Information aggregation, with application to monotone ordering, advocacy, and conviviality

Thesis by

Ben Klemens

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2003
(Submitted June 2, 2003)

Acknowledgements

Thanks to my advisors: Kim Border, Peter Bossaerts, Matt Jackson, and Richard McKelvey. Thanks also to Ed McCaffrey and Asya Pervukhin for legal advice, Chris Hitchcock for his insights into causality, and my fellow graduate students for everything else.

Abstract

This dissertation is in three parts.

I. Chapter 1 presents a convenient notation for describing methods of aggregating information to form posterior distributions, allowing a description of both Bayesian updating and many of the cognitive errors people commit in the lab. Chapter 2 looks at the monotone ordering problem: if the prior distributions are ordered in some manner, what updating operations will preserve that ordering? Bayesian updating is a member of a small class of operators which preserve the monotone likelihood ratio property, but is not in the class of functions which preserve first-order stochastic dominance. This chapter also considers ordering distributions by their medians, which is useful for Political Science and other decision making applications.

II. Chapter 3 presents a literature review of existing models of information aggregation from one party and gives the very weak conditions under which one or two biased advocates will always reveal full information. Chapter 4 then presents a model of a trial, in which events are grouped into causal stories. Each story may point to a specific verdict, but the judge has leeway in selecting a verdict when multiple stories are shown to simultaneously be sufficient to explain an event. Two judges may be 'perfect Bayesians', share the same priors, and still arrive at different verdicts for the same trial. Unlike the information revelation literature to date, there may be apropos stories and facts that neither party will want to reveal in equilibrium.

III. Chapter 5 presents a simultaneous model of goods or actions which demonstrate conformity effects. Previous models of such goods universally describe people as acting in sequence; actors in the model here act simultaneously, so they must decide what to do based only on prior information about the distribution of tastes in the

society. The shape of this distribution (e.g., centered around zero, skewed upward, or fat-tailed) predicts the number of people who will act in some systematic ways, which I catalog here.

Advisors: Matthew O Jackson (primary), Kim C Border, Peter L Bossaerts, Richard D McKelvey

Contents

Foreword

This dissertation confronts a few of the problems I have encountered in my readings of the economic literature.

The first problem is in monotone ordering. Much of the existing literature discusses a certain result about the monotone likelihood ratio property (MLRP), that if two distributions satisfy the MLRP and are then updated using one piece of news, then the two resulting posterior distributions will also satisfy the MLRP. The problem with this result is that the MLRP is basically useless (save for a caveat below)—outside of mathematical models, it is difficult to believe that one person's beliefs about a lottery would satisfy the MLRP with respect to another person's, nor is there any easy way to test or observe this.

Another ordering, first-order stochastic dominance (FOSD), *is* useful. If one person's beliefs about a lottery FOSD another's, then the first person gains more utility from the lottery, regardless of the details of the utility function. The MLRP only says anything about people to the extent that it implies FOSD. So Chapter 2 goes directly to FOSD, asking what operators preserve the FOSD ordering, the way that Bayesian updating preserves the MLRP. In order to answer this question, I introduce a notation for defining general updating operations in Chapter 1.

The next problem is in information provision by one or two biased advocates. If we assume a simple data gathering framework and allow actors to find the game theoretic equilibrium, then they *always* get the payoff from full information revelation. Especially with two advocates, this is a very robust result. Unfortunately, it does not ring true. No lawyer I have spoken to feels that this describes the outcomes they have observed in the courtroom, where advocates carefully select the information they wish to reveal. We are similarly hard-pressed to find people who feel that having opposing lobbyists is sufficient for legislators to garner all the information they need, which would imply that the democratic process would be better served with better-funded lobbyists and fewer confounding signals from the populace. Although there are times when layman observations are corrected by Game Theory, I do not think this is such a case: full information is not always revealed.

There are two modifications we can make. The first is to assume that decision makers do not play a game toward a Nash equilibrium, but naïvely process information. This is the approach of Chapter 3, which reviews some of the literature that takes this approach and describes a few other models apropos to modeling lobbyists under this regime.

The other modification is to leave the machinery of Nash equilibria and Bayesian updating in place, but make information more complex. There must be an asymmetry between information providers for the full information equilibrium to be broken. That is, the decision maker must process information differently when she hears it from one person than when she hears it from another. This is realistic, but seems on the surface to be anathematic to the ideas of rational information processing. The model of Chapter 4 reconciles this: the decision maker is capable of correct Bayesian updating, but in certain situations still has leeway to process information differently. The key is to take into account the question of causality, which is outside the realm of statistics.

The final problem is in group behavior, notably the emulation of others. This is a common human trait, but is not well treated in the economic literature. The standard model claims that people emulate others because of limitations in information gathering: unable to research a question themselves, they observe consumption by their cohort, and take that as a recommendation for the product consumed. Eventually, there is a critical mass of recommendations, and everyone thereafter consumes the same product—with probability one. The reasoning is not descriptive of how many consumption decisions are made, and the prediction is clearly false in innumerable cases in which emulation of others is an issue.

In Chapter 5, I offer an alternative model. As with the models above, the actors follow the rules of Statistics and Game Theory closely, but they explicitly get utility from emulating others. This is mostly consistent with the information cascade model described above, but allows individual preferences to have greater reign: if somebody strongly dislikes a product, they may not buy even if everybody else does. It also clarifies the importance of expectations, especially in situations of partial information. Finally, the distribution of tastes for a good predict what percentage of people will consume, with extreme outcomes like 100% consumption occurring only in extreme situations.

The last two sections also indirectly address another difficulty in the rational information aggregation literature: advertising. We all know the product the Coca-cola Company makes, and all of its relevant characteristics regarding taste, caffeine content, et cetera—so why does Coca-cola blanket the world with advertising? Further, why is the advertising basically contentless? If consumers are purely data-oriented and only take from an ad the fact that there exists a product with certain features, then advertising for Coca-cola and virtually any other product would look very different from what we see. The forms of advertising Coca-cola uses in the real world derive naturally from the type of advocate described in Chapter 4 and the type of consumer described in Chapter 5.

Chapter 1

Updating operators introduced

This chapter will introduce a general notation for updating operators. Such an operator takes two sets of beliefs or information as inputs, and produces a posterior set of beliefs as an output. Especially within the behavioral literature, there exist an abundance of models of learning and updating which each have their advantages. In this chapter, I will not advocate for one over the others, and will not propose new ones. I will simply discuss the need for a framework which includes Bayesian updating as a special case of a more general process, and will offer a consistent notation for this process.

