A model of trials and causal stories

May 16, 2005

Abstract

This paper models a situation where a judge is advised by two opposing advocates, such as a legal trial. Simple models of decisionmaking via advocates predict that all relevant information will be presented to the judge, yet trial lawyers generally omit a great deal of potentially apropos information. This paper reconciles the theory of rational decisionmaking with the observed behavior via a model in which the judge weighs not only facts but causal stories. Two judges in the model may be perfect Bayesians, share the same priors, and still rationally arrive at different verdicts for the same trial. This variance in opinion may lead to partial information by the advocates.

1 Introduction

This paper presents a model of a trial. Two advocates present arguments to a decision maker,[1] who then decides which arguments to accept, and determines a final verdict.

What arguments will the advocates make, and how can we expect the decision maker to evaluate them? To answer these questions, I present a model of causal stories and their evaluation. The model includes both events of the usual form (‘C occurred’), and events of the form ‘C causes E’. The data-oriented models, of which the literature is primarily composed, are a special case of this model. The additional embellishments beyond the data-oriented model are not necessarily ‘irrational’ or proxies for cognitive limitations; instead, they address questions of the perception of causality, which lie outside the framework of statistics.

[1] In the U.S. court system, there are two types of decision maker, who have different tasks: jurors and judges. For the purposes of this paper, there is no benefit to making a distinction between the two, so I will use the terms ‘decision maker’ or ‘judge’, even though the person involved is sometimes an actual judge and is sometimes a juror.

For comparison, a strictly data-oriented model works as follows: the state of the world is described by a list of facts, and advocates simply provide information so that the decision maker can update her prior beliefs about the veracity of the facts to more accurate posterior beliefs. After taking factual statements from the advocates and doing the appropriate updating, she decides upon the most likely state of the world based only on her own posterior beliefs about the facts. The prediction of the data-oriented model is simple, theoretically robust, and basically implausible: opposing advocates will always lead the judge to the full-information verdict. A full discussion is in Section 3.3.

The prediction of the causal model is that each advocate will be much more selective in deciding what information he will provide. It is possible that there are apropos stories which neither party will want to tell.

Although partial information revelation may seem intuitively obvious or easily explained, the literature suffers from a paucity of models based on a structure which allows partial information revelation in equilibrium. For example, the conditions needed for full information revelation discussed in Section 3.3 below allow for decision makers who have short memories or use randomization devices.

The primary difference between the model presented here and the data-oriented models is that it is often the case that two different explanations of an event are simultaneously sufficient. For example, 'She stabbed him because she had hated him for months', and 'she stabbed him because she was drunk and out of control' are two different stories which imply two different penalties. So what does the judge do if both stories are proven to be true? Probability will not help her, because she already knows the state of the world (the defendant was both drunk and had hated him for months), so the final verdict is left to the judge's opinion. It is possible that she will base her opinion on who presented the argument; if she does so, then there may exist stories that will be detrimental to the presenter, regardless of whether the plaintiff or defendant presents.

Normatively, there is a significant difference between these two models: under the data-oriented model, all relevant information is always fully revealed, so the decision maker is in the ideal position to make a decision; but in a causally oriented model, the decision maker may easily reach a different decision from the one she would make given full information.

Applications The model here may be applied to any situation where two opposing advocates attempt to convince a decision maker that theirs is the correct viewpoint. This includes advertising, where two vendors vie for a purchase from one consumer; or lobbying, where lobbyists at either extreme argue that a legislator should vote in favor of their side.

Also, this model may generally tell us more about how people make decisions. Since it allows for the separation of beliefs about facts from opinions about motives or causes, it facilitates study of the properties of those opinions.

Outline of the paper Section 2 gives a review of the literature of causality, from the perspective of statisticians, psychologists, and legal philosophers. I discuss the advice given by the texts on legal argument, which articulate the importance of presenting parsimonious arguments in trial. Section 3 proves that under the simplest of assumptions, full information is always revealed by two advocates in a data-based framework. Section 4 gives an overview of an alternate story-based model and its elements, including causal weights. Section 5 then asks which stories the advocates will choose to reveal, and gives testable comparative statics describing when full information is more or less likely to be revealed.

2 Background

This section will discuss the prior work on causality, and then present a few excerpts from some of the law textbooks which taught argument to many of the lawyers this paper hopes to describe.

2.1 Prior work on causality

2.1.1 Statisticians

The statistician and the judge are in different positions. The statistician must determine which events may be counted as causes of the event in question. In so doing, he faces a number of paradoxes and problems which have been cataloged by previous authors; Thalos [2002] gives a good overview. But the decision maker here has the causal stories handed to her by the advocates, and need only evaluate them.

Another significant difference is that the statistician has an abundance of data with which to calculate correlations. Granger's concept of causality [Granger, 1969], also known as 'screening off', is the basis for the statistical causality literature (e.g., Perl [2000]). It is oriented toward statistical prediction: if C reduces the variance of our best predictor of E, then we class C as a cause of E and can measure its efficacy as such. Granger, with his statistician's wit, begins his exposition on causality with: "Let U_t be all the information in the universe accumulated since time $t-1$" [Granger, 1969, p 428]. Such a quantity of evidence will not be brought to the courtroom.

Finally, the *perception* of causality is not causality. There are philosophers and statisticians who claim that causality does not exist, or that it is restricted to a few well-defined cases. But the word 'because' is fundamental to language, because we perceive causality in myriad cases and base our thinking upon these observations.

For all of these reasons, the model presented here focuses on the perception of causality and the associated subjective probabilities, and is thus oriented toward behavioral prediction instead of statistical prediction.

2.1.2 Psychologists

The literature on stories as perceived by jurors is primarily empirical. My claim that people process information via stories exactly matches that made by the Story Model of Pennington and Hastie [Pennington and Hastie, 1986, 1988, 1992]. Their model of a story includes ingredients such as goals, initiating events, actions, and consequences. The model here is more abstract and discusses only sets of causes. As with any abstraction, this allows more general analysis and application to many situations, at the cost of being unable to make some of the more specific statements that derive from the more detailed Story Model.

