How to Think About Errors, Costs, and Their Allocation

*By Ronald J. Allen**

Note from the Editor:

In December 2010, the Federalist Society heard from a number of federal judges and civil procedure experts about amendments to the Federal Rules of Civil Procedure, including the process that would be undertaken to amend the rules and some proposed amendments that might be offered. Based on the comments and perspectives received, the Federalist Society determined that it could add value to the broader discussion over amending the rules by asking experts to flag issues or perceived problems with the rules as they currently exist, and to identify the range of solutions that are being offered to address these problems. This back-and-forth culminated in four papers, one of which follows. A version of these papers will appear in the Florida Law Review, and they are published here with permission.

There is an ongoing robust debate about the structure of litigation in general, and in particular, about access to the courts. For a considerable period of time, the mantra that the courts should be readily available to all to present claims that their rights have been violated has dominated both academic discourse and perhaps significantly influenced the structure of litigation.[1] The conventional view that the courts should be freely open to all was dealt a blow by the *Iqbal*[2] and *Twombly*[3] decisions, which imposed greater gatekeeping responsibilities on the federal district courts. Predictably these decisions provoked a storm of protest, in large measure because they may indeed make it more difficult for many petitioners to have their petitions considered on the merits.[4] However, whether that result is a social harm or a positive good depends on matters in addition to simply winnowing the field of potential disputants, a point neglected by much of contemporary scholarship in civil procedure. That scholarship has had a laser-like focus on facilitating the bringing of claims, and in doing so makes two serious errors. It neglects that litigation is one small part of a larger social optimization problem, and has a peculiar conception of errors and costs and how they should be allocated. In this brief paper, I provide the analytical background to these assertions.

Primary and litigation behavior are conventionally conceived of as distinct spheres with internal logics of their own, the former articulating rules governing everyday actions, from social interaction to structuring efficient economic behavior and the latter governing that peculiar set of actions involved in litigation. Facilitating appropriate primary behavior is the overriding goal of social organization, and one of its main tools is the substantive law. Litigation behavior is the effort to resolve disputes about inappropriate primary behavior or to reestablish the status quo following disruptions of the social fabric.

Resources devoted to litigation appear to most legal commentators as wasted resources, adding no value to society. Since litigation itself does not produce useful good, litigation should obtain to correct results as efficiently as possible. These aspirations are reflected in Rule 1 of the Federal Rules of Civil Procedure that the rules of civil procedure "should be construed and administered to secure the just, speedy, and inexpensive determination of every action and proceeding," and Rule 2 of the Federal Rules of Evidence: "These rules shall be construed to secure fairness in administration, elimination of unjustifiable expense and delay, promotion of growth and development of the law of evidence to the end that the truth may be ascertained and proceedings justly determined." The principle animating these provisions is access to justice, in particular the principle that even the indigent should be not disadvantaged when the adversary is wealthy. Regrettably, life, as always, is complicated. Costs cannot be eliminated, and thus the most important question is their allocation.

Primary behavior does not produce goods cost-free. There are both waste products produced and the risk of harm to others. Lowering the cost of the production of an item encourages its production, as raising the cost has the opposite effect. Consequently, if the producer can externalize some of its cost (dumping waste in the river or on a neighbor's property), the cost of the good will not reflect its true social cost, which means that there will likely be over-production of the good in question. By contrast, optimal production of social goods is facilitated by having them produced at their true social cost. This is why it is important for the substantive law to align costs with behavior.

Litigation costs are generally believed to be socially perverse, as they act as a tax on productive behavior. To some extent this is true, but a costless legal regime would stimulate the production of its product like any "manufacturer," and the result could be overproduction of this good, as well. Although this may appear counter-intuitive, remember that decisions must be made as to how to dispute—in simple terms whether to sue or negotiate. Everyone comes into contact with numerous instances in which this decision must be made. Perhaps a neighbor plays music too loudly or neglects to dispose of trash correctly. If litigation were costless, rather than negotiate, one could simply sue. The costs of litigation affect the manner in which people relate, and those effects can be beneficial or perverse. The costs of litigation, in short, may counter-intuitively produce social goods through the incentive effects they create for modes of disputing.