In Chapter 2, I will show that some operators preserve certain orderings of sets of beliefs, while other updating operators preserve other orderings.

1.1 Beliefs versus data

Conside a consumer's beliefs about the likelihood that an advertised good is useful or not, or a person's beliefs about the likelihood of good or bad weather tomorrow. We can write these beliefs as probability density functions (PDFs), just as we can use a PDF to summarize the data which the weatherman uses to determine the likelihood of rain. But although they can both be written as PDFs, distributions of beliefs are not distributions of data.

Data involves events which have actually happened and—apart from potential doubts about our measurement instruments or misapplications of the Heisenberg

Uncertainty Principle—are not in dispute. If we are unsure about how the data was produced, we can potentially break open any black boxes and understand the mechanism to as much precision as our capabilities allow.

Beliefs often come from unbreakable black boxes. For example, information providers may have trade secrets which they will not reveal, or they may prefer not to let people know how much of a guess their beliefs really are. Consumers may not know why they believe that the red dongle is better than the blue dongle, and even if scientists could put together a model of the brain that could explain these opinions, it is probably impossible to know enough about a person's experiences or past training to put the model to use. Finally, there are the simple practical considerations, that we ask the beliefs of information providers in situations where our time or abilities preclude gathering the data ourselves. If we had the ability and resources to understand the sources of an expert's data, then we wouldn't have bothered asking the expert. So unlike black boxes which produce data, which we may be able to take apart and understand better, the head-shaped boxes that produce beliefs can not be taken apart, for social, biological, and practical reasons.

This leads to another important difference between beliefs and data. It is an easy matter to put together a method of aggregating data. Bayes's rule is easily applied to produce Bayesian updating, which, given the right setup, is a mathematically correct method for assimilating new data. But since beliefs differ from data for all the reasons above, it is difficult to justify using the same method to aggregate them.[1]

1.2 Definition of an updating operator

This section presents a notation which captures the idea of taking two PDFs as input, and then concluding with a posterior PDF.

Let $\tau(x)$ be the true, unobservable PDF of x. Say that one source claims that for $x = 1$, $\tau(1) = a$, and another source claims $\tau(1) = b$. The decision maker must

[1] For an entirely different discussion of why probabilities describing beliefs are not to be treated like probabilities describing random events, including some interesting history on the subject, see the introductory chapter to Shafer [50].

amalgamate these two data points into one belief, so let $op(a,b)$ be the updating operation, mapping $\mathbb{R}^+ \times \mathbb{R}^+$ to $\mathbb{R}^+$.

The same may be done for any value of x: let $a(x)$ be one source's claims about the true distribution of x, and let $b(x)$ be the other's. Then $op(a(x), b(x))$ defines an implicit function mapping x to $\mathbb{R}^+$. A few more caveats ensure that this will lead to a valid output given valid PDFs as inputs.

Definition 1 *An updating operator is any two-variable function $op(a,b) : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$, where $op(a,b) \geq 0$ for all $a \geq 0$ and $b \geq 0$; and if $\int_{-\infty}^{\infty} a(x)dx = 1$ and $\int_{-\infty}^{\infty} b(x)dx = 1$, then $\int_{-\infty}^{\infty} op(a(x), b(x))dx$ is finite.*

I will restrict attention to operators which are continuous in both variables. It will also be occasionally useful to restrict attention to those updating operators which are increasing in both arguments. That is, if $\alpha > \beta$, then $op(\alpha, \gamma) > op(\beta, \gamma)$ and $op(\gamma, \alpha) > op(\gamma, \beta)$ for all γ. This is sensible because if one source modifies its report from $a(x_1) = \beta$ to $a(x_1) = \alpha$, thus putting more weight on x_1, a listener should take that into account when aggregating the information and put a little more weight on x_1 as well.

Although the discussion to this point has been oriented toward symmetric applications, the first and second inputs into the updating operation will have different interpretations for many of the applications below. Further, symmetric operators will be shown to have some additional implications over the general specification of Definition 1, where $op(a,b)$ may or may not equal $op(b,a)$.

There is no reason why $\int_{-\infty}^{\infty} op(f(x), g(x))dx$ should be one, meaning that the actual posterior requires normalization.

Definition 2 *The posterior distribution given updating operator $op(\cdot,\cdot)$ and inputs $f(\cdot)$ and $g(\cdot)$ is*

$$post(f(\cdot), g(\cdot), op(\cdot,\cdot), x) = \frac{op(f(x), g(x))}{\int_{-\infty}^{\infty} op(f(y), g(y))dy},$$

which is guaranteed to integrate to one.

In this notation, Bayesian updating is equivalent to $op(f(x|\cdot), g(\cdot)) = f(x|\cdot)g(\cdot)$, for any fixed x, giving a posterior which matches the definition given below in Equation 2.1. For example, let x be an observable variable and t be a parameter of x such as the mean. A Bayesian researcher would begin with prior beliefs that t has a PDF $g(t)$, that the data has distribution $f(x|t)$. The researcher then runs an experiment which provides a value of x, χ. At this point, the researcher applies $op(f(\chi|t), g(t)) = f(\chi|t)g(t)$ to arrive at a posterior $post(t|\chi)$, which is a function of t but not x.

Shafer [50] describes how this is equivalent to Dempster's rule within the probability framework here.

For a posterior which averages the priors, $op(f(\cdot), g(\cdot)) = f(\cdot) + g(\cdot)$, and the posterior is

$$post(f(\cdot), g(\cdot), op(\cdot, \cdot), x) = \frac{f(x) + g(x)}{\int_{-\infty}^{\infty} f(y) + g(y) dy} = \frac{f(x) + g(x)}{2}.$$

Finally, notice that if k is any positive constant,

$$\frac{k \cdot op(f(x), g(x))}{\int_{-\infty}^{\infty} k \cdot op(f(y), g(y)) dy} = \frac{op(f(x), g(x))}{\int_{-\infty}^{\infty} op(f(y), g(y)) dy}.$$

In other words, any statement below about the properties of $op(f(\cdot), g(\cdot), x)$ also applies to $op'(f(\cdot), g(\cdot), x) = k \cdot op(f(\cdot), g(\cdot), x)$. Having made this note here, I will state all results below ignoring the fact that the operators can trivially be multiplied by any positive constant.

