful to recall the complexity of optimal information collection when there are many possible clumps of information, as there often are. Here, specialized agencies have a comparative advantage in that they hire experts and develop experience that substantially assists

---

<sup>86</sup> Note further that an agency can have a winning case under a structured decision procedure even when  $H \leq B$ , as elaborated in section II.B. An agency in that situation may choose not to bring such a case, in which event its behavior would deviate even less from the prescriptions of optimal information collection.

in information collection (not just decisionmaking). Indeed, the daunting nature of this task is relevant to deciding whether to establish such agencies in the first place.

Once a government agency or a private plaintiff finds itself in a U.S. (federal<sup>87</sup>) court, however, the game changes significantly.<sup>88</sup> If a complaint's adequacy is challenged in a motion to dismiss, the only question before the court is whether the challenger has stated a plausible claim.<sup>89</sup> Under a pure balancing test, the plaintiff must allege that  $H > B$ . (For example, in a negligence case, it must be alleged that the expected harm that would have been avoided by the omitted precaution exceeds the cost.<sup>90</sup>) By contrast, under the structured decision procedure the plaintiff must allege instead that  $H > H^*$ .<sup>91</sup> It is unclear how much this difference matters in practice or, when it does, which hurdle would be easier to overcome (which depends on, among other things, the magnitude of  $H^*$ , a point revisited in section B).

When a motion to dismiss is denied or none is filed, the case proceeds to discovery. Ordinarily, the scope of discovery covers all issues and all types of evidence, subject to limits regarding burdensomeness, or what is now called "proportional to the needs of the case."<sup>92</sup> The key point is

---

<sup>87</sup> The present investigation is guided entirely by consideration of the Federal Rules of Civil Procedure and judicial interpretations thereof. Of course, U.S. state courts often, but not always, function in roughly similar ways. And specialized federal courts have their own rules.

<sup>88</sup> Class certification is set to the side here. Interestingly, this constitutes a realm in which there seems to be more deviation from standard procedures, insofar as there is some bifurcation between consideration of facts bearing on class certification and the merits (despite often-present subject matter overlap).

<sup>89</sup> See *Bell Atl. Corp. v. Twombly*, 550 U.S. 544 (2007); *Ashcroft v. Iqbal*, 556 U.S. 662, 675–87 (2009).

<sup>90</sup> It is possible, under a balancing test, that a plaintiff would only be required to allege that  $H > 0$ . That might make sense in settings in which it seems plausible or is common that  $B = 0$ , although presumably an allegation of simple negligence does require at least advancing some plausible basis for believing that the defendant failed to take a cost-justified precaution.

<sup>91</sup> It is generally understood that, under structured decision procedures, it is the defendant who has the burden of proving, or at least producing evidence in support of, the proposition that  $B > B^*$ , so the plaintiff need not make an allegation regarding step 2 (or, a fortiori, step 3), although variations could require otherwise.

<sup>92</sup> FED. R. CIV. P. 26(b)(1) ("Parties may obtain discovery regarding any nonprivileged matter that is relevant to any party's claim or defense and proportional to the needs of the case, considering the importance of the issues at stake in the action, the amount in controversy, the parties' relative access to relevant information, the parties' resources, the importance of the discovery in resolving the issues, and whether the burden or expense of the proposed discovery outweighs its likely benefit."). The conceptual and operational meanings and significance of this limitation are not substantially clarified by the Advisory Committee's notes and background materials or by judicial opinions (the quantity of which is rather limited relative to the amount of discovery that transpires since few discovery disputes generate written opinions). Observe that, for anything to be literally "proportional," there needs to exist some thing that, in turn, has to be proportional to some other thing, and there also has to be a factor of proportionality. We suspect that the intuitive sense of the two requisite "things" would be notions like diagnosticity and cost, so that in a loose sense it may be that a judge's thinking about the matter is guided by a sort of diagnosticity/cost ratio. Regarding

that, unless a judge or magistrate chooses to engage in substantial case management, the ordinary conduct of discovery does not involve sequencing. Specifically, it does not adhere to the principles of optimal information collection developed in section III.A, which would require interim assessments and stopping decisions that are associated with particular decisions regarding liability. Nor does discovery follow the dictates of structured information protocols, which would call for discovery only on  $H$ , followed by a determinative resolution of whether  $H > H^*$ , which would have to be answered affirmatively (requiring complete factfinding) before proceeding to discovery pertaining to  $B$ . Thus, when some scholars advance structured decision procedures because they may save the costs of collecting information on  $B$  in the course of civil litigation, it is mysterious what they have in mind.

the critical ratio (proportion) — that is, the amount of cost that seems warranted in light of the predicted diagnosticity — we suspect that it would derive from a judge's sense of the importance of the issue being illuminated, which would often be related to the stakes but sometimes to other factors as well. In this author's reading, there is no indication of an awareness of the more explicit framework presented in section III.A regarding the value of information. Nor is there an awareness of the significant further subtleties developed in some of the academic literature regarding the value of accuracy in adjudication. *See, e.g.,* Kaplow, *Value of Accuracy*, *supra* note 46; Kaplow, *Information and Adjudication*, *supra* note 46, at 1332–44; *see also infra* note 104 (regarding limitations on judges' capacity to manage discovery and potential ways to address them). Indeed, it is not even clear that the concept of proportionality is doing any work, rather than being taken as a synonym for “reasonableness,” which is to say, an invitation to use common sense and experience without offering any further, particular guidance (beyond the list of considerations contained in the rule, which themselves seem to be an invitation to consider any seemingly relevant facts and circumstances in whatever way seems appropriate). In any event, the proportionality prescription seems generally to be taken as a cost-benefit test of some sort, *see* DUKE LAW CTR. FOR JUDICIAL STUDIES, REVISED GUIDELINES AND PRACTICES FOR IMPLEMENTING THE 2015 DISCOVERY AMENDMENTS TO ACHIEVE PROPORTIONALITY 42–44 nn.20–21 (Mar. 20, 2017), [https://judicialstudies.duke.edu/sites/default/files/centers/judicialstudies/civil\\_rules\\_project-mar.pdf](https://judicialstudies.duke.edu/sites/default/files/centers/judicialstudies/civil_rules_project-mar.pdf) [<https://perma.cc/KK6H-3FEP>], and, as suggested above, we would suppose that the benefit involves a notion of diagnosticity, perhaps weighted by a sense of the importance of what might be illuminated.

It is also noteworthy that most of the attention over the years to the regulation of discovery has been directed not at how best to craft a sequential process but rather at drawing the line as to what should and should not be permitted, as if the decision regarding what discovery to allow would be made entirely at the outset. But there are exceptions, particularly in material that has appeared after the 2015 amendments. *See id.* at 19 (“Practice 5: In many cases, the parties will start discovery by seeking information relevant to the most important issues in a case, available from the most easily accessible sources. In a case in which the parties have not done so, or in which discovery is likely to be voluminous or complex, or in which there is likely to be significant disagreement about relevance or proportionality, the parties and the judge should consider and discuss starting discovery with the subjects and sources that are most clearly proportional to the needs of the case. The parties and the judge can use the results of that discovery to guide decisions about further discovery.” (footnotes omitted)); *id.* at 24 (“Practice 9: When proposed discovery would not or might not be proportional if allowed in its entirety, the judge should consider whether it would be appropriate to grant the request in part and defer deciding the remaining issues.”). An impression is that the addition of the term “proportional” in the 2015 amendments to the Federal Rules of Civil Procedure and the related jiggling of Rules 26(b)(1) and 26(b)(2) were primarily meant to give judges a further nudge (beyond the attempts in prior revisions in 1983 and 1993) to exercise greater control over discovery, rather than to dictate how that might best be accomplished.

After discovery, a party may move for summary judgment. Under either balancing or a structured decision procedure, this would ordinarily involve a motion by the defendant claiming in essence that there is a negligible evidentiary basis (sufficient to create a “genuine dispute”) for believing that  $H$  is nontrivial.<sup>93</sup> Note, as just explained, that at this point discovery would ordinarily have been completed on all issues, so even if the motion is granted, all information gathering pertaining to  $B$  will have occurred even under a structured decision rule.

There are other summary judgment possibilities as well, although ones that in most settings are less frequent. With balancing, a defendant could argue that there is no genuine dispute that  $B$  is fairly large and, moreover, there is no genuine dispute that  $H$ , even if nontrivial, is smaller than  $B$ . With the structured decision procedure, a defendant may argue that even though  $H$  is certainly positive, it is undisputedly not large enough to exceed  $H^*$ ; and it may also argue that, even if  $H > H^*$ , it is true both that  $B > B^*$  and also that  $H \leq B$ , both beyond genuine dispute. In addition, as noted in the margin, there are circumstances in which plaintiffs might move for summary judgment, which is infrequent in the areas of law subject to structured decision procedures.<sup>94</sup>

<sup>93</sup> FED. R. CIV. P. 56(a); see *Celotex Corp. v. Catrett*, 477 U.S. 317 (1986); *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242 (1986); *Matsushita Elec. Indus. Co. v. Zenith Radio Corp.*, 475 U.S. 574 (1986). An interesting and underexplored feature of summary judgment concerns the decision standard when there will be a bench trial. In essence, a judge is told that it is impermissible to predict how a factfinder would weigh the evidence, limiting the analysis to what a reasonable factfinder might conclude. See John T. McNaughton, *Burden of Production of Evidence: A Function of a Burden of Persuasion*, 68 HARV. L. REV. 1382 (1955) (offering the classic elaboration of this idea by offering an amendment to Wigmore’s famous diagram). However, when the judge him- or herself is the factfinder, this restriction is strange. Of course, the presentation at a trial — with pretrial briefs, live testimony, cross-examination, and so forth — may facilitate thoughtful assessment of the facts, but this is not the reason ordinarily given. And it is hardly clear that bench trials are organized in the fashion that is most conducive to a judge sorting through the evidence. Interestingly, in many continental systems, decisions are based more on the paper record, and in yet other jurisdictions, including some common law jurisdictions, it is not uncommon, for example, to have the two sides’ experts testify together, and in other ways to deviate from the standard script employed in U.S. civil litigation. See, e.g., CIVIL JUSTICE COUNCIL, CONCURRENT EXPERT EVIDENCE AND “HOT-TUBBING” IN ENGLISH LITIGATION SINCE THE “JACKSON REFORMS”: A LEGAL AND EMPIRICAL STUDY (2016), <https://www.judiciary.uk/wp-content/uploads/2011/03/cjc-civil-litigation-review-hot-tubbing-report-20160801.pdf> [<https://perma.cc/U4F8-DT2U>]; Steven Rares, *Using the “Hot Tub” — How Concurrent Expert Evidence Aids Understanding Issues*, 95 INTELL. PROP. F.: J. INTELL. & INDUS. PROP. SOC’Y AUSTL. & N.Z. 28 (2013).