The precise policy prescriptions following a deeper understanding of the problem of social cost are ambiguous because they depend in part on the relative values of resolving different kinds of disputes in different ways. It may be sensible to nudge certain kinds of disputes toward formal dispute resolution and others away from it. Maybe commercial disputes differ from family disputes, and maybe discrete commercial

** John Henry Wigmore Professor of Law, Northwestern University; President, Board of Foreign Advisors, Evidence Law and Forensics Sciences Institute, China University of Political Science and Law.*

transactions differ from antitrust actions systematically. Life, in short, is complicated, and one of the tasks for the legal system is sorting out that complexity.

I believe the history of the federal rules of procedure and evidence at least implicitly reflects these analytical points; they were enacted in part to offset what were believed to be distorting aspects of the systems that they replaced.[5] The previous systems were believed to disadvantage plaintiffs by raising their costs much too high. The solution to this was to simplify pleading requirements and allow cases to get on to what was believed to be low-cost discovery, followed by low-cost trials. Discovery costs would be low because the assumption was that knowledge of the typical cases was shared by both parties and thus a substantial investment in discovery would not be required. In addition, both parties would have the incentive to keep costs of discovery to their necessary minimum. It is immediately obvious how these conceptions map onto the previous analytical points. In a world of symmetrical information and low transaction costs, the federal rules perhaps accomplished the goal of facilitating the accurate and efficient resolution of disputes without distorting the underlying substantive law, values that the procedural regime the federal rules replaced did not adequately secure. If the original assumptions about litigation are true, procedural wrangling serves no purpose. Moreover, costs were not and could not be lowered to zero, so there remained reciprocal incentives to avoid litigation through other means of resolving disputes.

Note the historical contingency of the era that adopted the Federal Rules of Civil Procedure. It involved substantive assumptions about the relative positions of plaintiffs and defendants that were empirically true but not logically entailed. Thus, changes in the relative positions of plaintiff and defendant from the pre-Rules situation may justify changes in the procedural context, which could entail among other things a reallocation of costs. Perhaps originally the procedural regime favored defendants and thus subsidized socially wasteful activity; perhaps now in some set of cases it favors plaintiffs with the opposite effect. In such cases, defendants will be deterred from productive activities not by the law but by litigation costs that increase the *in terrorem* value of even meritless suits that put pressure on a defendant to settle and burden otherwise lawful conduct. Potential defendants will engage in litigation avoidance tactics that are likely to be socially wasteful, and they will settle to avoid litigation costs rather than risk liability on the merits. This increases the cost of socially useful activity that cannot be distinguished from socially costly activity at an acceptable price through litigation. The alternative is to buy peace through settlements even though the underlying primary behavior is perfectly acceptable. The effect is a tax on useful behavior.

To generalize, the interactive effects of primary and litigation behavior must be taken into account by the legal system. The effect or consequences of primary behavior on litigation behavior is often noted, but litigation behavior affects primary behavior as well. This means that the regulatory problem is unlikely to be solved by simple slogans such as those concerning access to court. Before addressing how to approach regulating such a complex problem, another issue involving the inadequacy of the conventional understandings of the litigation matrix needs to be addressed. In addition to inadequately considering the relationship between primary and litigation behavior, the conventional conception of an error is inadequate.

The conventional conception of an error is composed of two parts: denying a petitioner access to an adjudication on the merits (through narrowing the court house door)[6] and a belief that Type I (false positive) and Type II (false negative) errors are roughly equivalent and that the procedural goals should be to treat parties roughly equally and to minimize total errors.[7] Although these ways of thinking have been around for a considerable period of time, it is plain that they suffer from serious defects.

First, each time an undeserving litigant imposes costs on an adversary, an error has been made, a point that seems rather remarkably to have been neglected by those who have complained of the recent Supreme Court forays into procedural matters. The image of the federal or any other court system being constantly open and easily accessible for all neglects that a plaintiff walking through the courthouse door imposes costs on a defendant. If the defendant has behaved inappropriately by reference to the substantive law, these are costs the defendant should bear. But, as elaborated above, if the defendant has not behaved inappropriately by reference to that same substantive law—if a plaintiff's claim is unjustified, in other words—the costs imposed on defendants are errors that impose taxes on productive behavior, and thus likely socially perverse. The point is so obvious as to need little further elaboration. An undeserving plaintiff deprives a deserving defendant of its assets, and the best-case scenario is that the deserving defendant passes those costs on to a hapless public. The best-case scenario, in short, is decidedly unappealing. The point, of course, is that the conventional view seems dominated by the belief that there are no wrongful complaints filed, which is ludicrous. More importantly, in an era of asymmetrical costs, where filing a complaint can generate enormous costs on the part of the defendant, the defendant will be consistently in the position discussed above of having to minimize extra costs attached to socially useful behavior and passing whatever costs cannot be avoided on to someone else if possible.