1.3 A few examples

This section takes examples from the behavioral literature and shows what they imply about the form of the updating operator. It is not a comprehensive or complete survey of the literature, but will give the reader familiar with the various effects discussed in behavioral economics an idea of how those effects are described by the framework here.

1.3.1 Order effects

There is extensive experimental evidence that the order in which information is received will affect the way in which it is processed; see, e.g., Camerer [12] for a survey. For example, Tversky and Kahneman [54] found 'anchoring' effects where subjects place more weight on evidence presented earlier than on evidence presented later. Such behavior can not be modeled by a symmetric updating operator: it must be the case that the order of the inputs to the function matches the order in which the data is presented, and $op(a, b) \neq op(b, a)$.

This experimental evidence tells us nothing more about the form of updating operator. For example, the meaning of 'weighting' depends on the choice of operator: a multiplicative operator could take the form $op(a, b) = a^k \cdot b,\ k \geq 1$; an additive operator could take the form $op(a, b) = k \cdot a + b,\ k \geq 1$.

In a slightly different framework, such asymmetric operators will allow partial information revelation on page 53.

1.3.1.1 Learning models

'Cournot learning' is the creation of beliefs based on the last play observed. In other words, given a new round of information, the updating rule is to throw out past rounds and simply take the new round as a new set of beliefs. 'Fictitious play', in the context here, involves a learning model which averages outcomes over all past rounds. The assumptions of both of these learning models can be subsumed into one model, by a weighted averaging operator

$$op(a, b) = a + \gamma b,$$

where a is the new probability of an event, and b is the old beliefs about its probability (the output of the updating operation from the last round). If $\gamma = 0$, then this is

Cournot learning, and if γ is greater than to one, then it is closer to the learning associated with fictitious play. See Fudenberg and Levine [22] or Cheung and Friedman [14]; or see Camerer and Ho's Experience-Weighted Attraction [11], which is based on this method of updating.

This form will make a reappearance in Section 2.6.3.

1.3.2 Framing effects

Prior information can affect the absorption of subsequent information. People who strongly believe a theory may ignore information which counters their pet theory, and update only using information which agrees with their priors. Camerer ([12], introduction) gives an extensive review of the various studies which have observed such effects.

The most direct way to model this is to say that there is a perception function, $f(a, b)$, which transforms new information b based on the prior beliefs a. A standard updating method may then be applied to the transformed information. An unbiased observer would update using $op(a, b)$, while an observer who suffers these encoding biases would update using $op(a, f(a, b))$.

For example, say that the mean of a variable is a hotly contested issue. People who believe that the mean is positive ignore new information which claims that it is negative, and vice versa. Then

$$f(a(x), b(x)) = \begin{cases} U(x) & \text{mean}(a) \cdot \text{mean}(b) < 0 \\ b(x) & \text{mean}(a) \cdot \text{mean}(b) \geq 0 \end{cases},$$

where $U(x)$ is an uninformative prior (defined formally in Section 5.4.3). Normal Bayesian updating, as above, is the updating operator $op(a, b) = a \cdot b$, so Bayesian updating given framing effects can be modeled by $op(a, f(a, b))$, or be rewritten as a function only of a and b:

$$op'(a, b) = \begin{cases} a & \text{mean}(a) \cdot \text{mean}(b) < 0 \\ a \cdot b & \text{mean}(a) \cdot \text{mean}(b) \geq 0 \end{cases},$$

since Bayesian updating based on the uninformative prior results in a posterior which equals the prior.

1.4 Conclusion

We have only one tool for updating beliefs about the probability of events using data: Bayes's rule. But if we only have beliefs presented to us by consultants, advertisers, newspapers, or other humans, then it is inappropriate to use Bayes's rule as if these are objective observations about potential states of the world. It is often physically impossible to gather enough information to reduce these opinions to objective data, and replacing the missing information with beliefs about the received beliefs, beliefs about those beliefs about beliefs, et cetera, leads to an analytic quagmire which is beyond the computational ability of most humans, and makes it impossible for the modeler to make serious predictions or to say anything descriptively interesting about people and their processing of beliefs.

There are a number of methods of assimilating information which have shown themselves to be descriptive of how people in experimental labs and in the real world assimilate data.

Because of both the theoretical problem that beliefs are not data and can not be reduced to data, and the results from the lab, it is valuable to admit different types of updating operations into our theoretical models. This chapter introduced a notational framework for doing this, and gave examples of how this corresponds to certain effects observed in the behavioral literature. The chapters which follow will use this framework to determine which updating operators are admissible given certain pieces of evidence or theoretical requirements.

Chapter 2

Gathering information from multiple parties

2.1 Introduction

First-order stochastic dominance (FOSD) is a useful property for a set of probability distributions to have. A lottery whose distribution FOSDs[1] that of another lottery can be said to be 'better' or 'more optimistic', in the sense that the expected utility given the the dominant lottery is higher than that of the dominated lottery. This statement assumes that a higher payoff is better, but assumes nothing about risk preferences, returns to scale, or anything else about the form of utility functions.

For some applications, FOSD can be used as a sanity check: the distribution of a stock's expected ask price should always FOSD the distribution of the expected bid, and as new information comes in, the updated ask distribution should still FOSD the updated bid's distribution. In other applications, it is a useful assumption: one bidder may have a higher expected value for a good than another, which we can model by saying that the first bidder's distribution of the good's expected value FOSDs that of the second.

What sort of information from the auctioneer would or would not affect this ordering?

[1] The phrase 'f FOSDs g' may be read as 'f first-order stochastically dominates g'. The clarity of the abbreviation more than makes up for any æsthetic shortcomings it may have. Similarly, the reader may take 'f MLRPs g' to mean 'f satisfies the monotone likelihood ratio property [defined below] with respect to g'.

Similarly, Section 5.5.2 below asks when one advertisement can be classed as 'better' than another. If one person begins with some prior and were to update with advertisement A to posterior A, or were to update with advertisement B to posterior B, then if posterior A FOSDs posterior B, we can say that A is a better advertisement than B. This is Milgrom's application in his 'good news, bad news' paper [37].