2.1.3 Lawyers

Legal philosophers are familiar with the idea of separating necessary causes from sufficient causes, e.g. Honoré [1999]. But as a qualitative field, legal philosophy stops with these classifications. This paper extends the idea by placing quantifiable causal weights on individual causes. This also allows us to specify exactly which part of a ruling is the subjective evaluation of probabilities (which can be audited by outside observers) and which part is pure opinion (which can not be audited); and allows for empirical tests, in the lab and from actual cases.

2.2 Stylized facts about trials

This section will present some of the facts we observe in law as practiced in the United States, which the model will use or explain.

2.2.1 Appellate lawyers exist

We commonly say that a judge's role is to determine the truth, but this is only approximately the case. The judge's role is to weigh the cases presented by both sides and select which she deems to be the best one.

The evaluation of stories explains why lawyers before the appellate and Supreme courts exist. The judges are former lawyers and therefore well-versed in the minutiæ of legal procedure, and the cases are full information throughout, since findings of fact in a lower court may not be disputed. That is, the trial is not about unearthing new evidence, but about processing existing evidence. To a truth-seeking judge, the lawyers are redundant: a clerk could photocopy the appropriate papers for the judge at a fraction of the cost of hiring two lawyers and staging a trial. The Bayesian/Platonic appellate judge can only explain her need for opposing lawyers and a trial as a show of fairness, tradition, or public record. But the appellate judge whose job is to weigh two sides of a case and select the more compelling of the two tautologically relies upon lawyers to provide the arguments she will weigh.

2.2.2 The story told by the law texts

For a look at how lawyers argue, we can look at two textbooks for trial lawyers which commonly appear on U.S. law school reading lists: Mauet [2002] and Murray [1995].

Mauet is clear on advising the law student to become a storyteller: "Trials involve much more than merely introducing a set of facts; those facts must be organized and presented as part of a memorable story." [p 27] Nor must this story be burdened by too much detail: "A theory of the case is a clear, simple story of 'what really happened' from your point of view [...] If you can not state your theory of the case in a minute or two, it needs more work." [p 24]

Murray [p 9] corroborates this testimony:

> The trial lawyer's task can be analogized to the role of raconteur at a social party or similar informal gathering. The storyteller entertains his audience by recreating in the consciousness of the listener an image of an event or episode from another time and place.

A key aspect of storytelling is selecting which facts will be used to tell the story. Both authors cited above advocate against throwing everything possible at the judge or jury: "Do not overprove your case", Mauet advises [p 516]. Similarly from Murray [p 9]: "Even the simplest and shortest of events contains a myriad of potential details. [...] The trial lawyer must rigorously select those details that are essential to the image to be presented and must find efficient ways by which to convey them."

2.2.3 Discovery

This advice comes into especially stark contrast when we consider the discovery process, which precedes the trial itself. No facts or arguments may be presented in trial unless they appear in the briefs filed before the trial. The custom is therefore for the advocate to list every conceivable item in the briefs, to avoid the possibility that he needs some fact in trial but is barred from using it because he failed to include it in his briefs. For the game theorists, this is fortunate because the universe of items which can be presented in trial is a well-defined, common knowledge set: items listed in the briefs. For the discussion here, it is noteworthy that this set of items is customarily as large as possible, often including multiple—and potentially contradictory—stories or legal categorizations, so Mauet's and Murray's advice that the advocate should pick one story and prove it with a small subset of the items listed in the briefs is especially striking.

3 Conditions for full information revelation

This section presents a very general corollary of the Minimax Theorem that shows that full information will be revealed in all of the models presented in the literature of Section 3.1. It provides contrast and motivation for the main results, which appear later in this paper.

3.1 Prior work on information provision

The literature describes two types of advocates, classified by the clarity of their motivations. The first type includes consulting firms or subordinates entrusted with data gathering, whose biases exist but are unknown to the decision maker. Such advocates are discussed in works by Crawford and Sobel [1982], by Dewatripont and Tirole [1999], and in the literature based on these papers. However, they are not discussed here.

This paper focuses on a second type of advocate, such as lobbyists or lawyers, whose bias is clearly known to all involved, and whose goal is focused on influencing the decision maker to implement their preferred outcome. Milgrom and Roberts [1986], Austin-Smith [1998], and Calvert [1985] model this type of advocate. Their informational models could be taken as a special case of the purely Bayesian model discussed here, although they use these structures for different applications and to derive different results.

Glazer and Rubenstein [2001] have a unique model of debate among two advocates with a clear bias. However, their model assumes a limit to the amount of information that an advocate may provide, making it inappropriate for answering the question of when full information will be revealed.

Finally, all of this literature is divergent from the mechanism design literature in the structure of information to be extracted. In most models for mechanism design, the facts sought are unverifiable information, such as one's desire for a public good, so people are able to lie about these facts. The setup in all of the models in this paper and those cited above includes an honesty rule of some sort, stipulating that advocates may remain silent, but they may not lie. The question is then only about how much information the advocates will be willing to reveal, not the accuracy of the information.

3.2 Definitions

The conditions required for full information revelation are few and simple. The first is that advocates are in opposition:

Definition 1 *Denote the mth outcome of a game as $D_m \in \mathcal{D}$, where $\mathcal{D}$ is any set of outcomes (or decisions). Let the utility of that outcome to player n be $U_n(D_m) \in \mathbb{R}$. A* strictly competitive game *is a two-player game where $U_1(D_1) > U_1(D_2)$ if and only if $U_2(D_2) > U_2(D_1)$.*

That is, if one player prefers D_1 to D_2, the other must prefer D_2 to D_1.

This is a formalization of the description of strictly competitive games given in, e.g., Luce and Raiffa [1957]. As they point out, any strictly competitive game can be transformed into a zero-sum game: replacing $U_2(D)$ with $-U_1(D)$ will not affect player two's decisions, since both utility functions order outcomes in the same manner.

The next condition reflects an interesting feature of full information: once one player reveals it, the strategy of the other player can be ignored and questions of game theory become irrelevant.