<sup>94</sup> A plaintiff, under a structured decision procedure, could move for partial summary judgment regarding step 2, arguing that the defendant’s alleged  $B$  is indisputably less than  $B^*$ ; if successful, this would mean that the defendant could not advance  $B$  at trial, so that the plaintiff would win if it can convince the factfinder that  $H > H^*$ . It could, moreover, move for a full judgment if it further established as well that  $H$  indisputably exceeded  $H^*$  and (unless it can show that the defendant’s proffered  $B$  should be disregarded because it is indisputably below  $B^*$ ) that  $H$  indisputably exceeds  $B$ . And, under unconstrained balancing, establishing just the latter would suffice. (In some case law and commentary, there is discussion of the burden shifting to the defendant. See Kaplow, *supra*

If a case goes to trial, any decision under either balancing or a structured decision procedure will typically not be made until the end. (The exception involves a motion for judgment as a matter of law at the conclusion of the plaintiff's case, which relates to the core situation in which a defendant should prevail on a motion for summary judgment.<sup>95</sup>) In that event, even under a structured decision procedure, in a case in which there is no liability because step 1 fails (that is, the factfinder ultimately concludes that  $H \leq H^*$ ) there will not even be a savings in trial costs as long as there had been a genuine dispute about whether this was so. The only savings would be in the factfinder's final deliberation efforts, for a judge or a jury may enter a finding of no liability if it concludes that step 1 fails, without deciding steps 2 and 3.<sup>96</sup> As a practical matter, verdicts and (with bench trials) opinions often complete all of the steps, in part because of the possibility of an appeal.<sup>97</sup> Therefore, the belief that structured decision rules economize substantially on litigation costs seems to be a mirage. (In such a system, neither does balancing, but it is not suggested otherwise when the two approaches are compared.)

Another belief sometimes advanced in discussing structured decision procedures (and more broadly) is that, when the plaintiff establishes its case at trial, requiring that  $H > H^*$  under the structured decision procedure, the burden then shifts to the defendant regarding step 2.<sup>98</sup> It is

---

note 1. Yet it is hard to see, except after a trial, how this could happen unless, as just stated, the plaintiff can establish that there is no genuine dispute that  $H$  exceeds  $H^*$ . Merely establishing a genuine dispute to avoid summary judgment for the defendant hardly means that the plaintiff has met its burden of persuasion.)

<sup>95</sup> Under *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, the standard for summary judgment under FED. R. CIV. P. 56 is the same as that for judgment as a matter of law under FED. R. CIV. P. 50. *Anderson*, 477 U.S. at 249–50. For this reason, when a judge rules for the defendant at the close of the plaintiff's case, it would seem that a defendant's earlier motion for summary judgment (if made) should have been granted. Some judges may have refrained from ruling on such a prior motion or may have resolved doubts in favor of the plaintiff but subsequently conclude that the matter is no longer in doubt once the plaintiff's case has been fully heard. Obviously, substantial costs are incurred as a consequence of delaying such a decision.

<sup>96</sup> On reflection, it is bizarre that a structured decision procedure that results in a substantial range of incorrect decisions and, if taken seriously, would also badly distort the information-gathering process, would nevertheless be advanced because of cost savings when such are reaped, at best, in the sliver of cases that proceeds to the end of trial and, even then, constitute only a sliver of the total costs in such cases.

<sup>97</sup> Similarly, appellate courts sometimes cover all of the steps, or at least step 2, in any event.

<sup>98</sup> See, e.g., Andrew I. Gavil, *Burden of Proof in U.S. Antitrust Law*, in 1 ISSUES IN COMPETITION LAW AND POLICY 125, 126 (Wayne Dale Collins ed., 2008) ("Another way of posing the question is to ask, 'at what point should a presumption of unreasonableness arise, such that the burden of production should shift from the plaintiff to the defendant?' Similarly, how much and what kind of evidence is sufficient to shift the burden of production back to the plaintiff, who of course bears the ultimate burden of proof?"). Regarding the more surprising notion that the

unclear how this shift would be manifested; after all, the factfinder, whether a judge or a jury, does not at any point during the trial announce that the plaintiff has won (or even has the lead) on step 1. (The suggestion brings to mind a referee who, perhaps, initially points a flag toward the plaintiff, but when its persuasion burden is met, then points it toward the defendant. There is a referee, but no flag and no pointing.) Even if a defendant moves for judgment as a matter of law after the plaintiff's presentation and that motion is denied, this ruling means only that the defendant has not shown definitively that  $H \leq H^*$ ;<sup>99</sup> the plaintiff's evidence is merely deemed to be sufficient to create a genuine dispute on the matter, keeping open the possibility that a factfinder may nevertheless conclude that  $H \leq H^*$ .<sup>100</sup> It is only when a finding of  $H > H^*$  is actually made, at the end of the trial, that step 1 is ever established.

In all, we can see that structured decision procedures often call for different, incorrect outcomes, but it is not clear that they have much effect on information gathering or the process of decisionmaking (short of the final decision) in the course of civil litigation. Nor is information gathering through discovery, as ordinarily conducted, guided by the prescriptions of optimal information collection that derive from unconstrained balancing. This section now examines other ways that the difference between the two types of legal rules may matter and then briefly considers possible reforms.

It is natural to ask whether nonexpert factfinders, particularly lay juries, might benefit from the guidance offered by structured decision procedures. Although possible, a conjecture is that this is unlikely, both because such procedures, when followed systematically, lead to incorrect conclusions (the focus of section II.B) and because balancing — the weighing of pros and cons — is a decision rule that untutored factfinders are familiar with.

As a thought experiment, consider deploying the three-step structured decision procedure in an ordinary negligence case. The jury would be told first to consider whether the magnitude of the expected harm that the omitted precaution(s) would have avoided exceeds some threshold,  $H^*$  (which, as under actually advocated and employed structured

---

burden may shift prior to trial, perhaps at summary judgment, see note 94. In light of this paragraph's discussion and that in the cited footnote, it is hard to understand just where and how a burden shift can actually occur, unless (as discussed below) proceedings are bifurcated.

<sup>99</sup> Another oddity is that, at this point in the trial, the defense has not yet presented any of its case, including its evidence that  $H \leq H^*$ . Hence, even if a judge ruled at that point that, if the trial were over, the plaintiff would prevail on step 1, the defendant would still have no affirmative burden on  $B$  in the sense that it could instead attempt to rebut the plaintiff's showing on  $H$ .

<sup>100</sup> Contrast the depiction offered in Gavil, *supra* note 98, at 139 n.63 ("Under a litigation process approach, however, the court would not consider any evidence of a defense (and the defendant would not be required to proffer any defense) until after the plaintiff had already met a *burden of production* and *effectively shifted* its burden to defendants." (emphasis added)).



schemes, is usually described verbally). If not, it should not assign liability. At this point, we should ask ourselves how high the jury might imagine  $H^*$  to be, or what the jury might best be told it is. Is \$10,000 large enough? Or should it be at least \$1,000,000? Or perhaps \$10,000,000? Keep in mind that this inquiry is done without looking ahead to  $B$ , which here concerns the cost of the precaution (for example, should the defendant have made a minor tweak in its operations? or employed a highly advanced and extremely expensive new safety technology?). Then, if the step 1 test is passed, the jury would be asked to examine  $B$  (about which it had already heard evidence but had been told to ignore until now) and compare it to some  $B^*$ . This comparison raises similar questions. Finally, if step 2 is also passed, it would be instructed to perform the negligence rule's familiar balancing test.

This approach is not used in negligence cases, and it seems far-fetched to suggest that it would be an improvement even granting that, say, in medical malpractice and airplane crash cases, factfinding is extremely challenging. Although the determination of which mode of instruction is best presents an empirical question, it is hardly clear that using a structured decision procedure in antitrust (where rule of reason cases may well be decided by juries) would be useful.<sup>101</sup> Perhaps we might defend structured decision procedures more as heuristics,

---

<sup>101</sup> Interestingly, the ABA Antitrust Section's 2005 model jury instructions on the rule of reason employ a structured decision rule, but, in a note that is not part of the central instruction itself, suggest that the multistep sequence of tests is not actually part of the law; rather, the instruction is meant to aid the factfinder. See SECTION OF ANTITRUST LAW, AM. BAR ASS'N, MODEL JURY INSTRUCTIONS IN CIVIL ANTITRUST CASES, 2005 EDITION, at A-4 & n.1 (2005). (This note no longer appears in the subsequent edition where the analogous instruction appears. See *id.*, 2016 EDITION 3 (2016).) By contrast, in Title VII disparate *treatment* cases, a structured decision procedure is undoubtedly part of the law, as developed by the Supreme Court, but it is *not* included in jury instructions, in part because it is regarded to be too confusing. See, e.g., MANUAL OF MODEL CIVIL JURY INSTRUCTIONS FOR THE DISTRICT COURTS OF THE NINTH CIRCUIT, 2017 EDITION 214 (2018), <http://www3.ce9.uscourts.gov/jury-instructions/model-civil> [<https://perma.cc/DDD4-S4PR>]; Kenneth R. Davis, *The Stumbling Three-Step, Burden-Shifting Approach in Employment Discrimination Cases*, 61 BROOK. L. REV. 703, 706-07 (1995) ("The complicated, three-stage, burden-shifting scheme shifts the burden of incomprehensibility to the jury. Recognizing this problem, some courts do not even instruct the jury on the *McDonnell Douglas* approach, fearing that a *McDonnell Douglas* charge may sabotage the factfinding process by leading jurors 'to seize upon poorly understood legalisms.' The efficacy of a system is suspect when judges distrust the jury's ability to understand it." (footnotes omitted) (quoting *Loeb v. Textron*, 600 F.2d 1003, 1016 (1st Cir. 1979)); Barrett S. Moore, *Shifting the Burden: Genuine Disputes and Employment Discrimination Standards of Proof*, 35 U. ARK. LITTLE ROCK L. REV. 113, 123 (2012) ("Courts now universally accept that *McDonnell Douglas* is not a matter for the jury."). Another reason for not giving instructions about the structured decision rule is that, to the extent the steps involve mere production burdens, which have already been resolved by the judge, they would be inapposite if the case is being given to a jury.

checklists if you will,<sup>102</sup> but in that case they might best be presented in just that form.<sup>103</sup>

Finally, consider some implications of the analysis in section III.A on optimal information collection for the management of litigation under the existing Federal Rules of Civil Procedure (or more informal arbitration tribunals) or for suggesting avenues of reform. Central to an optimal process is active decisionmaking at the outset (choosing which clump(s) of information to collect first), followed by the decisionmaker's assessment of what is learned either to reach a decision on liability or to collect additional information — and, if the latter, what clump(s) of information to collect next — and so on until a final decision is made.

Occasionally, federal judges and magistrates do some sequencing of discovery.<sup>104</sup> Note, however, that if no decisions can be made along the

---

<sup>102</sup> Lists have an order, which might have some subtle effect on how information is processed, especially by nonexpert decisionmakers. Cf. Adam M. Samaha, *Starting with the Text — On Sequencing Effects in Statutory Interpretation and Beyond*, 8 J. LEGAL ANALYSIS 439 (2016) (focusing on whether it matters in statutory interpretation if a judge starts with the text).

<sup>103</sup> Another possible argument for structured decision procedures is that they are more comprehensible to primary actors who need to understand the law to guide their actions, but this too is dubious. Balancing costs and benefits is surely a more familiar form of reasoning for businesses, regulatory agencies, and many others. Moreover, if advised that a realm of action is governed by a structured decision procedure of the type examined here, the actor would need to know the magnitudes of  $H^*$  and  $B^*$ , but, as mentioned, these are usually stated qualitatively rather than quantitatively. They could be made quantitative, which would provide sharper guidance in cases in which  $H$  clearly falls above or below  $H^*$  but it is uncertain whether  $H > B$ , although guidance would be murkier when the latter is clearer than the former. The analysis in section II.B implies, however, that the guidance thereby provided by the structured decision rule would often lead actors to behave in socially undesirable ways. Note also that, in many settings, potential actors may be more knowledgeable about  $B$  than  $H$  (because the benefit inures to themselves whereas the harm is caused to others), which makes prediction challenging under both approaches.