If the prior distributions satisfy the monotone likelihood ratio property (the MLRP, defined below), then the posterior distributions after Bayesian updating with common information will also satisfy the MLRP, regardless of what common information is used to update. Since any pair of distributions which satisfy the MLRP also satisfy FOSD, MLRP priors guarantee Bayesian posteriors which are ordered by FOSD.

This is the state of the literature. This result appears in a discrete form in Whitt [58], in Milgrom [37], and in some of the papers that they cite. Bikhchandani, Segal, and Sharma [8] do a similar thing with a simplified setup.

This chapter explores this result. I first show in Section 2.3.2 that the MLRP is both sufficient, as discussed in the literature, and necessary for posteriors which are guaranteed to be ordered by FOSD. Then in Section 2.4, I modify the theorem to describe a situation where posteriors are ordered not by FOSD, but by statistics such as the median. This result requires that public information be log-concave (for example, the common normal distribution), and private information be from a translation family. For those applications which focus on the median, such as political questions or some types of jury situations, these conditions are all that is necessary.

Within the context of the general notation for updating operators introduced in Chapter 1, the obvious next question is: how far can these results be generalized beyond Bayesian updating?

Calvert [10] claims that "[...] there is nothing magical about Bayes's rule that should cause us to believe, in advance, that a different rule would qualitatively change our conclusions about the rational use of [...] information." Section 2.6 shows that there *is* something magical about Bayes's rule, and that other reasonable methods of updating do not have the consistency property of preserving the MLRP for which

Bayes's rule is frequently cited in the literature. If we believe that the MLRP should be preserved, then we believe that Bayesian updating is the correct method of aggregating information.

But as proven in Section 2.3.2, Bayesian updating does not guarantee posteriors ordered by FOSD given arbitrary priors ordered by FOSD—the priors need the more restrictive monotone likelihood ratio property. Theorem 16 shows this again, by describing the full set of updating operations which will guarantee FOSD posteriors given FOSD priors. This set of operators consists of those which take linear weighted averages of their inputs, and conspicuously excludes Bayesian updating from the set.

2.2 Preliminary notes

Say that there are two competing researchers, who both know the conditional distribution $f(x|t)$ (with $x, t \in \mathbb{R}$), but have private probability density functions for t, $g_1(t)$ and $g_2(t)$. Assume that all functions have strictly positive values for all $x \in \mathbb{R}$ and $t \in \mathbb{R}$ (that is, they have full support).

Given a probability density function (PDF) $g_i(t)$ describing the parameter t, Bayes's rule tells us how to use $f(x|t)$ to update to posterior distributions:[2]

$$post_i(t|x) = \frac{f(x|t)g_i(t)}{\int_{\mathbb{R}} f(x|\tau)g_i(\tau)d\tau}. \tag{2.1}$$

Typically, t is a parameter such as a mean, and x is new data such as experimental observation. Let $\tilde{x}$ be the observed values of x; after $\tilde{x}$ is observed, it has a fixed value, so $post_i(t|\tilde{x})$ is a function of only t.

That said, questions of whether the various functions $g_i(t)$ and $post_i(t|x)$ are ordered of course require some methods of ordering distributions.

[2]'Bayesian updating' means different things to different people. For example, Majumdar [34] and Whitt [58] prefer a form which is applicable to a finite number of states of the world. New information eliminates states of the world from consideration, and thus changes the probabilities placed on the other states. Here, I use a continuous form, and new information comes in the form of a likelihood function with full support.

2.2.1 Ways to order a family of distributions

Since this section gives general definitions, take $p(t)$ to be any PDF, mapping from values of $t \in \mathbb{R}$ to $\mathbb{R}^+$.

Often, we have a family of related PDFs. For example, each researcher may have beliefs which differ slightly from the beliefs of other researchers. Let $i \in \mathbb{R}$ be an index, and denote a member of the family of distributions as either $p_i(t)$ or, equivalently, $p(t|i)$. A family of PDFs is ordered when any two members $p_i(t)$ and $p_j(t)$ are ordered.

MLRP, the monotone likelihood ratio property: if $i > j$,

$$\frac{p(t|i)}{p(t|j)}$$

is an increasing function of t; if $i < j$ it is decreasing, and is constant (one) if $i = j$. Notice that these are strict inequalities.

Having two functions $p_1(x)$ and $p_2(x)$ is equivalent to having one function $p'(x|i)$ where $p'(x|1) = p_1(x)$ and $p'(x|2) = p_2(x)$. Therefore, to say '$g_1(t)$ MLRPs $g_2(t)$' means that $g_1(t)/g_2(t)$ is an increasing function of t.

FOSD, first-order stochastic dominance: $p_1(t)$ FOSDs $p_2(t)$ iff for any constant k:

$$\int_k^\infty p_1(t)dt > \int_k^\infty p_2(t)dt. \tag{2.2}$$

Alternatively, a family $p(\cdot|t)$ satisfies FOSD iff $p(\cdot|t_1)$ FOSDs $p(\cdot|t_2)$ for all $t_1 > t_2$.

single-crossing: $p_1(t) > p_2(t)$ for all t less than some point K, and $p_1(t) < p_2(t)$ for all $t > K$.

translation: A translation family of distributions is a set of PDFs $p(t|i)$ such that $p(t - i|i) = p(t|0)$. For example, the Normal distributions, $T \sim \mathcal{N}(i, 1)$, are a translation family. I will often refer to $p(t|0)$ as simply $p(t)$, which is the underlying distribution which is being translated by i units to produce $p(t|i)$.

These terms allow the following definitions:

Definition 3 *Consider PDFs f, g_1, and g_2, all of which have full support over the reals. Then an updating operation* preserves FOSD *iff $g_1(t)$ FOSDs $g_2(t)$ implies that $post_1(t|x)$ FOSDs $post_2(t|x)$ for all likelihood functions $f(x|t)$ and any fixed x.*

Definition 4 Preserving the MLRP *is similarly defined.*

2.3 Bayesian updating preserves the MLRP

As discussed above, FOSD provides an ideal method of ordering distributions. But unfortunately, it is easy to construct examples which show that Bayesian updating does *not* preserve FOSD. This section shows how the problem can be surmounted using the MLRP.

Diagrammatically, Figure 2.1 shows the trick used in the literature to ensure that posteriors satisfy FOSD: MLRP priors imply MLRP posteriors, which in turn imply FOSD posteriors.