Definition 2 *A strategy is* unilateral *iff it may be chosen by any player under any conditions, the outcome is the same regardless of which player plays the unilateral strategy, and the outcome given that one player chooses the unilateral strategy does not depend on what the other player does. The outcome given the unilateral strategy is the* unilateral payoff*.*

For example, let $\mathcal{I} = \{i_1, i_2, \ldots, i_n\}$ be the set of available items of information. The strategy set for both advocates consists of all subsets of $\mathcal{I}$; let the group of information items chosen for presentation by player one be G_1 and that chosen by player two be G_2. These subsets may be the empty set or all of $\mathcal{I}$. Let their union be $\mathcal{G} \equiv G_1 \cup G_2$.

The decision maker has a decision function which maps from sets of information G into final decisions $D(G)$. For concreteness, take $D(G) \in \mathbb{R}$, and let the first advocate's utility be increasing in the value of $D(G)$, while the second advocate's utility is decreasing in $D(G)$. In the game here, the advocates present her with $\mathcal{G}$, so she decides $D(\mathcal{G})$.

The game is strictly competitive with respect to the advocates' payoffs from the final decision. $G_j = \mathcal{I}$ is a unilateral strategy: G_1 is defined to be a subset of $\mathcal{I}$, so if $G_2 = \mathcal{I}$, then $G_1 \cup G_2$ is the same ($\mathcal{I}$) regardless of what G_1 chooses, and similarly for $G_1 = \mathcal{I}$.

This is a model of gathering information in its purest form: the decision maker simply seeks a series of facts which she is unable to gather herself, due to restrictions on time, ability, or other resources. Once the information is handed to her, she is entirely unconcerned with the source of the information, and she independently arrives at her decision. It features a unilateral strategy and competing advocates.

3.3 The full information outcome

The outcome of the game described above will be determined by the following lemma, which is a simple corollary of the Minimax Theorem.

Lemma 1 *In a strictly competitive game with a unilateral strategy, players will always receive the unilateral payoff in equilibrium.*

Proof: Regardless of what player one does, player two can play the unilateral strategy, meaning that his optimal response will give a payoff as large or better than the unilateral payoff. The minimum among these maxima is the unilateral payoff. Similarly for the other player: the maximum among his minimum payoffs is the unilateral payoff. The minimax payoff for player one matches the maximin payoff for player two, and the game is strictly competitive. The Minimax Theorem (as in, e.g., Luce and Raiffa [1957]) tells us that the unilateral payoff is therefore always an equilibrium payoff.

Assume there exists another equilibrium with a different payoff. One player or the other will be doing worse than the unilateral payoff, since the game is strictly competitive. That player will defect to the unilateral strategy, meaning that the assumption of an equilibrium with a different payoff was false. $\diamondsuit$

Notice that the proof was in no way dependent on the example: it applies to *any* strictly competitive game with a unilateral strategy. Also notice that the theorem said nothing about how the decision maker goes about making a decision; in the context of the example, we did not need to make any assumptions about the function $D(G)$. The function may or may not include strategic considerations made by the decision maker beforehand, and makes no assumptions about how new information changes the final decision. For example, the decision process may include flipping coins, systematically forgetting or ignoring information, or systematically mis-processing information; the only requirement is that a unilateral strategy exists for the given decision rule.

By most definitions of the term, full information revelation is a unilateral strategy. If advocates are only partially revealing information, and one or the other could do better by revealing full information, then he always has the option to do so, and his opposition can do nothing to obfuscate facts after they have been brought to light. Therefore, as per the proof of Lemma 1, neither player will ever accept a payoff less than that he could get by revealing full information.

Nothing prevents another outcome from having a payoff which is the same as that for full information revelation. But full information is a focal equilibrium and seems more likely to be chosen than an alternate (possibly mixed) strategy whose payoff is the same. Therefore, although there is no theoretical guarantee that full information will generally be a unique equilibrium, it seems reasonable to presume that it is the outcome that will prevail if the conditions of Lemma 1 hold.[2]

4 The model

As discussed in Section 3.3, a model based purely on statistical calculation derives equilibrium advocate behavior very different from that of the stable state of actual advocate behavior described in Section 2.2.2. Therefore, this section proposes a model which adds causal machinery to the estimation of subjective probabilities, resulting in theoretical equilibria that match observed behavior much more closely.

This section will first introduce the individual components of the model: stories, sufficient causes, the judge's beliefs about these components, causal weights, and the verdict function.

4.1 A timeline of the trial model

Here is a brief outline of the trial process. All terms will be defined and discussed in the sequel.

There exists a universe of stories (the briefs) from which the advocates may select the stories they wish to present to the judge. The plaintiff and defendant simultaneously[3] present as many stories as they wish.

A story is *accepted* if the judge determines that the story is true and sufficient to explain E.

If a judge accepts none of the advocates' stories as true and sufficient, then she would either express her disappointment with the advocates informally, or call a mistrial. I will model this here by presuming that if the judge selects the state of the world where none of the stories are true, she re-draws another state from the same probability distribution. She will thus eventually settle on a positive number of true stories with probability one.

Having accepted some stories and rejected others, she places a weight on each of the stories

[2] If we were so inclined, we could use this as a definition of irrelevance: those facts which are not revealed in equilibrium do not change the outcome, and are therefore irrelevant.

[3] In an actual trial, the plaintiff goes first, then the defendant, then both sides are allowed a rebuttal in the same order. I feel that explicitly modeling the turn-taking would add little value over simply finding the equilibria of the simultaneous game here.

which she did accept, meaning that the judge feels that some stories were more important in explaining E, even though any one of them would be sufficient by itself. Having determined her opinion (expressed in the weights and defined formally in Definition 6 below), she uses that opinion to calculate the final verdict.

The plaintiff receives a payoff equal to the verdict, and the defendant receives a payoff equal to the negation of the verdict. The judge does not receive a payoff, meaning that she will base her verdict not on an expected utility calculation, but using a process which follows guidelines outlined below. The advocates are assumed to have uninformative prior beliefs about the judge's opinion; informative priors would temper the results below but not change their substance.

4.2 Objects

Let $\mathcal{S}$ be the set of stories. Stories are binary (either true or false), and each story $S \in \mathcal{S}$ has a probability $P(S)$ of being true.

Let E signify the event which brought the trial about, and which the advocates seek to explain.