There is a second fundamental error in the conventional thinking about errors. It focuses on just two of the decisions that can be reached at trial—an error for one side or the other—but there are four decisions that can be made at trial, and all have social benefits or costs. In addition to errors, correct decisions can be made. Neglecting correct decisions is peculiar. For example, in civil cases, the error equalization policy is satisfied by making errors in every single case, so long as the base rates of cases that go to trial include roughly the same number of deserving plaintiffs and defendants.

The relationship between the four possible outcomes at trial and procedural regulation is itself more complicated than it appears on the surface. In general, without knowledge of the base rates of deserving parties that go to trial and the relationship between the assessments of fact finders and true states of affairs, there is literally no way to predict the effect of procedural regulation on correct or incorrect decisions. For example, implicit in the conventional discourse is that a finding

that the probability of liability is .8 means that in eight out of ten similar cases, the true facts are consistent with liability. However, there could be any relationship at all between fact finders' findings of probability and true states of affairs. In the set of all cases where fact finders find there to be a .8 probability of liability, it could be true that all cases in that subset are cases where no liability should be found. Similarly, if everyone who goes to trial is guilty or liable, there can be no convictions of the innocent or mistakes against deserving plaintiffs, no matter how low the standard of proof, and vice versa.

The conventional discourse on procedural regulation also assumes a static system, whereas in fact it is dynamic. One aspect of this dynamism is that parties decide which cases to take further into the procedural system, and can adjust their decision in light of changes in the rules. Thus, the simple assumption that changing the burden of pleading or persuasion, or whatever, causes more errors of one kind than another, or any other suggested cause and effect relationship between regulations and outcomes, is obviously not analytically true; it depends on how the system responds to the change.

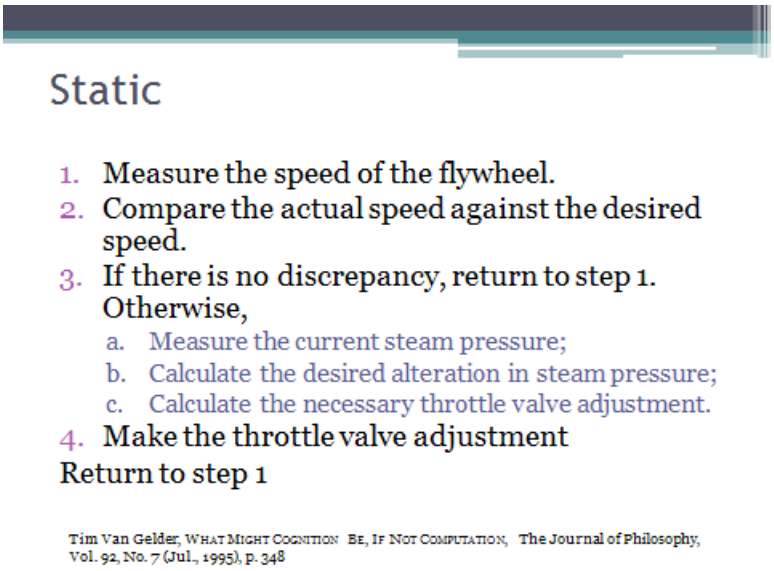
The combined effect of the neglect of the interactive relationship between primary and litigation behavior and the curious conception of an error is obvious. The result is to obscure that trial decisions are only one part of the output of the legal system. Parties negotiate outcomes in both civil and criminal cases. They do so in the shadow of trials, among other things, but the outcomes in those cases are part of the total social welfare effects of a legal system. In addition, those decisions are made in a dynamic not static environment, which leads to the question how to most effectively regulate such complex processes.

In the abstract, I think the answer is clear. How to translate the abstractions into feasible regulation is another matter. First, the abstract answer is addressed in the quote below from my recent Meador Lecture, which is followed by my further reflections on social optimization of the procedural system:

> [T]he reality of the legal system is not nice, tidy, simple, and static, contexts but instead bubbling cauldrons of messy, complicated, organic, evolutionary processes. The standard tool used to regulate this bubbling mess is rules, and it is the friction between that tool and many of the uses to which it is put that explains in general why fact finding and legal regulation are viewed as so often problematic. This same relationship is explanatory of many legal puzzles, such as, in ascending order of importance, the curious implications of standard legal error analysis, the rules v. standards debate, and the meaning of "law."