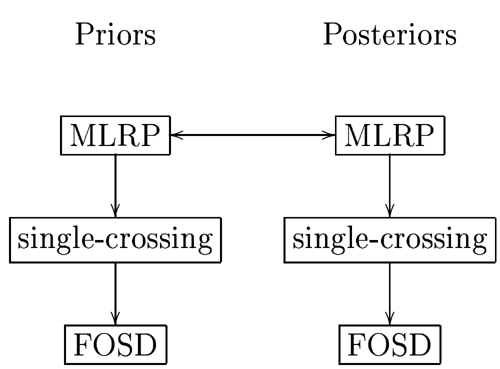

Figure 2.1: Relations to be proven in the sequel

2.3.1 Ordering posteriors with the MLRP

The validity of the arrows in Figure 2.1 are proven in the following statements, all of which have appeared in the literature and/or are easily proven.

Lemma 1 *Let $t \in \mathbb{R}$, and $p_1(t)$ and $p_2(t)$ be continuous PDFs. Define K to be the set of ts such that $p_1(t)/p_2(t) = 1$. If $p_1(t)$ MLRPs $p_2(t)$, then K is a single point. Also, $p_1(t)/p_2(t) < 1$ for all $t < K$ and $p_1(t)/p_2(t) > 1$ for all $t > K$. In other words, MLRP $\Rightarrow$single-crossing.*

[This and other proofs not given in the body of the text are in the appendix, Chapter 7.]

Lemma 2 *If $p_i(t)$ is a family of single-crossing distributions then the family satisfies FOSD.*

Lemma 3 *Given $x \in \mathbb{R}$ and $t \in \mathbb{R}$, and a functions $f(x,t)$ with full support over the ranges of x and t. Then if $f(x,t)$ satisfies the MLRP with respect to t [that is, $f(x,2)$ MLRPs $f(x,1)$], then it also satisfies the MLRP with respect to x [that is, $f(2,t)$ MLRPs $f(1,t)$].*

Lemma 4 *Bayesian updating is done using a draw from x and the single distribution $f(x|t)$. Then a distribution family $g_i(t)$ satisfies the MLRP $\Leftrightarrow$ within the support of $f(x|t)$, the family of posterior distributions satisfies the MLRP with respect to t.*

Proof: This is proven in a more general setting by Proposition 10 and Lemma 11. $\diamondsuit$

Note also that by Lemma 3, statisfying the MLRP with respect to t is equivalent to satisfying the MLRP with respect to x.

Chaining the above together gives the following result:

Theorem 5 *MLRP priors (on t) $\Rightarrow$ FOSD posteriors on t for any fixed x, within the support of $f(x|t)$. That is, $g_1(t)$ MLRPs $g_2(t)$ implies $post_1(t|x)$ FOSDs $post_2(t|x)$ for any x.*

Proof: MLRP priors on $t \Rightarrow$ MLRP posteriors on t (by Lemma 4); MLRP posteriors $\Rightarrow$ single-crossing posteriors (by Lemma 1); single-crossing posteriors $\Rightarrow$ FOSD posteriors on t (by Lemma 2). $\diamondsuit$

2.3.2 Necessity

For any set of FOSD likelihood functions $g_i(t)$, there is *some* prior distribution $f(x|t)$ such that updating g with f always results in FOSD posteriors—just use the Uniform distribution over the support of the gs, for example.[3]

However, if the MLRP is not satisfied (for a set of points greater than measure zero), then there is always a function $f(x|t)$ for which the posteriors will not be ordered by FOSD for some value of x. This section will show how to construct such a function, thus demonstrating the 'only if' part of the following:

Proposition 6 *A pair of priors $g_1(t)$ and $g_2(t)$ such that $g_1(t)$ FOSDs $g_2(t)$ can be updated using* any *likelihood function $f(x|t)$ to an FOSD family of posteriors iff $g_1(t)$ MLRPs $g_2(t)$.*

Proof: Theorem 5 proved that if $g_1(t)$ MLRPs $g_2(t)$, then these prior distributions satisfy FOSD, and can be updated by any likelihood function to an FOSD family of posteriors. To prove the other direction, assume the MLRP is not satisfied. Then there is some range (a, b) over which $g_1(t)/g_2(t)$ is decreasing. Define

$$g_1' = \begin{cases} g_1(t) & t \in (a,b) \\ 0 & t \notin (a,b) \end{cases}, \qquad \text{and} \qquad g_2' = \begin{cases} g_2(t) & t \in (a,b) \\ 0 & t \notin (a,b) \end{cases}.^4$$

Then $g_2'(t)/g_1'(t)$ is (weakly) increasing everywhere the ratio is defined, so Theorem 5 can be applied to show that regardless of the distribution $f(x|t)$ and the observed x, $post_2'(t|x)$ FOSDs $post_1'(t|x)$ for any fixed x.

Let $f(x|t)$ be the Uniform$[a, b]$ distribution for all t. Then $post_1(t) = post_1'(t)$ and $post_2(t) = post_2'(t)$, since they match throughout the support of $f(x|t)$ and are zero elsewhere. Therefore $post_2(t)$ FOSDs $post_1(t)$—so $post_1(t)$ clearly does not FOSD $post_2(t)$. Thus, for any family of distributions which does not satisfy the MLRP for some range, there is a distribution with a small support which can be used to update to get non-FOSD posteriors. $\diamondsuit$

[3] For the support of $(-\infty, \infty)$, a Uniform$(-\infty, \infty)$ distribution is defined on page 93.

[4] These distributions need to be multiplied by a constant to integrate to one, but the constant is clearly irrelevant.

The Uniform distribution is certainly exceptional, having limited support and only two values of $p(t)$. So as a further embellishment, consider this distribution:

$$T \sim \begin{cases} \frac{\epsilon(1-e^{(y-a)})}{2(b-a)} & y < a \\ 1-\epsilon & a \leq y \leq b \\ \frac{\epsilon(1-e^{(b-y)})}{2(b-a)} & y > b \end{cases} . \tag{2.3}$$

The distribution approaches the Uniform distribution as ϵ goes to zero, but it has full support, no discontinuities, is weakly single-peaked, and would still serve as a counterexample for an appropriately small ϵ. So there is nothing unique about the support or form of the Uniform distribution.