Let $s(S)$ signify the statement 'S is sufficient for E to occur'. Such claims of sufficiency are objects, just like the stories. Let the space in which these objects lie be $\mathcal{SC}$, and $P(s(S))$ be the probability that the statement is accepted by the judge (which is not necessarily related to the veracity of S itself).

4.3 The judge's beliefs

The judge's beliefs will be divided into two portions: the likelihood of accepting objects (stories and causal statements) as true, and the weights assigned to accepted stories.

Definition 3 *Let $J : \mathcal{S} \times \mathcal{SC} \rightarrow [0, 1]$ map from sets of stories and sufficient causes to the judge's subjective belief that the elements of these sets are true.*

For notational simplicity, I will sometimes discuss the judge's beliefs over fewer than two dimensions. For example, her belief about the probability of an event S is represented by $J(S, \mathcal{SC})$. I will abbreviate this unduly cumbersome notation to $J(S)$, leaving it to be understood that a space which is not specified is unrestricted.

4.3.1 Acceptance

A judge *accepts* a story if she both believes that the story is true, which will occur with probability $J(S)$; and believes that the story is sufficient to explain the event E, which will occur with probability $J(s(S))$. It will be convenient to define a separate notation for this action, since acceptance plays so heavily in the sequel.

Definition 4 *A story S is* accepted *if it is judged to be true and to be a sufficient explanation for E. The probability of acceptance is $PA(S) \equiv J(S, s(S))$. Let $PA(\neg S)$*[4] *signify the probability of rejection, $1 - J(S, s(S))$.*

For example, if the judge believes that truth and sufficiency given truth are statistically independent, then

$$PA(S) = J(S) \cdot J(s(S)).$$

4.4 Weights on sufficient causation

Although many of the examples given here and elsewhere in the causal literature are factual in nature, the story told by a lawyer will consist of both facts about past events and legal arguments. A sample story: Ms. A was driving carefully but hit ice; driving errors fall under a negligence rule; negligence rules in driving have been applied in similar past cases.

Taking legal argument as an exercise in classification (as many legal philosophers do, e.g. Sunstein [1996]) opens up the possibility that multiple stories explain the same simple event, and are all correct even though they may each imply a different verdict. For example, there may be one statute that classifies driving under a negligence rule (so Ms. A is not liable), and another which classifies operating heavy machinery in inclement conditions under strict liability (so Ms. A is liable). Both of these stories may simultaneously be true and sufficient to explain the events under the law.

In the model, the means by which a judge processes multiple accepted causal stories is via a weighting scheme. The weight placed on any given story depends on which stories have been accepted as true and sufficient. Therefore, let $\mathcal{S}$ be the set of stories which have been accepted as both true and sufficient, and express the weight placed on S_1 when $\mathcal{S}$ is the set of true stories as $W(S_1, \mathcal{S})$. Since these are weights, they should be positive, sum to one, and be zero for all S not in $\mathcal{S}$. Formally:

Assumption 5 *If E is an event, and $\mathcal{S} = \{S_1, \ldots, S_n\}$, is the set of stories which have been accepted as true and sufficient to explain E, then for each $S_i \in \mathcal{S}$:*

[4] In this context, the symbol '¬' signifies negation, not the complement, so $P(\neg S_1)$ is the probability that S_1 is false, not the probability that $\{S_2, \ldots, S_n\}$ is true.

(i) $W(S_i, \mathcal{S}) \geq 0$, $\forall i$
(ii) $\sum_{i=1}^{n} W(S_i, \mathcal{S}) = 1$
(iii) $W(S_i, \mathcal{S}) = 0$, $\forall S_i \notin \mathcal{S}$.

As a special case of Assumption 5, if only one story is accepted as an explanation for E, then $W(S_1, \{S_1\}) = 1$.

With two stories, there are three possible states: $(S_1 \cap S_2)$, $(S_1 \cap \neg S_2)$, and $(\neg S_1 \cap S_2)$. The state $(\neg S_1 \cap \neg S_2)$ is not relevant to the judge's task. Each state implies a different weight placed on S_1 and S_2:

$$\begin{array}{rcl} \mathcal{S} = \{S_1\} & \Rightarrow & W(S_1, \mathcal{S}) = 1 \quad \& \quad W(S_2, \mathcal{S}) = 0 \\ \mathcal{S} = \{S_2\} & \Rightarrow & W(S_1, \mathcal{S}) = 0 \quad \& \quad W(S_2, \mathcal{S}) = 1 \\ \mathcal{S} = \{S_1, S_2\} & \Rightarrow & W(S_1, \mathcal{S}) + W(S_2, \mathcal{S}) = 1. \end{array}$$

There is nothing in the realm of statistics which will tell us how the judge will distribute the weights in the third case. Returning to an example from above, she may decide that 'she was drunk' was the primary cause, or she may decide that 'she had hated him for months' was the primary cause. That is, any allocation of $W(S_1, \mathcal{S})$ and $W(S_2, \mathcal{S})$ which follows the above assumption is consistent with the laws of statistics and rational information processing.

4.4.1 Opinion formally defined

The notation for weights is valuable because it isolates the allocation of causal weights from the calculation of subjective probabilities.

Definition 6 *A* judge's opinion *is a set of causal weights for any given set of stories. That is, for any given* $\mathcal{S} = \{S_1, \ldots, S_n\}$*, the judge's opinion has a set of weights* $W(S_1, \mathcal{S})$*, ...,* $W(S_n, \mathcal{S})$*.*

We can require consistency from the judge in terms of her evaluation of probabilities, and can accuse her of bias if her evaluation is too far from a more objective standard. But the opinion as defined here is indeed an opinion, and we can expect that it will vary widely from judge to judge, or even occasion to occasion.

4.4.2 Verdicts

Let each story have a value, $v(S) \in \mathbb{R}$, representing the verdict (in dollars paid by the plaintiff to the defendant) which would be given if the judge decided that S were the sole cause of E.

Given that $\mathcal{S} = \{S_1, \ldots, S_n\}$ are accepted as true and sufficient to cause E, the final verdict is assumed to be a linear combination of each story's value weighted by its causal weight:

$$V(\mathcal{S}) = \sum_{i=1}^{n} W(S_i, \mathcal{S}) \cdot v(S_i). \tag{1}$$

So given that there are a number of true stories which conclude with the event, the final verdict will be a linear combination of the individual verdicts.