<sup>104</sup> This author is unaware of empirical evidence on the frequency and methods by which this is done. From informal discussions, it seems that substantial management of the sort described here is infrequent. In part this may reflect that greater judicial resources would be required, such as are present in some continental systems that do proceed more sequentially. See PETER L. MURRAY & ROLF STÜRNER, GERMAN CIVIL JUSTICE 639 (2004); Rolf Stürner, *The Principles of Transnational Civil Procedure: An Introduction to Their Basic Conceptions*, 69 RABELS ZEITSCHRIFT FÜR AUSLÄNDISCHES UND INTERNATIONALES PRIVATRECHT 201, 227 (2005) (suggesting that the trend toward more intensive judicial case management in the United States has been limited by its relatively small number of judges). And it may reflect an appreciation of limitations on judges' abilities. For example, merger cases are highly complex along a number of dimensions and involve testimony presenting the use of technical, specialized methodologies, yet most federal district judges only rarely preside over a merger case. One way to relax the resource constraint and enhance the court's institutional competence in this realm would be to use a court-appointed expert or an expert magistrate to help manage the process. See, e.g., SECTION OF ANTITRUST LAW, AM. BAR ASS'N, PRESIDENTIAL TRANSITION REPORT: THE STATE OF ANTITRUST ENFORCEMENT 17–19 (2017) (proposing consideration of this reform in the antitrust context); see also Frank H. Easterbrook, *Discovery as Abuse*, 69 B.U. L. REV. 635, 644–45 (1989) (advocating greater judicial involvement in discovery); Richard A. Epstein, *Essay, Of Pleading and Discovery: Reflections on Twombly and Iqbal with Special Reference to Antitrust*, 2011 U. ILL. L. REV. 187, 206–07 (proposing staggered discovery); *supra* note 92 (discussing the principles that judges probably use when they do manage discovery, and their relationship to the principles of optimal information collection).

way, the benefits are significantly circumscribed. Put simply, it may not much matter whether clump one or clump two is collected first if both will be collected regardless. Implicitly, then, judges who manage sequential discovery are making decisions to cut off some paths of inquiry, contingent on what has been learned so far. If we imagine that there is, say, some diagnosticity/cost ratio that determines what is worth collecting, examining initially collected information generates revised estimates of the value of the as-yet uncollected information, which can lead to more or less being collected than if all possibilities had to be assessed and decided upon up front. Additionally, even though an involved judge is not allowed to make decisions regarding liability during the course of discovery (or even at summary judgment, except in limited circumstances), a judge could make implicit decisions, as per the dictates of optimal information collection, and use them to decide on the ordering of information collection and on when to stop, declaring that enough is enough.<sup>105</sup>

A substantial constraint of the current system is that it is not designed to facilitate such a process. Notably, making sensible decisions on what, if any, information should be collected next requires real assessment of the information that has been collected so far to determine (as explained in section III.A) whether the case falls in the middle range of  $H - B$  rather than at the high end (calling for an immediate decision of liability) or the low (negative) end (no liability). Likewise, determining the diagnosticity of as-yet uncollected information requires significant assessment of the information in hand.

Gains may also be possible through the bifurcation of proceedings, which is occasionally employed. Bifurcation, however, is usually done by issue.<sup>106</sup> The advantages of this method are more limited than meets

---

<sup>105</sup> Another source of leverage is that a judge can defer ruling on a motion to dismiss, allow limited discovery, and, if it reveals the plaintiff's case to be sufficiently weak, then grant the motion. See Kaplow, *Multistage Adjudication*, *supra* note 6, at 1283–88. The formal basis for such a ruling would be that the plaintiff's allegation is implausible. In principle, that is to be decided without discovery. But a judge who is not sure about the matter may wish to give the plaintiff the benefit of the doubt, but only to the extent of and in the domain of that doubt.

<sup>106</sup> *Cf. supra* note 88 (mentioning that bifurcation may be done in the course of class certification). Notwithstanding some of the criticism that follows, such bifurcation may well make sense when the split is between liability and damages, as long as the evidence neatly separates, which it may in many torts cases but often not in other contexts (the principal reason being that liability often depends on  $H$ , which might be closely related to the measure of damages). A noted lower court antitrust case, *Rebel Oil Co. v. Atlantic Richfield Co.*, 51 F.3d 1421 (9th Cir. 1995), involved a district court judge's use of bifurcation. "[T]he district court limited discovery solely to the issue of whether ARCO had sufficient market power to charge prices above competitive levels. The district court justified the limited discovery on the ground that, absent a showing of market power, Rebel could not demonstrate that it suffered 'antitrust injury.' There was no discovery on predatory pricing, intent and collusion." *Id.* at 1432 (internal citations omitted). The difficulty here is that the appropriate meaning and relevance of market power, properly understood, depend on the matters that the district court had deferred, as explained in Louis Kaplow, *Recoupment, Market Power, and Predatory Pricing*, 82 ANTITRUST L.J. 167 (2018). See generally Louis Kaplow, *On the Relevance of Market Power*, 130 HARV. L. REV. 1303 (2017).

the eye and, if not done carefully, it can lead to adverse outcomes. The basic reason is that, as section III.B emphasizes, clumping is often best done by source, not by issue. Moreover, often the same evidence bears on both  $H$  and  $B$ . Actually implementing the bifurcation seemingly prescribed by the structured decision procedure was revealed to be extremely problematic for many reasons. Of course, there may be some types of cases, or particular cases of a given type, where issue bifurcation makes sense, especially when compared to the default regime in federal courts of collecting “all” or nothing. Nevertheless, greater attempts to follow the actual prescriptions of optimal information collection seem worth considering.

Regarding any of these possibilities for greater management of litigation, it would be best in terms of optimal information collection and final decisions if they were conducted under the auspices of a balancing test. However, substantial gains may be possible under structured decision rules as well. As mentioned in connection with agency processes, which tend to be much less constrained, it makes sense to substantially follow the prescriptions of optimal information collection even when unconstrained balancing is not the final decision rule.

### B. Screening

This section considers how step 1 in the structured decision procedure — the requirement that  $H > H^*$  — might be employed to screen cases: that is, to assign no liability early on when cases are sufficiently weak.<sup>107</sup> To focus on this feature, consider a modification to our three-step structured decision rule that omits step 2. That is, suppose that if step 1 passes, we go straight to balancing.<sup>108</sup>

Most of the implications of such a scheme are apparent from the preceding analysis. Elimination of step 2 avoids errors that can arise when liability is imposed due to that step’s failure — when  $B \leq B^*$  even though  $B > H$ . On the other hand, the implied structured information protocol still front-loads all information pertaining to  $H$  (which was due entirely to the existence of step 1), with all of its attendant costs and

---

<sup>107</sup> Because Part III on information gathering fully considers the optimal sequencing of information and timing of decisionmaking, as well as the respects in which structured decision procedures fall short on these dimensions, an analysis of screening is implicit in that treatment. This section highlights key features that pertain particularly to screening, which is sometimes viewed as a distinct function of legal rules and procedures. The discussion here is confined to substantive legal rules and standard procedural mechanisms, whereas other instruments (such as nontrivial filing fees) may have superior features in inducing informed plaintiffs to self-select in making their filing decisions. See Louis Kaplow, *Optimal Design of Private Litigation*, 155 J. PUB. ECON. 64 (2017); Kaplow, *Information and Adjudication*, *supra* note 46, at 1345–50.

<sup>108</sup> Note that this formulation may be close to some actual and proposed three-step structured decision procedures examined in Kaplow, *supra* note 1, to the extent that step 2 constitutes only a production burden or  $B^*$  is not very high.

oddties. It might nevertheless be thought that such a two-step procedure, perhaps with  $H^*$  set only modestly above zero, would quickly eliminate many weak cases and, more importantly, discourage such cases from being brought in the first place.<sup>109</sup> In light of adjudication costs, such an approach has appeal.

Obviously, the toughness of such a step 1 screen is related directly to the magnitude of  $H^*$ . If it is high, many cases might be screened, but many mistakes will also be made. If  $H^*$  is fairly low, few may be screened, although perhaps enough to be worthwhile, especially if the filing of many meritless cases is deterred. However, even if  $H^*$  is low, as long as the corresponding structured information protocol needs to be followed, all of its costs will still be incurred. An alternative would be to require nontrivial proof that  $H$  likely exceeds  $H^*$  (a sort of production burden), reverting to unstructured balancing if that preliminary demonstration is made.

Note further that, under an approach with a low  $H^*$ , errors would be limited in both frequency and magnitude. They would be relatively infrequent because, when  $H \leq H^*$  and  $H^*$  is low,  $H$  will not be greater than  $B$  very often. Furthermore, even when we nevertheless have  $H > B$ , the magnitude of our error, which is  $H - B$ , will be low because we know that  $H$  cannot be very large.<sup>110</sup>

For agencies, it is not clear how helpful such a step 1 would be. The best way to proceed is that developed in section III.A on optimal information collection. As we saw there, the clumps of information that make the most sense to collect first may pertain to  $H$ , to  $B$ , or often to both. The third is often the best option because much information pertains to both  $H$  and  $B$  in any event (and it would be counterproductive to ignore what we learn about  $B$  when screening cases) and because

---

<sup>109</sup> The latter might be thought to be particularly important if plaintiffs rather than defendants bear most costs associated with demonstrations of  $H$ . However, Part III provides a number of reasons why information regarding  $H$  and information regarding  $B$  are often intertwined. Moreover, even when they are distinct, in some important contexts much of the information on  $H$  is in the hands of defendants. See, e.g., *infra* note 114 (noting Title VII disparate impact cases).

<sup>110</sup> The present discussion, however, like most of this Article, abstracts from the endogeneity of *ex ante* behavior and the flow of cases into the legal system, see *supra* notes 5–6, a point now receiving some consideration in the literature. The present point about limited social harm from setting  $H^*$  modestly above zero needs qualification when taking a broader view. In particular, it is possible that a defendant may act undesirably in a way that causes a small degree of harm to each of a large number of individuals, which actions might be deterred if a class action suit were permitted (not a problem for the present argument if  $H$  is understood to represent the aggregate harm) or, if that were not feasible or permitted, by the prospect of a large number of individual (small  $H$ ) suits. Note further that, even if the latter would involve substantial litigation costs, most of those costs along with the harm itself might be avoided if deterrence were largely successful. See Steven Shavell, *The Social Versus the Private Incentive to Bring Suit in a Costly Legal System*, 11 J. LEGAL STUD. 333 (1982).

information synergies usually are by source rather than by issue. Moreover, the choice of which clumps to examine first if they do relate separately to  $H$  and to  $B$  is optimally guided by their diagnosticity/cost ratios and could readily favor collecting information on  $B$  first. Relatedly, if what we learn about  $B$  tells us that  $B$  may be particularly high, it would then make sense to be much more demanding about  $H$ , and conversely if  $B$  appears to be quite low.