2.4 Linear translation

This section finds conditions where any translation family (which may or may not satisfy the MLRP) leads to ordered posteriors. If $g(t|i)$ is a translation family as per the definition in Section 2.2.1, then $g(t|2)$ FOSDs $g(t|1)$. But as discussed above, FOSD priors are not sufficient to guarantee FOSD posteriors, so instead, a new ordering will be offered.

The trick is to set up an MLRP translation family for the function used to update, and then do an appropriate coordinate transformation. It relies on the following:

Proposition 7 *A function is log-concave iff its translation family satisfies MLRP.*

For a proof, see Eeckhoudt and Golier [19], section 4.1.[5] It will be useful to note for later reference that this result does not depend on the fact that the integral of a PDF over its entire domain is one.

Consider this alternate setup: $f_i(x|t)$ is a translation family, taking functions of the form $f(x|t)$ and matching them to a single function $g(t)$ in different positions, as dictated by i. Barring a few details to be discussed below, this is equivalent to

[5] They refer to the MLRP as the Monotone Probability Ratio (MPR) property, reserving the name monotone likelihood ratio property for something else.

having a translation family $g_i(t)$, whose members are matched to one and only one distribution $f(x|t)$.

This version is the model above, where $g(t|i)$ is specific to individual i and $f(x|t)$ is a single, common function. The symmetry of x, t, and Equation 2.1 allows this. However, the conditions of Lemma 4 have shifted: in the alternate setup, it was necessary that the family $f_i(x|t)$ satisfy the MLRP, but here the $g_i(t)$ family need only satisfy translation, in many ways an easier condition to satisfy.[6] However, the function $f(x|t)$ must now satisfy log-concavity (since its translation family had to satisfy the MLRP).

The statement about the posterior also changes in the modified coordinate system; to clarify, I present the following fact, which is true for any pair of functions f and g:

$$\int_A^B f(x)g(x+d)dx = \int_{A+d}^{B+d} f(x-d)g(x)dx. \tag{2.4}$$

Figure 2.2 shows this graphically: the left-hand side of Equation 2.4 multiplies the values of $g(x)$ in the region between $A+d$ and $B+d$ by the values of $f(x)$ between A and B. The right-hand side of the equation multiplies the same values of $g(x)$ in the region between $A+d$ and $B+d$ by the values of $f(x-d)$ in the region between $A+d$ and $B+d$. Since the values of $f(x)$ in the first region and $f(x-d)$ in the second are the same, the equation holds.

Let the CDF of $post_i(x|t)$ up to a be $POST_i(a|t)$; Equation 2.4 allows us to rewrite the numerator of $POST_i(a|t)$ (see Equation 2.1), using $A=-\infty$, $B=0$, and $d=i$. Making the substitution and then neatening the notation:

$$\int_{-\infty}^{0} f(x)g(x+i)dx \quad = \quad \int_{-\infty}^{i} f(x-i)g(x)dx,$$

[6]'Easier' should not be taken too literally here: there are families which satisfy MLRP but are not translation families (such as the Chi-squared distributions), and translation families which do not satisfy MLRP. However, it is easier to justify modelling two people having beliefs which are translations of each other than to explain why two people would have beliefs which MLRPed each other.

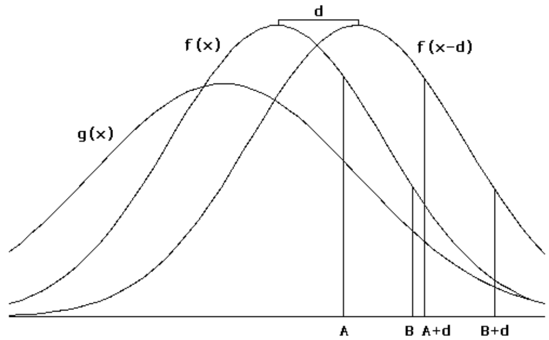

Figure 2.2: Visual aid for Equation 2.4

i.e.,

$$\int_{-\infty}^{0} f(x)g(x|-i)dx = \int_{-\infty}^{i} f(x|i)g(x)dx.$$

The left-hand integral is up to a fixed point in the coordinate space, and could be written as the numerator to $POST_{-i}(0)$. But the right-hand integral is up to the moving index for the translation family: $POST_i(i)$.

We know that the left side satisfies FOSD, and so is increasing as i increases (that is, as $-i$ falls); so the right-hand side must also be increasing as i increases.

Since $f_i(x|t)$ was a translation family which satisfies the MLRP for any fixed x, Proposition 7 says that $f(x|t)$ (aka $f_0(x|t)$) must be log-concave given x.

The following statement has now been proven in the new coordinate system:

Theorem 8 *If $g_i(t)$ is a translation family of priors and $f(x|t)$ is a log-concave likelihood function for any given x, then $POST_i(i)$ is an increasing function of i.*

This means that for low values of i, most of the distribution is above i (that is, i is below the median of the distribution given i), while for high values of i most of the distribution is below i (that is, i is above the median given i).

Log-concavity on non-PDF general functions How shall we interpret the log-concavity of $f(x|t)$ in terms of t? Consider the bivariate function $f'(x,t) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^+$, where $f'(x,t) = f(x|t)$. The only conceptual asymmetry between x and t is that

$\int_{\forall t} f'(x,\tau)d\tau$ must be finite, while $\int_{\forall x} f'(\chi,t)d\chi$ may take on any value. Some of the results about PDFs listed to this point apply when we hold x fixed at $\bar{x}$ but let t vary as when we hold t fixed at $\bar{t}$ and let x vary.

For example, any nonnegative, continuous, log-concave function can not have two local maxima. If $f'(x,\bar{t})$ satisfies log-concavity for any fixed $\bar{t}$, then $f'(x,\bar{t})$ is single-peaked. Similarly, if $f'(\bar{x},t)$ satisfies log-concavity for any fixed $\bar{x}$, then $f'(\bar{x},t)$ must be either single-peaked, monotonically decreasing, or monotonically increasing.

If $f'(\bar{x},t)$ is log-concave for any fixed $\bar{x}$, then $f'(\bar{x},t)$ satisfies the MLRP with respect to any translation of itself. For example, $f'(\bar{x},t)$ MLRPs $f'(\bar{x},t+7)$.

A pair of functions which satisfy the MLRP can not cross more than once. Since $f'(x,\bar{t})$ MLRPs $f'(x+7,\bar{t})$, the two functions satisfy single-crossing or do not cross at all; the same is true when $f'(\bar{x},t)$ MLRPs $f'(\bar{x},t+7)$.