If there is only one story which is accepted, then $W(S, \{S\}) = 1$, and the final verdict reduces to the definition of $v(\cdot)$: $V(\{S\}) = v(S)$.

The results which follow do not rely heavily on the linear form of the verdict function, but do depend on the assumption that the verdict is some sort of weighted aggregation of the verdicts for the accepted stories.

The expected outcome The *ex ante* expected outcome is the following sum over all the nonempty sets of stories S:

$$\sum_{\mathcal{S}_i \in 2^S} \left[\left(\prod_{S \in \mathcal{S}_i} PA(S) \right) \left(\sum_{S \in \mathcal{S}_i} v(S) W(S, \mathcal{S}_i) \right) \right]. \tag{2}$$

The product is the probability that all the elements of $\mathcal{S}$ have been accepted as true and an explanation of E; the sum is the weighted sum of the verdicts from Expression 1.

The defendant wants to minimize Expression 2, while the plaintiff wants to maximize it.

4.5 Equilibrium

Define an incremental change to be the addition or removal of any number of stories from the set of stories presented by an advocate. Define an equilibrium set of stories to be a set such that neither agent wishes to make an incremental change to it.

This is a non-restrictive definition of equilibrium, in that if we were to allow more drastic changes to be made, we may find fewer equilibria; this is comparable to finding local maxima instead of a global maximum. A more general analysis is not possible given the relatively sparse assumptions made here.

5 How many stories to present

This section reconciles Mauet and Murphy's rule that lawyers should parsimoniously present limited information with the above result that under very simple conditions, the full-information outcome prevails.

First, note that if the judge's opinion is appropriately restricted, then it will be irrelevant to the advocate's information revelation decision: the full-information outcome will prevail.

Lemma 2 *Let the set of stories which the advocates may choose to present consist of a fixed set of n stories: $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$. If the judge's opinion with respect to the subsets of $\mathcal{S}$ is fixed before the trial, then the final verdict will be the one that would be given if the advocates admitted all of $\mathcal{S}$.*

Proof: The model fits all of the requirements for the full information payoff given above in Section 3.3: the game among the lawyers is strictly competitive, as per Definition 1, and revealing all of $\mathcal{S}$ is a unilateral strategy, as per Definition 2. Apply Lemma 1.$\diamondsuit$

But if the judge arrives at her opinion at the end of the trial, it is possible that she will choose to base her opinion on who presents the story; full information revelation then becomes less likely. I will restrict attention to the case of three stories.

Theorem 3 *There are three stories which the advocates could submit: $\mathcal{S} = \{S_d, S_p, S_m\}$. Without suffering any loss of generality, let $v(S_d) \equiv 0$, $v(S_p) \equiv 1$, and $v(S_m) \in (0, 1)$.*

If the defendant has already presented S_d and the plaintiff has already presented S_p, and if the judge may base her opinion on who presented S_m, then there exist conditions and a judge's opinion such that the plaintiff would not present S_m in equilibrium; there exist conditions and a judge's opinion such that the defendant would not present S_m in equilibrium; and all of these conditions may simultaneously be true.

[Proofs not in the body of the text can be found in the appendix.]

The important point behind this theorem is that there are conditions where both sides could regret presenting S_m. If the plaintiff presents S_m, then the judge may select $W(S_p, \{S_d, S_m, S_p\})$ to be small, leaving moderate weight on S_d; while in the case where the defendant presents S_m, she may select $W(S_d, \{S_d, S_m, S_p\})$ to be small, leaving moderate weight on S_p. Therefore, we may easily construct cases where neither party wants to reveal S_m, and as a result the judge will hear only the extreme arguments.

5.1 An example: the aleatory judge

Let there be three stories: S_{10} leads to a ten-year sentence for the defendant, S_{20} leads to a twenty-year sentence, and S_{30} a thirty-year sentence.

The judge has the following aleatory[5] decision procedure: she first hears all cases by all sides, then decides which of the stories presented are acceptable explanations of the case. If only one is accepted, then that story's verdict is the trial's verdict. If only one advocate's stories are accepted, then she randomly selects among the stories presented by that advocate with equal probability. If both advocates have stories which are accepted, then she first randomly selects which advocate to listen to, then randomly selects a story among those he presented.

Why is this plausible? It takes into account that the judge has a rational means of determining the state of the world (that is, which stories among those presented are true). But once she has determined the state of the world, and has exhausted all information about what happened and which legal rules are apropos, she may still not have a verdict with certainty. From there, she may then base her decision on whether she finds the plaintiff or defendant to be more worthy of agreement, rather than deciding solely based on the arguments themselves, all of which she may determine to have equal merit.

Assume that all three stories, if presented, will be accepted with certainty. Clearly, the defense would prefer to present S_{10} over the other two stories or no story at all; the prosecution prefers S_{30} over the other options. If the judge accepts both stories, then she will flip a coin to choose whom to listen to, and the expected outcome to this situation is a $30/2+10/2 =$20 year sentence.

Would the defense present S_{20}? If he does so, and all three stories are accepted, then the judge will first flip a coin to choose whether to listen to the plaintiff or defendant, and if she chooses the defendant, will then flip a coin to decide which of the defendant's stories to listen to. The expected verdict is $30/2 + (\frac{10}{2} + \frac{20}{2})/2 =$22.5 years, so the defense is better off not presenting S_{20}. Similarly, the expected verdict if the plaintiff presents S_{20} is $(\frac{30}{2} + \frac{20}{2})/2 + 10/2 =$17.5 years, so the plaintiff is also better off not presenting S_{20}. The more moderate story will not be told.

5.2 Comparative statics

Many of the limits given in Theorem 3 move as we would expect them to.

Corollary 4 *The range of values of $v(S_m)$ which will induce the plaintiff to always reveal S_m increases as $v(S_p)$ decreases, and as $v(S_d)$ grows.*

[5]The OED defines aleatory as "dependent on the throw of a die", citing a fragment from Urquhart's translation of Rabelais: "So continually fortunate in that aleatory way of deciding Law Debates".