More broadly, learning even a modest amount about both  $H$  and  $B$  can be extremely helpful in screening — in a manner suggested by the tradeoff involved in determining how high to set  $H^*$  if a step 1 test of whether  $H > H^*$  is to be employed. It is really information on both, even if limited, that is ordinarily most probative of whether, on further examination, a decisionmaker might ultimately conclude that  $H > B$  (and by how wide a margin). All things considered, imposing a preliminary requirement that  $H > H^*$  as a screening device, before thinking in balancing terms, is counterproductive in an unconstrained agency context.<sup>111</sup>

Consider next how this two-step decision rule might operate to screen cases in court. Regarding a motion to dismiss, section A questioned whether it is easier to allege as plausible that  $H > H^*$  or that  $H > B$ , with no obvious answer. Surely, if  $H^*$  is sufficiently high, the former is more difficult, but here we are supposing that  $H^*$  may be set fairly low, operating purely for screening purposes. A difference is that, if we employ only the two-step procedure, then the plaintiff has to allege both  $H > H^*$  and  $H > B$ , which can only be harder than alleging the second alone — that is, if we hold constant the standards by which each is judged. Nevertheless, if  $H^*$  is not very high and we are now supposing that  $H > B$  must be plausibly alleged as well, inclusion of the first step does not obviously impose a significantly tougher screen.

Furthermore, instead of employing step 1, it would be possible to stick with unconstrained balancing but be somewhat more demanding about what is regarded to be plausible.<sup>112</sup> In general, if we wish to lean

---

<sup>111</sup> An implication of the analysis in subsection C.2 of rules designed as constraints on balancing, however, suggests that if there is a concern with an agency abusing its power or perhaps simply tending to be overzealous, and if moreover some external check is imposed on its activities (for example, to prevent it from imposing large burdens on a target of an investigation entirely in its discretion), then we face the sort of question considered next regarding screening by courts.

<sup>112</sup> It is familiar that the meaning of the plausibility standard from *Twombly* and *Iqbal* is obscure. See, e.g., Robert G. Bone, *Plausibility Pleading Revisited and Revised: A Comment on Ashcroft v. Iqbal*, 85 NOTRE DAME L. REV. 849 (2010); Kaplow, *Multistage Adjudication*, *supra* note 6, at 1252–59; Adam N. Steinman, *The Pleading Problem*, 62 STAN. L. REV. 1293 (2010). As Kaplow, *Multistage Adjudication*, *supra* note 6, discusses, there are issues regarding what information is available to plaintiffs before filing, the tradeoffs in making plaintiffs undertake more effort on their own prior to filing (the benefit being screening and the cost being that defendants may be able to produce the information more cheaply), and the ability of a judge to defer deciding a motion to

more heavily one way or the other — here, we are contemplating leaning more in favor of defendants — it tends to be optimal to do so directly. That is, it is better to impose a higher burden on the right question rather than to create a separate burden on some additional but incorrect question.<sup>113</sup>

Reference is often made to information that is in the hands of defendants. In some settings, it may be that information regarding  $H$ , which might be suffered by the plaintiff, would be more accessible to the plaintiff before filing a lawsuit than would be information about  $B$ , which often includes costs borne by the defendant.<sup>114</sup> When that is so, it may make sense, even under unconstrained balancing, to be more skeptical early in a proceeding of a plaintiff who is murky regarding  $H$  than of one who offers skimpy allegations regarding  $B$ .<sup>115</sup>

Regarding summary judgment and final decisions at the conclusion of a trial, the reasoning (combined with that in section A) is similar. The key point, of course, is that an important motivation for screening is to limit the ability of a plaintiff to impose substantial costs on a defendant when there is little merit but the potential for significant expense or the extraction of a substantial (above-merits) settlement, the prospect of which chills beneficial conduct. Much of the potential benefit from screening will already have been lost if the screening kicks in only at summary judgment, and all of it will have been forfeited if a case reaches the end of trial. Moreover, if balancing is applied properly, any changed final decisions due to a failure to satisfy step 1's demand that  $H > H^*$  will tend to be mistakes. But if  $H^*$  is set fairly low, step 1 usually will not matter much at later stages in any event.

Step 1's additional demand that  $H > H^*$  will, of course, deter filings if plaintiffs anticipate more negative outcomes down the road or, especially, in deciding motions to dismiss. As mentioned, however, demanding more at the motion to dismiss stage under unconstrained balancing

---

dismiss in order to allow a plaintiff limited discovery to illuminate the plausibility inquiry. *See id.* at 1206–08, 1229 & n.114, 1247, 1283–88.

<sup>113</sup> *See* Kaplow, *Multistage Adjudication*, *supra* note 6, at 1229–35; Kaplow, *Information and Adjudication*, *supra* note 46, at 1330–32. Concretely, instead of requiring that there be some indication that  $H$  exceeds some modest  $H^*$  as a first step, we might instead require an indication that  $H$  exceeds  $B$  by some modest amount. (For example, instead of asking if  $H > k$ , where  $k$  is some small constant, we could ask if  $H > B + k$ .)

<sup>114</sup> For the significant body of information bearing on characterization that simultaneously and relatively pertains to  $H$  and  $B$ , this distinction is inapt. More broadly, the matter varies greatly by context. For example, in Title VII disparate impact cases, information on  $H$  (usually proved by analyzing, say, an employer's hiring data) is generally in the hands of the defendant, although some rough indications might be more readily observable.

<sup>115</sup> For example, as a matter of Bayesian reasoning, we would draw a more negative inference from a limited proffer when the proffering party is likely to know a great deal about a matter than when it probably knows quite little.

would typically be a superior way to accomplish this deterrence, if indeed that would be desirable. (It should be clear from this section's discussion that it is being assumed for purposes of the analysis that there is a net benefit from toughness in screening, which may or may not be so in any particular context.<sup>116</sup>)

### *C. Reflections on Balancing*

In a number of legal settings, there is a reluctance to engage in explicit balancing. As a consequence, there may be an inclination to favor decision procedures that substitute for balancing, minimize the need to do so, or disguise any that takes place. These impulses seem to be part of the explanation for the embrace of structured decision procedures.

Subsection 1 addresses queasiness about balancing that arises from the difficulty of quantifying one or both of the desiderata that need to be balanced or from the felt inappropriateness of expressing the two in a common metric (often referred to as a problem of incommensurability). Here, the main argument is that, however great the challenges, something akin to balancing is the only plausible way to proceed when there are competing considerations, each sometimes powerful enough relative to the other to be decisive. Subsection 2 examines the qualitatively different notion that sometimes rules should constrain balancing because of concerns about institutional competence or the trustworthiness of the decisionmaker. Although these problems are important in some contexts, structured decision procedures of the type considered in this Article are ill-suited to the disciplinary task.

*1. Queasiness About Balancing.* — In order to balance the harm, *H*, that would be avoided by liability and the possible benefit, *B*, that might be forgone, it is necessary to quantify *H* and *B* and to have some basis for comparing the two results. Both tasks are challenging even in many conventional applications of balancing tests. In negligence cases or in the cost-benefit analysis of health and safety regulations, the balance often involves the cost of employing some technology or method to reduce risks and a probabilistic benefit to the life or well-being of some individuals. Measuring each can be complex and speculative, particularly for new techniques on the cost side and with respect to predictions that may involve imperfectly understood causal processes on the benefit side. Moreover, the proper method for trading off financial costs and benefits to life or the quality of life is not obvious. Nevertheless, explicit balancing in such realms is now routine.

This subsection is not the place to revisit all of these challenges or the different ways that they might be overcome. But since queasiness about balancing along these dimensions is common — and is undoubtedly a powerful force in fields of law that employ structured decision

---

<sup>116</sup> See, e.g., *supra* note 110.



---

---

procedures (antitrust, Title VII, and constitutional law) — it is worth examining the core ideas that underlie the balancing process, even in realms where the aforementioned challenges arise.

*Quantification.* — Quantification is the easier of the two sets of concerns: easy in the sense that it is not hard to articulate the response, recognizing that implementation may be a headache. In our cost-benefit example, like in much of the discussion in Parts II and III, we estimate (or guesstimate) as best we can, given the information available, and Part III addresses how the information-gathering process fits in. We employ expected values. In some contexts, it may be appropriate to adjust for risk (such as by using individuals' expected utilities rather than expected monetary outcomes). And so forth.

These points can also be illustrated with the medical treatment example. It is often difficult to quantify both the costs and the benefits of some treatment options, as each may involve predictions about efficacy, side effects, pain, and much more. Furthermore, depending on what subsequently happens, additional treatments might then be required. Yet, when a decision must be made, there is no alternative to making the best estimates we can.

Returning now to our structured decision procedure, note first that it does not come close to fully avoiding quantification. To compare  $H$  with  $H^*$  requires an estimate of  $H$ . Likewise with  $B$  and  $B^*$ .<sup>117</sup> When step 1 fails, of course, this second task is avoided, although as explained in section II.B this short-circuiting is not a virtue. Consider that we could refuse to administer any medical treatment whenever the benefits are below some threshold, avoiding the need in all such cases to quantify the costs. Or refuse treatment whenever the costs are above some threshold, avoiding the need to quantify the benefits. Would anyone go to such a doctor? Or wish to have a regulator use such a decision procedure to decide whether to approve new drugs?

Furthermore, when steps 1 and 2 pass, the structured decision procedure does require that  $H$  and  $B$  be balanced, which in turn requires

---

<sup>117</sup> It might seem as if we might be able to avoid quantification because only a dichotomous decision needs to be made at steps 1 and 2 regarding whether the corresponding thresholds are exceeded. To be sure, in easy cases precise quantification of  $H$  or of  $B$  may not be required, and the same is true in easy cases of direct balancing. However, to address the range of cases and, in particular, close cases, quantification is required. Moreover, the thresholds  $H^*$  and  $B^*$  need to be expressed in quantitative terms (which they generally are not) in order to make the required comparisons. Indeed, returning to easy cases, can we really say that  $H$  obviously exceeds or falls short of  $H^*$  without ever having articulated  $H^*$  in quantitative terms?

In operation, structured decision procedures may succeed in allowing decisionmakers (such as judges) to avoid having to *state* any of their quantitative conclusions because, at step 1 and step 2, they can merely announce whether the (unquantified) thresholds are exceeded. Moreover, even if they believe that both steps pass so that balancing is required, they can obfuscate by adjusting one of their conclusions at an earlier step — for example, by stating that step 2 fails if they believe that  $H > B$ , thereby avoiding the need to articulate how they balanced  $H$  and  $B$ .

that each has been quantified. In all, the structured decision procedure in principle requires that both  $H$  and  $B$  be quantifiable and, in practice, always requires quantification of  $H$  and sometimes of  $B$ .

*Comparison.* — The commensurability of  $H$  and  $B$  involves a qualitatively different challenge when they involve different types of things. Title VII disparate impact decisions may involve trading off discriminatory effects and workplace productivity, and some constitutional law settings may involve a tradeoff of free speech and public safety or national defense.<sup>118</sup> By contrast, some antitrust applications are less demanding along this dimension, such as in comparing anti- and procompetitive effects (both expressed, at least on the surface, in the common denominator of competition or, as often applied, a common notion of welfare, such as effects on consumers).<sup>119</sup>

The question of how appropriately to make various of these tradeoffs is context-specific and often quite challenging. Recall that now-routine cost-benefit analysis often trades off dollars and lives. The question of how to value life has generated a huge literature and continues to be contentious (“life is priceless,” “putting a price tag on life degrades its value”), although most now accept the need for some form of valuation, at least in certain contexts.<sup>120</sup>

One reason for the ascendancy of the technique is the recognition that its use enables more lives to be saved for a given expenditure.<sup>121</sup> If, for example, we are spending \$100,000 at the margin per (statistical) life saved with one regulation and \$100,000,000 at the margin per life saved with another, then we could save many more lives at the same cost:

<sup>118</sup> See Kaplow, *supra* note 1.