So although $f'(\bar{x},t)$ is not a PDF, many of the implications of log-concavity still hold, sometimes with minor modification.

2.4.1 When both theorems hold

Intuitively, as more information appears, one may expect that the various private posteriors converge toward a single distribution. One may easily construct counterexamples where this is false, but in the case where both of the above theorems hold, then this is indeed the case. This section states the joint result formally and then gives a detailed description of how the distributions compress into one distribution.

Corollary 9 *If the priors are a translation family which satisfy the MLRP, and are updated using a log-concave distribution $g(t)$, and $k \in \mathbb{R}$ is any constant, then both*

$$\int_k^\infty post_i(\tau)d\tau \text{ and } \int_{-\infty}^{i+k} post_i(\tau)d\tau$$

are increasing functions of i for all k.

Proof: The premises of both Theorems 5 and 8 hold, and therefore the conclusions of both theorems are simultaneously true. $\diamondsuit$

For some intuition as to what this means, consider the location of the medians under the two conclusions above. The priors were a translation family, meaning that if the median of $g_0(t)$ is at 0, then the median of $g_1(t)$ is at 1, the median of $g_2(t)$ is at 2, et cetera.

As for the posteriors, say that the median of $post_0(t)$ is zero, i.e.,

$$\int_0^\infty post_0(\tau)d\tau = \frac{1}{2}.$$

The following lines prove by contradiction that the median of a posterior distribution with index greater than zero must be greater than zero when Corollary 9 holds. Consider $post_1(t)$ integrated from $-k$ to infinity, where $-k$ is any negative number:

$$\begin{aligned}\int_{-k}^\infty post_1(\tau)d\tau &> \int_0^\infty post_1(\tau)d\tau \\ &> \int_0^\infty post_0(\tau)d\tau \qquad (2.5) \\ &= \frac{1}{2}.\end{aligned}$$

Inequality 2.5 is an application of Theorem 8 (which restates Theorem 5). This sequence of inequalities demonstrates that it would be a contradiction for the median of $post_1(\cdot)$ to be at any value $-k$ less than zero. More generally, the median must be an increasing function of the index whenever FOSD holds.

If the second inequality holds, then the median of $post_1(\cdot)$ can not be too much greater than the median of $post_0(\cdot)$. Again, assume that the median of $post_0(\cdot)$ is at zero, so $\int_{-\infty}^0 post_0(\tau)d\tau = \frac{1}{2}$. Then:

$$\begin{aligned}\int_{-\infty}^1 post_1(\tau)d\tau &> \int_{-\infty}^0 post_0(\tau)d\tau \qquad (2.6) \\ &= \frac{1}{2}.\end{aligned}$$

Inequality 2.6 is from the second half of Theorem 8 (which restates Corollary 9). So the median of $post_1(\cdot)$ can't be larger than one if the median of $post_0(\cdot)$ is zero. Combining both of these restrictions at once means that if the median of $post_0(\cdot)$ is zero, the median of $post_1(\cdot)$ is between zero and one. Similarly, the median of $post_{-1}(\cdot)$ is between negative one and zero, the median of $post_2(\cdot)$ is between zero and two, et cetera. Further, this holds not only for the median—that is, where the CDF is $\frac{1}{2}$—but where the CDF is any value between zero and one, e.g., the first quartiles will compress in upon themselves in a similar fashion. Thus, the distributions begin evenly spaced, but after updating, they compress in upon themselves.

This discussion applies to any of the 'weighted Bayes's rules' which will be defined and discussed below.

2.5 A few examples

Here are a few brief sketches of how theorems such as Corollary 9 could be applied in the design of models. The first is taken from consumer theory, and the second is from Political Science.

2.5.1 Expectations about a movie

Let the quality of a movie be $t \in \mathbb{R}$, and the number of positive movie reviews be $x \in \mathbb{N}$. Say that consumers have prior beliefs that their utility for a good will be $EU_i(t) = K_i + t$, where K_i is a constant scalar which is specific to individual i, and priors on $t \sim \mathcal{N}(0, 1)$. For example, an action movie may appeal to April more than Bob, so that $K_A > K_B$, but if the movie is well made then both are likely to enjoy the movie more. The random variable x depends on t via the joint distribution $f(x|t)$, which is a Bivariate Normal distribution known to all consumers. Then both April and Bob can observe x and then update their prior $EU_i(t)$ using $f(x|t)$. The premises of Corollary 9 hold, so we will find that $post_A(t)$ FOSDs $post_B(t)$, but the distance between the median of $post_A(t)$ and the median of $post_B(t)$ is less than $K_A - K_B$. In other words, April still prefers the movie over Bob, but they are closer to a consensus after counting reviews than before.

2.5.2 Public opinion about public opinion

A continuum of politicians are each attempting to place themselves at what they individually perceive to be the median of the distribution of public opinion regarding a question.[7] Say that initially, voters' positions are represented by points $t \in (-\infty, \infty)$ along the political spectrum, and $g_i(t)$ is the PDF of voters along that political spectrum as perceived by politician i. Further, assume that $g_i(t)$ is of the form $\mathcal{N}(i, 1)$, meaning that each politician believes that public opinion is normally distributed around his current position. New information comes in the form of a likelihood function $f(x|t)$ which is log-concave in t. This is exactly the setup in Theorem 8, and so if new public information appears regarding the distribution of opinion among the populace, then there is only one value t' such that if $i = t'$ then politician i will not change his position. All politicians with $i < t'$ will find that their median shifts upward to a point between i and t', while those with $i > t'$ will find that their median shifts downward to a point between t' and i.

If candidates believe that they should position themselves as close to the median as possible, then they may initially place themselves at disparate points along the political spectrum, because they have different perceptions of the distribution of public tastes. But as more information appears, we will see the positions of the candidates converge toward one point.