The simple concept of a rule as setting necessary or sufficient conditions from which outcomes may be deduced is an example of monotonic logic, in which the addition of postulates or assumptions simply adds to what may be deduced from the previous assumptions. Monotonic logics are powerful tools, as the rise of modern mathematics and the success of many scientific fields demonstrate. They work best when their operant assumptions accurately capture their domains, which means they work quite well, in Hayek's famous dichotomy, in made systems such as games and less well in grown or organic systems, which typifies much of the human condition.[8] A large part of debate over rules and their limits is often implicitly about the complexity of the relevant domain and one's tolerance for mistakes of different kinds. As the number of pertinent variables increases or when some of them are continuous rather than discrete, the deductive problem quickly becomes computationally intractable, even for computers let alone humans. And of course if a new variable pops up that was not previously anticipated, all deductive bets are off, as it were. In either case (computational intractability or failure of imagination), algorithmic approaches that rely on extant rules generate the standard critiques of the indeterminate nature of rules. In reality, it is not that rules are indeterminate but that they are being put to a task for which they are not optimal.[9]

I suggested in that lecture that the central problem of the legal system is similar to the central problem of rationality, which is the taming of complexity. In both cases, simple deductive tools were being put to uses that were suboptimal. That raises the important question what other approaches may be more fruitful. Inspired by a brilliant article by an artificial intelligence researcher, Tim Van Gelder, that I came across many years ago, one possible answer I ventured was that the struggle of rationality to tame complexity may be less like digital computation and more akin to a dynamic regulator, such as the Watt Centrifugal Governor that was a critical part of the industrial revolution.[10] Analogously, legal analysis may need to evolve to deal with the complexities of systems. Van Gelder's example is a metaphor rather than an argument for my purposes, for it provides just the suggestion of possibilities rather than a defined research program, but it is nonetheless interesting.



Here is the problem the dynamic regulator solved. The growth of the textile industry in England depended upon a consistent energy source with very limited variability. The steam engine provided the energy, but its pistons provided episodic bursts of energy rather than a smooth, continuous stream. Fly wheels were helpful, but still not adequate. As Van Gelder pointed out, one potential solution to this problem is computational (see the figure at the bottom of the previous page).

Unfortunately, this computational solution requires a costly person doing it, and it will rarely produce a smooth enough source of energy. James Watt solved this problem by placing movable arms on a spindle at the center of the flywheel, whose motion was transmitted instantaneously to the valve regulating the flow of steam. As the rotation of the fly wheel speeds up, the arms extend, which transmits to the valve and closes it until the proper equilibrium is reached, and vice versa (see the figure at the top of this page).

Dynamic

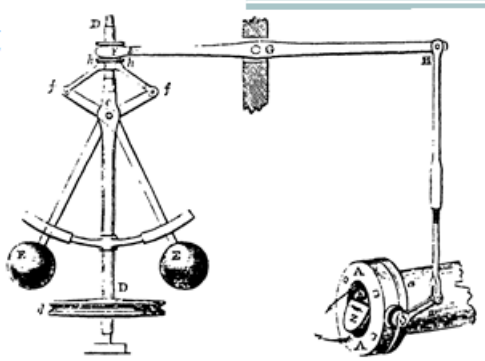


Figure 1[5]

[5] The Watt centrifugal governor for controlling the speed of a steam engine—from J. Farey, *A Treatise on the Steam Engine: Historical, Practical, and Descriptive* (London: Longman, Rees, Orme, Brown, and Green, 1827).

Tim Van Gelder, What Might Cognition Be, If Not Computation, The Journal of Philosophy, Vol. 92, No. 7 (Jul., 1995), p. 349

Regardless whether the centrifugal regulator captures something important about rationality, viewing the legal system with this metaphor in mind may be fruitful. The most dramatic point is that some problems can be solved other than through deductive arguments or simple rules; the contrary belief is a consistent constraint on legal scholarship generally. It is undoubtedly useful to break problems down into smaller parts, and so on, but at the same time that process can be counterproductive, disguising rather than highlighting the nature of the entity under examination. The alternative is to think of the legal system more, perhaps, like fluid dynamics treats the flow of liquids and gases, to embrace, in other words, the messiness of real life rather than abstract it away.