<sup>119</sup> See, e.g., FED. TRADE COMM’N & U.S. DEP’T OF JUSTICE, ANTITRUST GUIDELINES FOR COLLABORATIONS AMONG COMPETITORS 24–25 (2000) (“If the relevant agreement is reasonably necessary to achieve cognizable efficiencies, the Agencies assess the likelihood and magnitude of cognizable efficiencies and anticompetitive harms to determine the agreement’s overall actual or likely effect on competition in the relevant market. To make the requisite determination, the Agencies consider whether cognizable efficiencies likely would be sufficient to offset the potential of the agreement to harm consumers in the relevant market, for example, by preventing price increases.”).

<sup>120</sup> See, e.g., W. Kip Viscusi, *The Value of Individual and Societal Risks to Life and Health*, in 1 HANDBOOK OF THE ECONOMICS OF RISK AND UNCERTAINTY 385 (Mark J. Machina & W. Kip Viscusi eds., 2014). Similarly, much applied research on medical decisionmaking has focused on how to make difficult tradeoffs between, say, prolonging life and quality of life in various settings (as well as how to communicate pertinent information relating to quantification to patients). See, e.g., Dawn Stacey et al., *Decision Aids for People Facing Health Treatment or Screening Decisions*, COCHRANE DATABASE OF SYSTEMATIC REVIEWS, 1, Jan. 28, 2014. The substantial efforts with moderate returns are a testament to both the difficulty of the challenge and the ability to make progress when the tradeoff problem is confronted explicitly.

<sup>121</sup> This perspective is sometimes referred to as cost-effectiveness analysis rather than cost-benefit analysis. When there is a fixed pool of resources to be spent saving (statistical) lives, we can ask how to allocate the funds so as to save the most lives without answering the value of life question. But an answer is necessary to set the budget.

simply relax the latter regulation a bit — enough to save one less life, thus freeing up \$100,000,000 — and tighten the other regulation by enough that essentially spends that savings, which would save 1000 lives.<sup>122</sup> If we value both money and lives (or, really, any two things), getting the most bang for the buck requires making explicit, consistent tradeoffs.

This illustration and the argument for balancing more broadly involve an application of principles that were developed in the mid-twentieth century in work at the intersection of the fields of probability and statistics, economics, and game theory.<sup>123</sup> The central concept for present purposes is that, whenever we care about (at least) two factors, logical consistency requires that decisions be made *as if* we are engaged in balancing. Because the underpinnings of this idea offer a helpful perspective on the commensurability problem, they warrant further explication here.<sup>124</sup>

Suppose that a decisionmaker must be able to resolve a broad range of possible disputes wherein the only two considerations are *H* and *B*. Given the distribution of possible values of *H* and *B*<sup>125</sup> and the decisionmaker's judgment of their importance (that is, their relative weight),<sup>126</sup> assume further that at least sometimes *H* is large enough relative to *B* to warrant the assignment of liability and at least

---

<sup>122</sup> Due to diminishing returns, it may be that it would save far fewer than 1000 lives, although many more than 1. The core lesson stands.

<sup>123</sup> See generally LEONARD J. SAVAGE, *THE FOUNDATIONS OF STATISTICS* (1954); JOHN VON NEUMANN & OSKAR MORGENTHAU, *THEORY OF GAMES AND ECONOMIC BEHAVIOR* (2d ed. 1947). Although the exposition in the text focuses on commensurability, many of the relevant developments addressed aspects of what is here categorized under quantification: in particular, decisions under uncertainty, including where frequentist notions of probability seemed insufficient to ground probability judgments, giving rise to the concept of subjective probabilities (that is, a decisionmaker who makes choices as if particular probability values were taken to be true).

<sup>124</sup> A distinct approach to the challenge of commensurability with regard to comparison (setting to the side the difficulty of quantification) is to regard all that is socially relevant to be understood in terms of a common metric, such as welfare. This method — like those that may not require this assumption, which are the focus of the analysis in the text — requires normative defense. See, e.g., LOUIS KAPLOW & STEVEN SHAPELL, *FAIRNESS VERSUS WELFARE* (2002) (advancing welfare as the sole appropriate first-level principle); *id.* at 406–07 (explaining how concerns such as controlling government officials' behavior may call for rules that limit unconstrained welfare maximization by government agents, as discussed in subsection 2); Louis Kaplow & Steven Shavell, *Any Non-Welfarist Method of Policy Assessment Violates the Pareto Principle*, 109 J. POL. ECON. 281 (2001) (proving that giving any weight to any non-welfarist principle, if done consistently, entails sometimes choosing policies that make everyone worse off); Louis Kaplow & Steven Shavell, *The Conflict Between Notions of Fairness and the Pareto Principle*, 1 AM. L. & ECON. REV. 63 (1999) (same).

<sup>125</sup> Problems of quantification are set to the side in the present discussion.

<sup>126</sup> This could be a personal judgment, a social judgment that the decisionmaker observes and adopts, a weighting enshrined in an authoritative document (such as a constitution), or one that derives from any other source.

sometimes  $B$  is large enough relative to  $H$  to warrant no liability.<sup>127</sup> The claim established now is that, if the decisionmaker is to be consistent, the potential decisions as a whole will exhibit certain properties, and these properties are aptly described *as* those that would have been produced *if* some mode of balancing had been employed (hence the “as if” locution).<sup>128</sup>

To begin, consider some particular possible value of  $B$  — one that is high enough that sometimes there would be a decision of no liability (there exists some  $H$  low enough such that this is so) and is also low enough that sometimes there would be a decision of liability (there exists some  $H$  high enough such that this is so). Focusing on the aforementioned  $H$  that is low enough such that there is no liability, surely (continuing with our given value of  $B$ ) any lower  $H$  should likewise result in no liability. Similarly, if we consider the  $H$  that was high enough

<sup>127</sup> This is a maintained assumption for purposes of this Article, and one that seems widely embraced even by many who are queasy about balancing. See, e.g., Arthur Allen Leff, *The Leff Dictionary of Law: A Fragment*, 94 YALE L.J. 1855, 2123–24 (1985) (“[B]alancing. The metaphoric term generally used in the law to describe an exceedingly important conceptual operation. In almost all conflicts, especially those that make their way into a legal system, there is something to be said in favor of two or more outcomes. Whatever result is chosen, someone will be advantaged and someone will be disadvantaged; some policy will be promoted at the expense of some other. Hence it is often said that a ‘balancing operation’ must be undertaken, with the ‘correct’ decision seen as the one yielding the greatest net benefit. That some such process must be a part of any practical legal system is undeniable. But that should not blind us to the extreme danger of too facile a use of ‘balancing’ in a system of justice. . . . It is to say, however, that not all legal questions are thoroughly amenable to the same process by which lumps of matter are compared for gravitational attraction.”) It is logically possible that, even though both count, and either could in principle outweigh the other, it is believed that in the relevant domain it is always (or nearly always) true that one side outweighs the other.

It is also logically possible to view either  $H$  or  $B$  as trumping the other, that is, regardless of their relative magnitudes. Perhaps it is simply required that one be positive rather than zero to dictate the outcome. There are many familiar problems with such views. When there are multiple items — perhaps rights — that are each trumps, conflicts among them need to be resolved, which might be done with a hierarchy of trumps (if not, tradeoffs would need to be admitted). In addition, taking anything literally as a trump, although not illogical, can involve truly extreme prescriptions: if some right trumps budgetary savings, then the entire government budget should be spent to reduce by even a minuscule amount the probability that someone’s right might be infringed ever so slightly at some distant point in time. (This example also reveals a substantial weakness in suggestions that we might wriggle out of this problem by permitting the trump to be overridden when the government benefit is massive, such as in averting a catastrophe. That is,  $B^*$  is not infinite, but just very large. At that point, following this Article’s analysis, we want to know if this  $B^*$  is the same, regardless of the degree of infringement — so we require the same level of catastrophe to avoid the trump regardless of whether the infringement wipes out all civil liberties permanently or involves a slight, temporary infringement for a few individuals. If not, then we are engaged in balancing. See also *infra* note 134 (discussing the confusion that according high weight to an interest somehow renders balancing inappropriate.) More plausibly, we might attempt to ground rights as trumps, never to be traded off, as a form of rules designed to constrain balancing because a decisionmaker is untrustworthy, the subject of subsection 2.

<sup>128</sup> Stated more formally, there exists a transformation of the measure of  $B$  (or, alternatively, of  $H$ ) such that the results will accord with the assignment of liability if and only if  $H > B$  (again, setting ties to the side). See *infra* note 131.

for there to be liability, surely any higher  $H$  should likewise result in liability.

Reflection on these two observations suggests that, for our given value of  $B$ , there will be some critical value of  $H$ , above which there should always be liability and below which there should never be liability. As we gradually imagine raising  $H$  from some low level (where there is no liability) to higher levels, at some point the decision will switch to one of liability, and it will stay there for even higher levels of  $H$ .

To drive this point home, suppose that this were not so. Then it must be that, as  $H$  rises, the decision at some point switches to liability, but there would be some even higher level of  $H$  at which the decision changes back to no liability. This switch back obviously does not make sense. Put another way, when we suppose the contradiction, then there must exist two distinct values of  $H$ , a higher one involving no liability and a lower one involving liability (for the same  $B$ ).

This conclusion — that, for any given  $B$ , there must be some critical value of  $H$  that marks the switch from no liability (for all lower values of  $H$ ) to liability (for all higher values of  $H$ ) — applies to any value of  $B$  that we might have selected.<sup>129</sup> Note further that the mirror image conclusion follows by the same reasoning: For any given  $H$ , there must exist some critical value of  $B$  that marks the switch from liability (for all lower values of  $B$ ) to no liability (for all higher values of  $B$ ). And this holds for any value of  $H$  that we might have selected.<sup>130</sup>

When a decisionmaker contemplates the foregoing, for all possible values of  $H$  and of  $B$ , the resulting outcomes can be represented as a two-dimensional graph, with  $H$  on the horizontal axis and  $B$  on the vertical axis, as depicted in Figure 1. A curve (which we will not restrict to be a line<sup>131</sup>), rising from the lower left to the upper right, divides the space between decisions that assign liability (the lower right, where  $H$  is large relative to  $B$ ) and decisions that assign no liability (the upper left, where  $B$  is large relative to  $H$ ).<sup>132</sup>

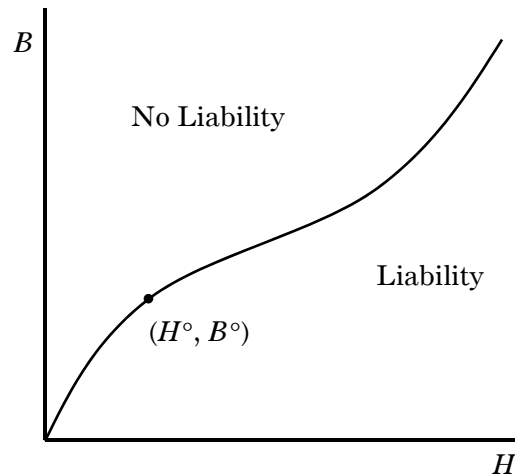
<sup>129</sup> For expositional convenience, the text sets to the side that there might exist some values of  $B$  so low that any value of  $H$  results in liability or some values of  $B$  so high that any value of  $H$  results in no liability.

<sup>130</sup> Subject to a caveat analogous to that expressed in the preceding footnote.

<sup>131</sup> As some will appreciate, the use of a line rather than any monotonically increasing curve would not really be a substantive restriction because we could, for example, taking as given the units in which  $H$  is measured, recalibrate the units of  $B$  (in a monotonic but nonlinear fashion) so that the diagram, with a suitably reinterpreted vertical axis, could be restated with a linear divider. This recalibration of  $B$  (or instead a recalibration of  $H$ , or of both) is what it means to put  $B$  and  $H$  in a common denominator. (To accommodate uncertainty, there are additional restrictions involved, which will not be considered further here.)