If we assume that acceptance of S_m, S_p, and S_d are independent events, then as $PA(S_p)$ increases, the range of values of $v(S_m)$ which will induce the plaintiff to always reveal S_m shrinks. Similarly for the defense.

Assuming independence, there is no change in the plaintiff or defendant's range for revelation given a change in $PA(S_m)$.

So if the plaintiff has a stronger initial argument, either because it is more likely to be true or because it induces a more favorable verdict, then there will be fewer conditions under which the plaintiff will certainly want to contribute another story. If the defendant has a stronger argument (measured by verdict, not probability), then there are more conditions under which the plaintiff will certainly want to add another story.

The comparative statics—especially the claim that $PA(S_m)$ will not change the range of revelation—may easily be tested in the lab.

5.3 An alternate specification

One reader commented that assuming a mistrial in the case where no stories are true may not be appropriate; for example, the judge may rule for the defense in such a case. In this case, Theorem 3 reduces to the following:

Corollary 5 *Assume the same situation as Theorem 3, but the judge rules for the defense in the case when no story is accepted as true. Then there does not exist a situation where any defendant will reveal S_m. Any prosecutor will reveal S_m iff*

$$v(S_m) > \frac{J(\{S_p, S_m\})}{(1 - J(\{S_d, S_m\}))}.$$

6 Conclusion

The idea that one thing causes another can not be accurately described by statistics. We are all familiar with the problems: Christmas card sales do not cause Christmas; my watch reading 12:01 does not cause your watch to read 12:01. Although we as human beings understand the true causal relationships among the events, it is impossible to robustly write down the relationships using only joint probabilities.

A decision maker may be a perfect Bayesian with respect to her evaluation of questions of probabilities, but this would say nothing about how she goes about placing weights on different causes. As such, two 'perfectly rational' judges may arrive at different decisions given the same case—or one judge may arrive at different decisions given two cases which the statistician would describe as equivalent.

For the question of information revelation, this may create problems. It is possible that a judge would place different weight on a story depending on who presented it, and this makes it possible that neither side will want to present certain pieces of important information. One often sees debates where only the most extreme and divisive arguments are presented, with both sides omitting the moderate, middle-ground arguments.

The model here explains why moderate arguments may not be revealed by opposing advocates, and supports the perceptions of what makes a good argument as described by lawyers themselves—that it is free of extraneous detail and clearly tells one and only one story. The key is not in any irrational failure to process information, but in the human perception of causality.

7 Appendix: Proofs

Here are the proofs omitted from the text.

Theorem 3 *There are three stories which the advocates could submit:* $\mathcal{S} = \{S_d, S_p, S_m\}$. *Without suffering any loss of generality, let* $v(S_d) \equiv 0$, $v(S_p) \equiv 1$, *and* $v(S_m) \in (0, 1)$.

If the defendant has already presented S_d *and the plaintiff has already presented* S_p, *and if the judge may base her opinion on who presented* S_m, *then there exist conditions and a judge's opinion such that the plaintiff would not present* S_m *in equilibrium; there exist conditions and a judge's opinion such that the defendant would not present* S_m *in equilibrium; and all of these conditions may simultaneously be true.*

Proof: With only S_d and S_p, the expected verdict is:

$$V(\{S_d, S_p\}) = \frac{\alpha}{1 - PA(\{\neg S_d, \neg S_p\})}, \tag{3}$$

where

$$\alpha = PA(\{\neg S_d, S_p\}) + PA(\{S_d, S_p\})W(s(S_p, \{S_d, S_p\})).$$

This is the payoff given the three possible states in which one of the stories told by one of the advocates is true. Since $v(S_d) = 0$, the terms $PA(\{S_d, \neg S_p\}) \cdot 0 + PA(\{S_d, S_p\})W(s(S_d, \{S_d, S_p\})) \cdot 0$ have been omitted from α. Since the event $\{\neg S_d, \neg S_p\}$

would involve a retrial, we scale the verdict by $1 - PA(\{\neg S_d, \neg S_p\})$. [To delve into a bit more detail, the verdict will be

$$\begin{aligned}
&\alpha + PA(\{\neg S_d, \neg S_p\})[\alpha + PA(\{\neg S_d, \neg S_p\})[\alpha + PA(\{\neg S_d, \neg S_p\})[\ldots]]] \\
&= \alpha + PA(\{\neg S_d, \neg S_p\})\alpha + [PA(\{\neg S_d, \neg S_p\})]^2\alpha + \ldots \\
&= \alpha/(1 - PA(\{\neg S_d, \neg S_p\})),
\end{aligned}$$

as in Equation 3.]

If S_m is presented, then there are eight possible states. To simplify the math, I will assume that the weights $W(S_p, \{S_p, S_d\}) = W(S_p, \{S_p, S_d, \neg S_m\})$, and similarly for $\{\neg S_p, S_d\}$, and $\{S_p, \neg S_d\}$. Weakening this assumption will make the bounds found below more stringent. Given this assumption, the verdict given all three stories will be:

$$V(\{S_d, S_m, S_p\}) = \frac{\alpha' + \beta}{1 - PA(\{\neg S_d, \neg S_p, \neg S_m\})}, \tag{4}$$

where

$$\alpha' = PA(\{\neg S_d, S_p, \neg S_m\}) + PA(\{S_d, S_p, \neg S_m\})W(s(S_p, \{S_d, S_p\})), \text{ and}$$

$$\begin{aligned}
\beta = \quad & PA(\{S_d, S_p, S_m\}) \cdot [W(s(S_m, \{S_d, S_p, S_m\}))v(S_m) + W(s(S_p, \{S_d, S_p, S_m\}))] \\
& + PA(\{S_d, \neg S_p, S_m\}) \cdot W(s(S_m, \{S_d, S_m\}))v(S_m) \\
& + PA(\{\neg S_d, S_p, S_m\}) \cdot [W(s(S_m, \{S_p, S_m\}))v(S_m) + W(s(S_p, \{S_p, S_m\}))] \\
& + PA(\{\neg S_d, \neg S_p, S_m\}) \cdot v(S_m).
\end{aligned}$$

If S_m is presented but rejected, then α' is the expected verdict; if S_m is presented and accepted, then the expected verdict is β. For the plaintiff to decide whether he should present S_m, he need only calculate the expected payoff before S_m is presented ($EV(\{S_d, S_p\})$), as in Equation 3, and the expected payoff after S_m is presented($EV(\{S_d, S_m, S_p\})$), as in Equation 4. If $EV(\{S_d, S_m, S_p\}) > EV(\{S_d, S_p\})$, then the plaintiff would benefit from presenting S_m, and if the reverse is true, then he would not.