How does this apply in the procedural context? Telling trial judges to behave as centrifugal regulators in order to optimize social productivity is probably not likely to yield satisfactory results. The second-best solution would be to assign the true costs of parties' actions to them. However, it is impossible to determine, practically and maybe theoretically, the "true" costs of litigation behavior. For example, when I ask for discovery, I may be trying to build my case or respond to the opponent's case. I should be responsible for building my case, but responding to my opponent's case perhaps is a cost that he should bear. When a lawyer cross-examines, whose costs are those? If it is pointing out the limits of the adversary's case, he should bear those costs; but if through cross-examination I am building my case, I should bear those costs. How could these different effects be sorted out into the categories of useful for one side or the other? A crude rule—opposing party pays for my costs of cross-examination—leads to potential manipulation. Nor is adopting a British-style loser pays system an obvious solution. Recent empirical work shows both that simple predictions about the effect of a "loser pays" system are likely false (can increase transaction costs), and people do not opt for the English rule in contract negotiations.[11]

Alternatively, the objective could be to structure the process so that the parties have the incentive to properly allocate costs, with when necessary the involvement of the trial judge. That would involve categorical cost allocation, with the possibility of relief from the trial judge. One category probably ripe for such treatment is discovery costs. Discovery costs generally benefit the party asking for the discovery, and also have been a cause of considerable injustice because of their increasingly asymmetric allocation. Plaintiffs simply by filing can impose enormous costs on defendants while bearing virtually none themselves. Note how far from the original conceptions giving rise to the Federal Rules of Civil Procedure the modern condition may be. If each side will have about the same amount of discovery costs, it makes perfect sense to let each side bear their own costs. That is identical to cost shifting, and any resources spent in shifting costs are simply wasted. Asymmetric costs, by contrast, cause skewed cost allocation and provide the opportunity for strategic exploitation. By contrast, placing the costs of discovery provisionally on the person asking for it, but allowing for judicial involvement to make adjustments, may both generally give incentives for the optimal production of information and permit a safety valve in the unusual case.

Although the possibilities are diverse, an example of an "unusual" case would be where there is good reason to believe that an adversary is acting strategically primarily in order to impose costs. In such a case, the "benefit" is to the adversary, and that is who should bear the costs. That would be accomplished by petitioning the trial judge for relief. In making such determinations, the judge's decision will be constructed by the adversarial process, and the parties will have the correct incentives to educate the trial judge. That is not a guarantee of perfection, but it provides some hope for reasonable outcomes. It exploits the advantages of both an initial "bear your own costs" scheme with the apparent inertia of trial courts that do not want to get involved with cost allocation or discovery regulation unless forced to. They would be forced to only when the situation was egregious enough to justify a well-grounded petition for relief.

Endnotes

1 *See* Arthur Miller, *From* Conley *to* Twombly *to* Iqbal*: A Double Play on the Federal Rules of Civil Procedure*, 60 Duke L. Rev. 1 (2010).

2 Ashcroft v. Iqbal, 129 S. Ct. 1937 (2009).

3 Bell Atl. Corp. v. Twombly, 550 U.S. 544 (2007).

4 See, e.g., the symposium on the cases in the Penn. St. Law Rev. ***

5 *See* Ronald J. Allen & Alan E. Guy, *Conley* as a Special Case of *Twombly* and *Iqbal*: Exploring the Intersection of Evidence, Procedure, and the Nature of Rules, *** Penn. St. L. Rev. ***.

6 *See* Miller, *supra* note 1.

7 *See, e.g.*, David Kaye, *The Limits of the Preponderance of the Evidence Standard: Justifiably Naked Statistical Evidence and Multiple Causation*, 1982 A.B.F.J. 487.

8 Friedrich A. Hayek, Law, Legislation, Liberty: Rules and Order 35-54 (1973).

9 Ronald J. Allen, Rationality and the Taming of Complexity, ** Ala. L. Rev. ***

10 Tim Van Gelder, *What Might Cognition Be, If Not Computation*, 92 J. Phil. 348 (1995).

11 Kong-Pin Chen & Jue-Shyan Wang, *Fee-Shifting Rules in Litigation with Contingency Fees*, 23 J.L. Econ. & Org. 519 (2007); Keith N. Hylton, *Rule 68, the Modified British Rule, and Civil Litigation Reform*, 1 Mich. L. & Pol. Rev. 73 (1996); Theodore Eisenberg & Geoffrey P. Miller, The English vs. the American Rule on Attorneys Fees: An Empirical Study of Attorney Fee Clauses in Publicly-Held Companies' Contracts, *available at* http://ssrn.com/abstract=1706054.