<sup>132</sup> For those comfortable with the mathematical representation, it will be apparent that taking  $H$  and  $B$  each to be unidimensional is an unnecessary simplification. Allowing either or both to be multidimensional permits the formal representation of implicit tradeoffs not only between  $H$  and  $B$  but also elements of  $H$  versus elements of  $B$  as well as elements of  $H$  versus each other and elements of  $B$  versus each other.

Figure 1: “As If” Balancing



If we reflect on the matter, we will appreciate the correspondence between Figure 1 and the aforementioned logic. If we pick any point on the curve, say  $(H^\circ, B^\circ)$ , it indicates that, for the given  $H^\circ$  and  $B^\circ$  that are a critical point at which the liability decision flips, the following are true: For the given  $B^\circ$ , any higher  $H$  than that critical  $H^\circ$  is associated with a decision of liability and any lower  $H$  with no liability. And for the given  $H^\circ$ , any higher  $B$  than that critical  $B^\circ$  is associated with a decision of no liability and any lower  $B$  with liability.

The possibility of such a representation of consistent decisionmaking is what is meant here by “as if” balancing.<sup>133</sup> The central ingredients are, as the foregoing suggests, fairly minimal. The decisionmaker is taken to care about both  $H$  and  $B$ , preferring (all else equal) less  $H$  and more  $B$ . Note, importantly, that the argument does not in any way restrict how the decisionmaker might value  $H$  and  $B$ , that is, the weights placed on each.<sup>134</sup> All that is required is that decisions be consistent in the sense elaborated here. When these conditions hold, balancing not

<sup>133</sup> The sense in which what has been described is formally equivalent to “balancing,” understood literally as assigning liability if and only if  $H > B$ , is addressed in notes 128 and 131.

<sup>134</sup> It is sometimes thought that balancing is inappropriate when one side, say,  $H$ , is much weightier than the other. This objection reflects a misunderstanding because there is no restriction on the relative weights (or even a requirement that simple, linear weights be employed, *see supra* note 131). If  $H$  is ten times more important than  $B$  is — a statement that can only be rendered meaningful if, among other things, we state the units in which each is measured — then we could either require instead that  $10 \times H > B$  or change the units in which  $H$  is measured (converting each of the original units into 10 of the transformed units), which would produce the same outcome. *See also supra* note 127 (on rights as trumps).

only makes sense<sup>135</sup> but really constitutes the only sensible mode of decisionmaking in such an environment.<sup>136</sup>

This depiction is consonant with the decision process in our illustrations involving cost-benefit analysis and medical decisionmaking. All else equal, greater costs — money, pain, or whatever — are undesirable; when they are very low, it makes sense to proceed, but when they are sufficiently high, not to proceed. And conversely for the benefits, be they in terms of saving (statistical) lives, improving the quality of life, or advancing some other value. Just as individuals choosing courses of medical treatment or regulators deciding on safety regulations and drug approvals cannot avoid attempts at quantification of both costs and benefits, however difficult that might be, neither can they sensibly avoid making tradeoffs. And consistent resolution of cases involving tradeoffs necessarily involves something tantamount to a balancing process.<sup>137</sup>

As an aside, it is curious that we often encounter great resistance to balancing in the legal realm in light of the fact that the longstanding symbol of justice is Lady Justice's scales, a central source of the balancing metaphor in legal settings.<sup>138</sup> Perhaps the blindfold, conventionally understood to depict impartiality, should be interpreted instead as reflecting her queasiness about balancing. But I doubt it.

Finally, keep in mind that structured decision procedures of the type under examination here do not avoid balancing altogether. If the objection were that balancing is impossible or inappropriate in principle, this would be a serious limitation to using these protocols in lieu of unconstrained balancing. To avoid balancing altogether, we could eliminate step 3, replacing it with a decision of no liability: that is, if step 1 passes

---

<sup>135</sup> For some further discussion of the notion of expressing two factors in a common denominator, which is implicit in the enterprise but need not reflect an explicit thought process of our decisionmaker, see note 131.

<sup>136</sup> Indeed, a difficulty in understanding some purported objections to balancing (particularly on grounds of incommensurability) is identifying what the alternative might be, particularly once we appreciate that (unless one side always trumps the other, see *supra* note 127) any other mode of decisionmaking must be internally inconsistent.

<sup>137</sup> A further question concerns how explicit the balancing should be. In spite of the understandable queasiness (consider in particular Title VII disparate impact law and strict scrutiny in constitutional law, see Kaplow, *supra* note 1), there are a number of familiar benefits of making the method of decisionmaking explicit: in particular, accountability and improving the quality of analysis (which is difficult when those undertaking it are unwilling to admit, and may not even recognize, what is actually required).

<sup>138</sup> When drafting this segment, I performed a search in Google Images for "justice," and each of the first fourteen images contained scales (although they subsequently account for only a substantial majority). See also T. Alexander Aleinikoff, *Constitutional Law in the Age of Balancing*, 96 YALE L.J. 943, 962 (1987); Dennis E. Curtis & Judith Resnik, Essay, *Images of Justice*, 96 YALE L.J. 1727, 1741 n.32 (1987) (tracing the ancient origins of the use of scales to symbolize decisionmaking as balancing).

and step 2 also passes ( $H > H^*$  and also  $B > B^*$ ), then there is no liability.<sup>139</sup> Of course, this expands the set of situations in which there would be erroneous outcomes (from the perspective of balancing).<sup>140</sup> Moreover, we need to ask how to go about choosing  $H^*$  and  $B^*$  in setting up the rule. This imposes quite a challenge if  $H$  and  $B$  are regarded as incommensurable in principle.<sup>141</sup> If  $H^*$  is set extremely high or  $B^*$  is set extremely low, the result is close to one of never assigning liability, and if  $H^*$  is set very low and  $B^*$  very high, the result is close to one of always assigning liability. Where in between should  $H^*$  and  $B^*$  be set? Raising or lowering each of these thresholds affects the range of circumstances (combinations of  $H$  and  $B$ ) in which liability and no liability are assigned. How could we defensibly set these thresholds without in some sense making tradeoffs (in the aggregate) between  $H$  and  $B$ ?<sup>142</sup>

2. *Rules Designed to Constrain Balancing.* — It is familiar that rules can be superior to unconstrained balancing for a variety of reasons.<sup>143</sup> For ease of exposition and emphasis, this subsection sets aside the use of rules to economize on decisionmaking costs, improve the compliance of primary actors (who may be better guided when there is greater ex ante specification of the law's commands), and reduce the costs of enforcement.<sup>144</sup> The discussion here focuses instead on rules designed for

<sup>139</sup> We could then append a less restrictive alternatives requirement and, moreover, confine consideration to those that are equally effective. See *supra* section II.C.

<sup>140</sup> Consider the case in which  $H$  barely exceeds  $H^*$  and  $B$  is barely below  $B^*$ , so that liability is assigned. And compare that to the case in which  $H$  is trivially lower (but no longer exceeds  $H^*$ ) and  $B$  plummets to 0. Can no liability really make sense? Similarly, consider the case in which  $H$  barely passes  $H^*$  but  $B$  now just passes  $B^*$ , so we would assign no liability. And compare that to the case in which  $H$  is ten times as large. Can no liability still make sense? Even though, without  $H$  being any larger, we would assign liability if  $B$  were trivially smaller?

<sup>141</sup> It is possible that some writers, including authors of judicial opinions, are confused about this matter or are perhaps subconsciously hiding their implicit balancing from the reader. These phenomena are facilitated by expressing  $H^*$  and  $B^*$  in loose, nonquantitative terms (like “significant,” “substantial,” or “compelling”) and also by failing to state a quantitative conclusion on  $H$  and  $B$  in the case at hand. All that is stated explicitly might be the verbal analogue to  $H \leq H^*$ , or, if step 1 passes, that  $B > B^*$  (or not). A conclusion can be based on an implicit, even subconscious balance of  $H$  and  $B$  and yet be expressed in a manner that denies that any balancing took place or at least fails affirmatively to acknowledge that it did. See also *supra* note 137 (noting the familiar reasons that being explicit is usually preferable).

<sup>142</sup> Moreover, as developed in section II.B, for any given view on the relative importance of  $H$  and  $B$ , how could we justify choosing a decision rule that predictably produces poor outcomes under that view?

<sup>143</sup> See, e.g., FREDERICK SCHAUER, *PLAYING BY THE RULES: A PHILOSOPHICAL EXAMINATION OF RULE-BASED DECISION-MAKING IN LAW AND IN LIFE* 135–66 (1991).

<sup>144</sup> See, e.g., Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557 (1992). Although these sorts of benefits from rules are not the present focus, they may complement the force of some of the considerations noted below. For example, a reviewing court might choose to employ rules involving more restriction or greater simplification than may seem necessary to limit the potential for abuse in order to economize on enforcement costs and to improve compliance by individual agents or other branches of government. Note further that all of these considerations, or analogues thereto (placing more emphasis on institutional competence than on corruption or



the purpose of constraining decisionmakers who may not be fully reliable for reasons of institutional competence or because they may be untrustworthy.<sup>145</sup> This choice of emphasis reflects in part the use of a structured decision rule in formulating constitutional law's strict scrutiny doctrine, wherein a reviewing court must decide whether to invalidate the actions of another branch of government or of particular government agents.

Most agents, whether inside or outside the government, are in various ways constrained in their actions, not being left free to pursue their principal's overall objectives as the agent best sees fit. Tax auditors do not determine, *de novo*, how much tax an individual should have paid; traffic officers do not define offenses and set fines; army privates do not choose which nations to invade; nor do assembly-line workers decide what to make or marketing vice presidents whether to merge. The reasons for these myriad rules — many limiting the subject matter of the agent's discretion and available actions — include each agent's limited domain of comparative advantage as well as concerns that, if left free, the agent would pursue a personal view of the principal's objective or raw self-interest and may be subject to corruption and bias. Even when rules are highly imperfect, the results under them may be superior to those when behavior is unconstrained.

---

other abuse), might be applicable to the formulation of rules designed to constrain juries. *See supra* section A. Structured decision rules in antitrust occasionally seem to have this motivation; much of the analysis in this subsection (particularly that below concerning the comparison of *H* to *H\** and of *B* to *B\**) has implications for that application as well.