If we show that when Equation 3 is at a maximum and Equation 4 is at a minimum, the plaintiff still wants to present S_m, then we have proven that the plaintiff will always present S_m regardless of the judge's opinion. The expected payoff before presenting S_m is at a maximum when $W(s(S_p, \{S_d, S_p\})) = 1$–that is, when the judge places maximal weight on S_p in all cases where she has leeway to do so. In this case, α reduces to $\alpha_{max} = PA(S_p)$, and $\alpha'_{max} = P(\{S_p, \neg S_m\})$. Notice that since $v(S_p) = 1$, these are actually expected verdicts, e.g., $\alpha_{max} = PA(S_p)v(S_p)$.

After presenting S_m, the expected verdict is at a minimum when $W(s(S_d, \{S_d, S_m, S_p\})) = W(s(S_d, \{S_d, S_m\})) = 1$, and $W(s(S_m, \{S_m, S_p\})) = 1$—that is, when the judge places as much weight as possible on S_d, and, when S_d is not an option, on S_m. In this case, β reduces to $\beta_{min} = PA(\{\neg S_p, S_m\})v(S_m)$.

The plaintiff will therefore always present S_m when:

$$\begin{aligned} V_{min}(\{S_d, S_m, S_p\}) &> V_{max}(\{S_d, S_p\}) \\ \frac{PA(\{S_p,\neg S_m\})+PA(\{\neg S_d,S_m\})v(S_m)}{[1-PA(\{\neg S_d,\neg S_p,\neg S_m\})]} &> \frac{PA(S_p)}{[1-PA(\{\neg S_d,\neg S_p\})]} \\ v(S_m) &> \frac{PA(S_p)}{PA(\{\neg S_d,S_m\})}\frac{[1-PA(\{\neg S_d,\neg S_p,\neg S_m\})]}{[1-PA(\{\neg S_d,\neg S_p\})]} - \frac{PA(\{S_p,\neg S_m\})}{PA(\{\neg S_d,S_m\})}. \end{aligned} \tag{5}$$

When Inequality 5 does not hold, then there exist circumstances such that the plaintiff prefers that S_m not be revealed. Notice that the right-hand side of this constraint may easily be greater than one, in which case there is no $v(S_m)$ for which the plaintiff will always reveal S_m.

Now consider the defense: if the verdict without S_m is at a minimum, and the verdict with S_m is at a maximum, and the defense still wants to present S_m, then he will want to present S_m regardless of the judge's opinion.

Before presenting S_m, the expected verdict is at a minimum when $W(s(S_d, \{S_d, S_p\})) = 1$. Then $\alpha = PA(\{\neg S_d, S_p\})$. After presenting S_m, the expected verdict is at a maximum when $W(s(S_p, \{S_d, S_m, S_p\})) = W(s(S_p, \{S_m, S_p\})) = 1$, and $W(S_m, \{S_d, S_m\}) = 1$. In this case, $\alpha' = PA(\{\neg S_d, S_p, \neg S_m\})$ and $\beta = PA(\{S_p, S_m\}) + PA(\{\neg S_p, S_m\})v(S_m)$. We want to find when:

$$\begin{aligned} V_{max}(\{S_d, S_m, S_p\}) &< V_{min}(\{S_d, S_p\}), \\ &\text{or} \\ \frac{PA(\{\neg S_d,S_p,\neg S_m\})+PA(\{S_p,S_m\})+PA(\{\neg S_p,S_m\})v(S_m)]}{[1-PA(\{\neg S_d,\neg S_p,\neg S_m\})]} &< \frac{PA(\{\neg S_d,S_p\})}{[1-PA(\{\neg S_d,\neg S_p\})]}. \\ v(S_m) &< \frac{PA(\{\neg S_d,S_p\})}{[1-PA(\{\neg S_d,\neg S_p\})]} \cdot \frac{[1-PA(\{\neg S_d,\neg S_p,\neg S_m\})]}{PA(\{\neg S_p,S_m\})} \\ &\quad - \frac{PA(\{\neg S_d,S_p,\neg S_m\})+PA(\{S_p,S_m\})}{PA(\{\neg S_p,S_m\})}. \end{aligned} \tag{6}$$

Under reasonable assumptions, neither Constraints 5 nor 6 will hold, meaning that the prosecution can envision a judge's opinion which will make presenting S_m undesirable, and the defense can do the same. For example, say that $P(S_m)$, $P(S_d)$, and $P(S_p)$ are statistically independent, and $P(S_m) = P(S_d) = P(S_p) = \frac{1}{2}$. Then Constraint 5 reduces to $v(S_m) > \frac{4}{3}$, and Constraint 6 reduces to $v(S_m) < -\frac{1}{3}$. Neither of these will be true given $v(S_m) \in (0, 1)$, so neither party is guaranteed that presenting S_m is beneficial regardless of the judge's opinion. $\diamondsuit$

Corollary 4 *The range of values of $v(S_m)$ which will induce the plaintiff to always reveal S_m increases as $v(S_p)$ decreases, and as $v(S_d)$ grows.*

If we assume that acceptance of S_m, S_p, and S_d are independent events, then as $PA(S_p)$ increases, the range of values of $v(S_m)$ which will induce the plaintiff to always reveal S_m shrinks. Similarly for the defense.

Assuming independence, there is no change in the plaintiff or defendant's range for revelation given a change in $PA(S_m)$.

Proof: Assuming independence simplifies the constraint in Inequality 5, which we can now write as:

$$L_p = \frac{PA(S_d)}{(PA(S_p) - PA(S_d) - PA(S_d)PA(S_p))(PA(S_d) - 1)}. \tag{7}$$

By inspection, the limit does not depend on $PA(S_m)$.