<sup>145</sup> *See, e.g.*, Wolff v. McDonnell, 418 U.S. 539, 558 (1974) ("The touchstone of due process is protection of the individual against arbitrary action of government." (citation omitted)); ARISTOTLE, THE NICOMACHEAN ETHICS bk. V, ch. 6 (J.L. Ackrill & J.O. Urmson eds., David Ross trans., Oxford Univ. Press rev. ed. 1980) (c. 384 B.C.E.) ("This is why we do not allow a *man* to rule, but *rational principle*, because a man behaves thus in his own interests and becomes a tyrant."); Meir Dan-Cohen, *Decision Rules and Conduct Rules: On Acoustic Separation in Criminal Law*, 97 HARV. L. REV. 625, 650 (1984) (discussing "the need to shape, control, and constrain the power wielded by [official legal] decisionmakers"); Russell Hardin, *The Utilitarian Logic of Liberalism*, 97 ETHICS 47, 47 (1986) (arguing that rights are important because institutions that leave agents free to judge outcomes based on their utility would be unreliable); Margaret Raymond, *Rejecting Totalitarianism: Translating the Guarantees of Constitutional Criminal Procedure*, 76 N.C. L. REV. 1193 (1998) (suggesting that post-World War II constitutional procedural doctrine is driven by concerns about totalitarianism that are analogous to the concerns about tyranny that originally led to the constitutional amendments); T.M. Scanlon, *Due Process, in NOMOS XVIII: DUE PROCESS* 93, 93-100 (J. Roland Pennock & John W. Chapman eds., 1977) (noting the "truism that due process is concerned with protection against arbitrary decisions," *id.* at 93, suggesting that due process is concerned with the improper exercise of power, and referring to due process as "one of the strategies through which one may seek to avoid arbitrary power," *id.* at 97). The idea that individual rights should be viewed instrumentally because they are important protections against government abuse has long been familiar. *See, e.g.*, JOHN STUART MILL, ON LIBERTY (1859); J.S. MILL, UTILITARIANISM 93 (Roger Crisp ed., Oxford Univ. Press 1998) (1861) ("We should be glad to see just conduct enforced and injustice repressed, even in the minutest details, if we were not, with reason, afraid of trusting the magistrate with so unlimited an amount of power over individuals.").

Government enforcement agencies and courts are similarly subject to such rules, many limiting their jurisdiction and others constraining their behavior in their given domains. Such rules, moreover, may appropriately restrict not only primary government actors (agency enforcers, lower court judges) but also those reviewing those actors, including the Supreme Court. Particularly when decisionmakers are not entirely reliable, rules that constrain balancing — and thereby sometimes result in decisions contrary to what a properly performed balance would prescribe — can be desirable.

It does not follow, of course, that any rule will be superior to balancing when such concerns are present. The challenge is to analyze the comparative advantages and potential dangers associated with various actors, including those engaged in the review of others, in a manner that helps identify what rules, if any, should partially or fully displace balancing. This perspective is important in many areas of law and is particularly central in the design of constitutions and the operation of constitutional law.<sup>146</sup>

In a constitutional scheme, we might decide that it is appropriate to have courts review some or all of the decisions of other branches of government for their constitutionality. If that choice is made, it becomes necessary to specify the terms of that review.<sup>147</sup> The constitution (if it is a written one) and doctrines developed by the reviewing courts generally employ various rules both to constrain other government actors and to govern the operation of the courts themselves. The focus here is primarily on two concerns regarding misbehavior.<sup>148</sup> First, other branches or actors may behave in ways that are designed not to advance the social good but rather to entrench themselves and expand their power (which are addressed, for example, by rules that limit executive

---

<sup>146</sup> An exemplary but insufficiently appreciated treatment from this perspective is JEREMY BENTHAM, *THE COLLECTED WORKS OF JEREMY BENTHAM: SECURITIES AGAINST MISRULE AND OTHER CONSTITUTIONAL WRITINGS FOR TRIPOLI AND GREECE* (Philip Schofield ed., Oxford Univ. Press 1990) (1822–23). It is notable that, as a thoroughgoing utilitarian, Bentham thought much about and argued strongly for robust constitutional limitations on government.

<sup>147</sup> The importance of engaging in this sort of multilevel inquiry to justify such a set-up and give it content is a central theme of ERWIN CHEMERINSKY, *INTERPRETING THE CONSTITUTION* (1987), and is elaborated in the context of proportionality analysis in NIELS PETERSEN, *PROPORTIONALITY AND JUDICIAL ACTIVISM: FUNDAMENTAL RIGHTS ADJUDICATION IN CANADA, GERMANY AND SOUTH AFRICA* 13–37 (2017). Note further the a priori indeterminate implications of the fact that constitutional review is a second-best mechanism in a realm characterized by various constraints for the proper role of balancing. On one hand, because we are in a second-best world with other constraints and are thus engaged in an indirect maximization exercise, any strong presumption that unconstrained balancing is best seems dubious. On the other hand, it does not follow from our wish to constrain some agents that others (their supervisors in a sense) should necessarily be constrained, similarly or at all. It is necessary to engage in substantial analysis, conceptual and empirical, in order to derive answers in this domain.

<sup>148</sup> These are chosen, instead of corruption and catering to special interests, to reflect the apparent concerns of constitutional law's strict scrutiny doctrine.

---

---

power and that protect the freedom of speech and of the press). Second, they may fail to give due weight — or even assign negative weight — to the interests of dispossessed groups.

Consider further the question of institutional design regarding the extent to which the reviewing courts should be mistrusted (for these reasons or others) in such settings. If there were no concern whatsoever, then even though rules may optimally be deployed to constrain balancing by others, there would be no need to constrain the reviewing courts.<sup>149</sup> But if there were concerns, then some (and perhaps substantial) restrictions would be appropriate.

With these brief reminders, turn now to the question whether structured decision procedures of the type examined here are good candidates in this domain, that is, the sort of rules that are well-devised to constrain the abuses with which we are concerned. It will be explained that the fit is not close — in contrast with certain other types of rules, such as a requirement to apportion House seats by population or the imposition of term limits. The central mismatch is that, although the structured decision procedure does constrain a conscientious decisionmaker, by requiring both bad decisions and inefficient information collection protocols, it does little to constrain the aforementioned sorts of abuse.

Consider how unconstrained balancing might lend itself to abuse. Even though everyone knows that the rule is to assign liability if and only if  $H > B$ , the untrustworthy decisionmaker may do otherwise and go undetected if it is possible to misrepresent the values of  $H$  and  $B$ . Perhaps to punish a political opponent, liability might be assigned even though  $H$  is much less than  $B$  (indeed, there may be no real  $H$ , only an invented one). Or perhaps because of an impermissible bias, the decisionmaker may assign no liability even though  $H > B$  (misrepresenting  $H$ ,  $B$ , or both).

For this to be a problem, it must be that review of such decisions is infeasible, or at least very costly or of otherwise limited efficacy. When, instead, direct reassessment of the untrusted agent's balancing can readily be undertaken and has the requisite transparency, balancing by the reviewing court may well be the appropriate way to constrain others so as to avoid abuse.

Problems arise primarily when the reviewing court cannot readily observe  $H$  and/or  $B$ . An important and related point is that the public

---

<sup>149</sup> It does not follow that the review would involve unconstrained balancing because, if the agents being reviewed were themselves subject to some other rule, review would typically determine whether the agents' behavior conformed to that rule's dictates. It is important to note that the use of and advocacy for structured decision procedures for constitutional review is not accompanied by imposition of the same protocol on the primary decisionmaker. For example, a legislature is supposedly free to engage in unconstrained balancing, whereas the reviewing court is the one that is constrained by the form of the rule. Such a juxtaposition might be justified by limitations on the latter's institutional competence and for other reasons, which will be considered below, but does not immediately follow from the primary argument offered here.

---

---

likewise may be unable to exercise oversight under such a rule. That is, it may be that the original agents, lower courts, and even the Supreme Court are untrustworthy, but if the decision is supposed to reflect whether  $H > B$ , public accountability will depend on the possibility that the  $H$  and  $B$  in some cases, or at least in important cases, are observable. When the true values of these components are opaque, this is not possible. (Compare, for example, the rule on apportionment of representation in the House or term limits; despite some imperfections with the former, significant deviation would typically be apparent.)

This core reason that a balancing test may fail to constrain abuse applies substantially to the sort of structured decision rule considered here. Step 1 asks whether  $H > H^*$ . If  $H$  cannot readily be observed by the reviewer, even when substantially manipulated by the untrustworthy agent or branch of government under review, the test does not get off the ground. There is the related problem that, if our reviewer is itself an unreliable agent, any onlooker may have trouble confirming its assessment of  $H$  (the magnitude in the particular case). Furthermore, the magnitude of  $H^*$  may be obscure as a consequence of the failure in practice to specify the decision thresholds in quantitative terms. That is, even if it is obvious what  $H$  is, or at least that  $H$  clearly exceeds  $B$ , perhaps the reviewing court was correct in assigning no liability because it found that  $H \leq H^*$ . Employing a step 1 with its associated  $H^*$  may make it harder, not easier, under the structured decision procedure to know whether the decisionmaker is following the rules.

The analysis of step 2, regarding the comparison of  $B$  and  $B^*$ , is similar. If  $B$  cannot be assessed by the reviewing court or observed even approximately by an onlooker, manipulation of the application of this step is also undetectable, and the problem is only worsened by any additional ambiguity regarding  $B^*$ . That is, both step 1 and step 2 — where the antibalancing action of the structured decision procedure lies — may fail to improve and perhaps might degrade the transparency and thus the ultimate accountability of decisionmaking. Finally, in those cases in which step 3 is reached, the requisite balancing is subject to essentially the same limitations applicable to an unconstrained balancing test.

Accordingly, the structured decision procedure under examination in this Article seems to be a poor candidate for a rule that would effectively constrain balancing when we are concerned about the reliability of original decisionmakers or reviewing courts. This limitation, however, is not absolute; there remain situations in which these protocols may be of some help.

Specifically, focusing on step 1, consider realms in which  $H$  is externally observable (at least sufficiently often and to a sufficient degree to matter) but  $B$  is opaque. Under this asymmetrical set of assumptions, step 1 can be operative as a constraint even when the review of an unconstrained balancing decision is infeasible due to the inscrutability of

*B*. That is, we are imagining that the reviewing court — and perhaps our onlooker — can determine reasonably well the magnitude of *H*, and in addition we are now supposing that  $H^*$  has some clarity. Then, if  $H \leq H^*$  — or, perhaps more plausibly and consequentially, if *H* is significantly less than  $H^*$  — it will be clear that there should be no liability under the structured decision rule. Any decision to the contrary would entail observable misbehavior, the prospect of which would tend to deter such abuse. Now, in cases in which  $H > H^*$  — and perhaps to a substantial degree — we would still be unable (under the present assumptions) to check abuse involving assignments of no liability because we are unable to verify pronouncements about *B*.

With this point in mind, it might sometimes be possible to fashion a structured decision procedure that operates as a partial constraint. In such cases — that is, when the relevant configuration of assumptions holds — the action will be in step 1 (not all three steps), and the central design question will be how high to set  $H^*$ .<sup>150</sup> Recall that, as a constraint,  $H^*$  is relevant when *H* turns out to fall below (perhaps significantly below)  $H^*$ . A higher  $H^*$  will bind more often, helping to deter or overrule abuse in a wider range of cases. On the other hand, we know from section II.B that the higher is  $H^*$ , the more often step 1 dictates incorrect outcomes, which is contrary to what is optimal in light of the actual levels of *H* and *B*. That is the basic tradeoff.<sup>151</sup>

Another variant would employ a step 1 that is categorical and jurisdictional. That is, instead of asking whether *H* is large enough to proceed to step 2, the rule would inquire whether *H* involves the type of harm that is subject to review. This formulation — which seems pertinent, for example, in considering the doctrine of strict scrutiny in constitutional law — is more typical of rules that are designed to constrain balancing. This approach often defines the relevant category either qualitatively (a given agent has power in realm *A* but not in realm *B*; decisions with trait *X* are subject to review but not those with trait *Y*) or using observable quantities (like population, age, or a number of terms). These sorts of rules, as is familiar, are not without their own problems: boundaries may be malleable, and because they often involve costly restrictions there is the temptation to bend if not break the restraints (often reflecting at least implicit balancing).<sup>152</sup> Once such barriers break down, little effective constraint may remain, even when more

---

<sup>150</sup> In the circumstances imagined in the text, step 2 should be omitted because it only constrains (and thereby distorts the actions of) conscientious decisionmakers.