Taking the derivative of the plaintiff's limit in Equation 7 with respect to $PA(S_p)$:

$$\frac{dL_p}{dPA(S_p)} = \frac{PA(S_d)}{(PA(S_p) - PA(S_d) - PA(S_d)PA(S_p))^2},$$

which is always positive. This is the lower bound of the range of values of $v(S_m)$ which induce the plaintiff to reveal S_m, so as it rises, that range shrinks.

To show that the effect of a rise in $v(S_p)$, we need only rescale the new verdicts, by dividing them by $v(S_p)$, and reapply Theorem 3. This will give us a limit in the new scale, L'_p, which is the plaintiff's limit in the case where $v'(S_p) = 1$. To find the plaintiff's limit in the original scale, we need only multiply L'_p by $v(S_p)$, which was assumed to be greater than one, so $L_p > L'_p$. Statistical independence was not required for this argument.

Similarly for the defense. $\diamondsuit$

Corollary 5 *Assume the same situation as Theorem 3, but the judge rules for the defense in the case when no story is accepted as true. Then there does not exist a situation where any defendant will reveal S_m. Any prosecutor will reveal S_m iff*

$$v(S_m) > \frac{PA(\{S_p, S_m\})}{PA(\{\neg S_d, S_m\})}.$$

Proof: In this case, the payoff given only two stories is:

$$\alpha_2 \equiv [PA(\{\neg S_d, S_p\}) + PA(\{S_d, S_p\})w(S_p, \{S_d, S_p\})],$$

and the payoff given three stories is:[6]

$$\begin{aligned}
&PA(\neg S_m)\alpha_2+\\
&PA(S_m)\cdot[PA(\{S_d,S_p,S_m\})\cdot[W(s(S_m,\{S_d,S_m,S_p\}))v(S_m)+W(s(S_p,\{S_d,S_m,S_p\}))]\\
&\quad+PA(\{S_d,\neg S_p,S_m\})\cdot W(s(S_m,\{S_d,S_m\}))v(S_m)\\
&\quad+PA(\{\neg S_d,S_p,S_m\})\cdot[W(s(S_m,\{S_d,S_m\})v(S_m)+W(s(S_p,\{S_d,S_m\}))]\\
&\quad+PA(\{\neg S_d,\neg S_p,S_m\})\cdot v(S_m)].
\end{aligned}$$

If the variables are such that the best case before revealing S_m has a higher payoff than the worst case after revealing it, then these variables guarantee that any defense would always reveal S_m. That is:

$$\begin{aligned}
V_{min}(S_d,S_p) &> V_{max}(S_d,S_m,S_p)\\
PA(\{\neg S_d,S_p\}) &> PA(\{\neg S_d,S_p,\neg S_m\})+PA(\{S_p,S_m\})\\
&\quad+PA(\{\neg S_p,S_m\})v(S_m)\\
PA(\{\neg S_d,S_p,S_m\}) &> PA(\{S_p,S_m\})+PA(\{\neg S_p,S_m\})v(S_m).
\end{aligned}$$

Since $PA(\{\neg S_d,S_p,S_m\})\leq PA(\{S_p,S_m\})$, this inequality will never be true, meaning that a defense attorney can always imagine a judge who will make revealing S_m a bad move.

For the prosecutor, he will always present S_m if the following worst-case inequality is true:

$$\begin{aligned}
V_{max}(S_d,S_p) &< V_{min}(S_d,S_m,S_p)\\
PA(S_p) &< PA(\{S_p,\neg S_m\})+PA(\{\neg S_d,S_m\})v(S_m)\\
PA(\{S_p,S_m\}) &< PA(\{\neg S_d,S_m\})v(S_m)\\
\frac{PA(\{S_p,S_m\})}{PA(\{\neg S_d,S_m\})} &< v(S_m).
\end{aligned}$$

$\diamondsuit$

References

David Austin-Smith. Allocating access for information and contributions. *Journal of Law, Economics & Organization*, 14(2):1431–1451, 1998.

[6]For consistency with the proof in Theorem 3, I assume that $W(S_p,S_d,S_p)$ is constant in the two-story and the three-story case. Again, if we allow this to change, the worst case only gets worse, so the result about the defendant does not change and the limit for the prosecutor becomes more stringent $(PA(S_p)(1-PA(\{\neg S_d,\neg S_m\}))/PA(\{\neg S_d,S_m\}))$.

Randall L Calvert. The value of biased information: A rational choice model of political advice. *The Journal of Politics*, 47(2):530–555, June 1985.

Vincent P Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, 50 (6):1431–1451, November 1982.

Mathias Dewatripont and Jean Tirole. Advocates. *Journal of Political Economy*, 107(1): 1–39, 1999.

Jacob Glazer and Ariel Rubenstein. Debates and decisions: On a rationale of argumentation rules. *Games and Economic Behavior*, 36:158–173, 2001.

C W J Granger. Investigating causal relations by econometric models and cross-spectral analysis. *Econometrica*, 37(3):424–438, June 1969.

Tony Honoré. *Responsibility and Fault*. Hart Publishing, 1999.

R Duncan Luce and Howard Raiffa. *Games and Decisions: Introduction and Critical Survey*. John Wiley & Sons, 1957.

Thomas A Mauet. *Trial Techniques*. Aspen Publishers, 6th edition, 2002.

Paul Milgrom and John Roberts. Relying on the information of interested parties. *RAND Journal of Economics*, 17(1):18–32, Spring 1986.

Peter Murray. *Basic Trial Advocacy*. Aspen Publishers, 1995.

Nancy Pennington and Reid Hastie. Evidence evaluation in complex decisionmaking. *Journal of Personality and Social Psychology*, 51(2):242–258, 1986.

Nancy Pennington and Reid Hastie. Explanation-based decision making: Effects of memory structure on judgement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3):521–533, 1988.

Nancy Pennington and Reid Hastie. Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology*, 62(2):189–206, 1992.

Judea Perl. *Causality*. Cambridge University Press, March 2000.

Cass R Sunstein. *Legal Reasoning and Political Conflict*. Oxford University Press, 1996.

Mariam Thalos. The reduction of causal processes. *Synthese*, 13l:99–128, 2002.