<sup>151</sup> Cf. *supra* section B (discussing the setting of  $H^* > 0$  as a screening device).

<sup>152</sup> Cf. Richard H. Fallon, Jr., *Strict Judicial Scrutiny*, 54 UCLA L. REV. 1267, 1272 (2007) (“But if doctrinal formulas acquire a life of their own, they can never achieve more than limited autonomy, for courts inevitably apply doctrine in purposive ways. However banal, this point is an important one. Probably in common with many other doctrinal formulas, strict scrutiny developed partly as a device of judicial self-discipline, but judicial self-discipline is always imperfect and fraught with ambivalence, as the history and practice of strict judicial scrutiny unmistakably teach.”).

explicit balancing of the benefits and risks might sometimes have been effective in maintaining some control where such is particularly valuable.

## V. CONCLUSION

The use of structured decision procedures in lieu of unconstrained balancing in some areas of law — with no contemplation thereof in most others — is a puzzling phenomenon. Also surprising and more problematic is the failure of courts and commentators to articulate structured protocols with any precision, identify how they diverge from balancing, and address how the differences might be justified. This Article analyzes these issues in general terms.

First, as rules of final decision, structured rules deviate from balancing in two ways: sometimes they assign no liability even though the harm that liability would avert exceeds any benefit that would be forgone, and sometimes they assign liability even though the harm averted is less than the benefit forgone. Moreover, in both instances (particularly the latter), some or all of the information indicating that the outcome is erroneous is before the decisionmaker and may have already been processed, but under the structured rule must be ignored.

Structured decision procedures are often favored because they avoid difficult balancing. But when we trace the protocols' actual implications, this supposition is revealed to be largely mistaken. Many of the avoided balances are the easiest. For example, if harm is established and there is no countervailing benefit, skipping the final balancing step is inconsequential. Difficult balances are sometimes avoided, but this is done by imposing an outcome that may well be mistaken. We may as well flip a coin to decide hard cases. Adhering to the protocols can also require making difficult decisions that, if made correctly, are inconsequential and, if made incorrectly, lead to erroneous outcomes in cases in which direct balancing would have been easier.

Second, as guides to information gathering, structured decision procedures are significantly deficient. They violate every lesson of optimal information collection, which is what would guide information gathering under unconstrained balancing. For example, they may call for collecting expensive and only marginally illuminating information before cheap and highly probative information. They call for separating the collection of types of information that are best gathered together. And their only interim stopping rule — allowing an assignment of no liability after collecting all of the evidence on harm — is inappropriate not only as a final decision rule but also as an interim one, among other reasons because it ignores information that often is already in hand concerning benefit. Perhaps most important, much probative evidence is relevant

precisely because of how it bears relatively on the plausibility of a harmful versus a beneficial explanation for an action under scrutiny, so it is incoherent to ask how it bears on harm in a vacuum.

It might be supposed that agency investigations are largely guided by the dictates of optimal information collection, even in realms governed by structured decision procedures. By contrast, U.S. civil litigation, as currently structured and typically conducted, is strongly inconsistent with both approaches to information collection. Notably, unless a motion to dismiss is granted, discovery pertaining to all issues is conducted, without sequencing by issue or by priority in terms of diagnosticity and cost, and without interim decisions. The next possible stopping point is summary judgment, by which time all of the information is in hand and much of it has been processed. Hence, it is obviously erroneous to suggest that structured decision procedures sometimes save the substantial costs of collecting information on a challenged act's benefits because the case can be disposed of if the plaintiff fails to demonstrate sufficient harm.

Regarding both information gathering and final decisionmaking, decisionmakers probably engage in some (conscious or subconscious) degree of reverse engineering under structured decision procedures. For example, in deciding a close case at step 1, where they are supposed to consider only whether harm exceeds some threshold, they may be more likely to assign no liability at that point if (peeking ahead, contrary to the protocol) they see that the benefit is large, and they may be more likely to proceed to step 2 when the requisite demonstration of harm may be somewhat weaker than required if they appreciate that there is little benefit. Furthermore, as emphasized throughout this Article, the key decision thresholds tend to be stated in an obscure manner, a practice that enables decisionmakers to deem them exceeded or not, as the case may be, without clearly violating the protocols' dictates. Again, this can be implicit or subconscious: for example, in a case with some harm and no benefit, the appropriate threshold for harm at step 1 may be perceived as less demanding than in cases in which the benefit appears to be large.

To the extent that such behavior occurs, structured decision procedures in action may be closer to unconstrained balancing than is recognized. Even so, the resulting lack of transparency interferes with accountability, communication of the law to primary actors, and critical assessment of legal rules and the system's operation. Poor guidance is provided for agency investigations and court proceedings — regarding both information gathering and decisionmaking — to the extent that official protocols misstate best practices and misrepresent actual behavior. This state of affairs is conducive neither to optimal conduct in a given case nor to the development of better approaches for future cases. Nevertheless, due to queasiness about balancing regarding both quanti-

---

---

fication and commensurability, some may regard the ambiguity of structured protocols and the associated ease of dissembling as a feature rather than a bug.

Structured decision procedures are a focus of advocacy and a characteristic of actual doctrine in important areas of law. Why it is that such protocols are advanced in some legal domains and not others poses a central and largely unexamined question regarding fundamental aspects of legal design. This Article explores how structured decision procedures, if taken seriously, lead to systematically inferior liability decisions and offer surprisingly poor prescriptions for the gathering of information in individual cases. This investigation is conceptual, seeking to offer a systematic framework for identifying and thinking clearly about the pertinent issues, which should prove helpful in interpreting and applying existing legal rules as well as in determining how both substantive rules and legal procedures might best be reformed.

The general properties of structured decision procedures elaborated in this Article are very much evident in antitrust law (rule of reason, mergers), discrimination law (Title VII disparate impact), and constitutional law (strict scrutiny, proportionality analysis). A sequel to this Article shows how the template developed here greatly illuminates positive and normative dimensions in each of these legal domains.<sup>153</sup> This is true despite the fact that the stated rules in each application differ to varying degrees from each other and from the stylized structured decision procedure that is analyzed here. Most existing rules are murky on the key decision thresholds at step 1 and step 2 and on the role played by less restrictive alternatives. In light of these core ambiguities and the concomitant opportunities for circumvention, as well as the poor performance of structured rules when followed faithfully, it is unclear the extent to which structured decision procedures in operation actually deviate from unconstrained balancing.

For concreteness, it is helpful to consider some highlights from the sequel. Beginning with antitrust law, the Supreme Court's canonical statements over the course of a century of what many regard as the central doctrine, the rule of reason, present it as a pure balancing test. Yet this rule is increasingly restated as a structured procedure that resembles the three-step stylization examined here. Both of the protocol's decision thresholds are unspecified (including whether the harm at step 1 must be "substantial," whatever that may mean), the sequential separation of harm and benefit is confusing (particularly since anticompetitive acts are often defined as those other than "competition on the merits"), the analysis of less restrictive alternatives is perplexing (both as to what is required and how it fits in with the rest of the protocol), and the actual conduct of litigation does not adhere to the information-gathering prescriptions, despite some claims to the contrary.

---

<sup>153</sup> See Kaplow, *supra* note 1.



In government agencies' and courts' review of mergers, a similar structure has emerged, although not articulated as such. Even though most nontrivial horizontal mergers a priori generate some upward pricing pressure, few mergers are challenged and some challenges fail in court, and this is so (according to conventional lore) without reaching the question of whether there are offsetting efficiencies (akin to step 2). This suggests a distinct step 1 with a high threshold. On those occasions in which this hurdle is overcome, government guidelines and practice then move to step 2 but in turn find that it usually fails, suggesting a significant threshold there as well, particularly since it is commonly accepted that many mergers are motivated by greater efficiency (which is the rationale for prohibiting so few, on account of step 1). Once placed in this Article's framework, the analysis of merger challenges can be better understood, criticized, and improved. It seems likely, however, that regarding both information collection and decisionmaking, agencies' internal analyses reflect unconstrained balancing more closely than their official pronouncements suggest.

Disparate impact cases under Title VII are, under prevailing Supreme Court precedent and the 1991 Amendments to the Civil Rights Act, governed by a structured decision procedure that bears some resemblance to that examined here. The first step of disparate impact inquiries, which focuses on the plaintiff's prima facie case (typically proved with statistics), indeed suffers greatly from the problem of attempting to separate harm from justification, in a way that few have recognized, and the height and nature of the initial threshold is somewhat obscure. The second step (in some respects collapsed with the third), regarding the defendant's business justification, is understood to raise a serious question regarding the magnitude of the threshold: specifically whether it is essentially zero ("job related" is taken to mean that any productivity benefit is sufficient), massive ("business necessity" is the operative phrase, and "necessity" means necessity!), or somewhere in between (perhaps to a degree that is akin to balancing). These steps, the role of alternative employment practices (the analogue to less restrictive alternatives), queasiness about balancing in this context, and other matters are all illuminated by this Article's framework.

In constitutional law, strict scrutiny and proportionality analysis have structures similar to the stylized protocol examined here. Central questions about strict scrutiny concern, not surprisingly, the two thresholds: their height and their nature. The first threshold is regarded to be qualitative and categorical, rather than quantitative, essentially limiting the domain of review. As suggested in subsection IV.C.2, it sometimes makes sense, particularly with regard to constitutional provisions, to employ constraints on balancing, here by the legislative and executive branches, and to focus judicial review on realms in which other governmental actors cannot be trusted. Once step 1 is triggered, step 2's requirement of a compelling state interest seems to entail a threshold that

---

---

mixes qualitative and quantitative features in a manner that is difficult to interpret or justify: tiny impediments of qualifying interests perhaps suffice to justify substantial infringements, whereas huge costs of non-qualifying types may fail to justify modest infringements. Moreover, application of step 2 may implicitly require quantification of the infringement, which common understanding of step 1 indicates is unnecessary. Strict scrutiny omits step 3, but may in part introduce balancing through its step 2 as well as its less restrictive alternatives analysis, referred to as narrow tailoring. It is unclear the extent to which balancing by the court is actually avoided (rather than disguised), and any such deviations seem difficult to square with the purposes of review.

Proportionality analysis, employed in a number of jurisdictions and advocated by some for use in the United States, is often embodied in a multistep regimen that is even more suggestive of the structured decision procedures considered here. Yet the stance taken toward the height of the thresholds seems inconsistent. Some of the elucidation and rationalization proceed as though the thresholds are tough, which allows many cases to be disposed of before reaching proportionality review's final balancing step. If so, there are potentially significant costs and essentially no benefits (because the balances avoided are either easy or are ones that may well have favored an opposite conclusion). Other descriptions, seeming to sense the problem that early truncation may involve erroneous outcomes based on incomplete analysis, imply that the thresholds are instead negligible, in which event we essentially have unconstrained balancing that renders all but the final step moot. Matching these protocols against this Article's stylized procedure and accompanying analysis makes these inconsistencies more apparent and illuminates the costs if structured proportionality review substantially deviates from unconstrained balancing.

The sequel, like this Article, neither advances definitive claims regarding the actual state of doctrine nor suggests what doctrinal formulations would be best. Instead, it leverages the present Article's conceptual framework to better understand the operation and implications of these legal rules. Much confusion concerning the meaning of doctrine is clarified, questions that the law fails to answer or addresses in inconsistent ways are revealed, and significant but submerged effects of these rules are identified, all of which are necessary to appreciate what the law really is and to engage in intelligent debate about what the law should be.