2.6 General updating operations

Say that a decision maker would like to know the likelihood of $x \in \mathbb{R}$, and toward that end two different authorities each present her with a different PDF for x, $f(x)$ and $g(x)$. The decision maker herself can only have one set of beliefs, so she must

[7] The assumption that a candidate should place himself at the median is prevalent in the Political Science literature, but note the caveats of Section 5.6.1.

amalgamate $f(x)$ and $g(x)$ into one posterior PDF. For example, it may be reasonable for the decision maker to simply take the average of the two densities:

$$post(x) = \frac{f(x) + g(x)}{2}. \tag{2.7}$$

Now say that the second authority has two distributions to choose from, $g_1(t)$ and $g_2(t)$. The same question as above can be asked: if $g_1(t)$ MLRPs $g_2(t)$, then what sort of updating operations will preserve the MLRP? This section shows that the result that the Bayesian updating operation preserves the MLRP can be expanded to a class of updating operations which I dub 'weighted Bayesian rules'—but this class does not include many reasonable updating operations such as the averaging operator above. We are guaranteed that if an updating operation that is not in this class is used, then there is some public information which will destroy the order of the posteriors, using the MLRP ordering. Meanwhile, the averaging operator of Equation 2.7 is a member of the class of functions which preserve FOSD, but which may destroy the MLRP ordering.

2.6.1 Operators which preserve the MLRP

Recall from Definition 2 that the updating operator must be rescaled to be a true PDF. However, the following statement tells us that we can simplify the problem by ignoring the rescaling.

Proposition 10 *The posterior distribution $post(f, g_1, op, x)$ MLRPs $post(f, g_2, op, x)$ iff the function $op(f, g_1, x)$ MLRPs the function $op(f, g_2), x$.*

That said, we can consider arbitrary positive functions without regard to whether they integrate to one. Then:

Lemma 11 *A continuous operator preserves the MLRP iff it is of the form*

$$op(a, b) = t(a) \cdot b^p,$$

with $p > 0$, and $t(\cdot)$ any transformation.

The fact that the first argument may undergo any transformation is not surprising: an operator preserves the MLRP if $op(f(x), g_1(x))$ MLRPs $op(f(x), g_2(x))$ for any pair of gs which satisfy the MLRP and *any* density $f(x)$. So naturally, if $op(\cdot,\cdot)$ preserves the MLRP given $f(x)$, it also preserves the MLRP given $t(f(x))$.

More interesting would be to restrict ourselves to operators which preserve the MLRP symmetrically: $op(f, g_1)$ MLRPs $op(f, g_2)$, and also $op(g_1, f)$ MLRPs $op(g_2, f)$.

Theorem 12 *A continuous operator preserves the MLRP symmetrically iff it is of the form*

$$op(a, b) = a^p \cdot b^q,$$

with $p, q > 0$.

Proof: In this case, we have two conditions: $op(a, b) = t(a) \cdot b^q$ and $op(a, b) = a^p \cdot t(b)$. Any operator which satisfies both of these conditions must take the form $op(a, b) = a^p \cdot b^q$. $\diamondsuit$

So the MLRP is symmetrically preserved in all cases only when decision makers update based on a monomial operator. The Bayesian updating operator, $op(a, b) = a \cdot b$, is the special case where $p = q = 1$.

The exponents p and q allow the decision maker to place more or less weight on the distributions $f(\cdot)$ and $g(\cdot)$. Say that $g(x)$ is a single-peaked distribution; the convex transformation of squaring would exacerbate the peak, reducing the variance of the distribution. In the context of receiving information from an advisor, this means that $op(f, g, x) = f^1(x)g^2(x)$ places more weight on the second advisor's claim and puts more of the posterior density around the center of that advisor's distribution. In a similar manner, $op(f, g, x) = f^{1/2}(x)g^1(x)$ discounts the first advisor's advice. At the extreme, $op(f, g, x) = f^0(x)g^1(x)$ is a long way of writing $op(f, g, x) = g(x)$, ignoring the input of the function $f(\cdot)$ entirely. This justifies the use of the term 'weighted Bayes's rule' to describe this class of updating operators, where the weights are the exponents p and q.

But the set of weighted Bayes's rules is a small subset within the class of possible methods of updating, excluding such intuitive methods as averaging and convolution. If the reader does not believe that preserving the MLRP is essential, then he or she should take the methods discussed earlier in this chapter with a grain of salt, since they describe only characteristics of Bayesian updating and slight variants thereof. But if the reader believes that the MLRP *should* be preserved by new information, then the above shows that the reader must also believe that people use Bayesian updating (possibly weighted) to assimilate new information.

2.6.2 Operators which preserve single-crossing

Here is a characterization of the set of updating operators which lead to posteriors ordered by FOSD:

Proposition 13 *Given any pair of priors g_1 and g_2 which satisfy single-crossing, and any function f, and an updating operator $op(f, g)$ which is monotonically increasing in f and g, the operator $op(\cdot, \cdot)$ provides posteriors ordered by FOSD iff the function is of the form*

$$op(f, g) = t(f) + qg,$$

where $t(\cdot)$ is a transformation function of any form, and q is any positive constant.

This also describes the set of updating operators which preserve single-crossing:

Lemma 14 *An updating operator $op(f, g)$ which is monotonically increasing in f and g preserves single-crossing iff the function is of the form*

$$op(f, g) = t(f) + qg,$$

where $t(\cdot)$ is a transformation function of any form, and q is any positive constant.

Notice that if $g_1(x) = g_2(x)$, then $t(f(x)) + qg_1(x) = t(f(x)) + qg_2(x)$, meaning that the point at which the priors cross is also the point at which the posteriors cross.

That is, if decision makers use an updating operator which preserves single-crossing, no new news can move the point of crossing.

2.6.3 Operators which preserve FOSD

The set of operators which preserve FOSD matches the set of operators which preserve single-crossing.

Theorem 15 *Within the class of operators $op(f, g)$ which are monotonically increasing in both arguments, an operator preserves FOSD iff it is of the form $op(f, g) = t(f) + qg$, where $t(\cdot)$ is a transformation function of any form and q is a positive constant.*

Thus, the only updating method which preserved first-order stochastic dominance in all cases is that of taking the weighted mean of the two sets, perhaps transforming the first before knowing the second. As established in section 2.3.2, Bayesian updating is not in this set.

As we did with the MLRP, it is worth considering the problem of symmetrically preserving FOSD as well.

Theorem 16 *Within the class of operators $op(f, g)$ which are monotonically increasing in both arguments, an operator symmetrically preserves FOSD iff it is of the form $op(f, g) = pf + qg$, where p and q are arbitrary positive weights.*

Proof: We require both $op(f, g) = t(f) + qg$ and $op(f, g) = pf + t'(g)$. The only operators which satisfy both of these conditions are those listed in the theorem. $\diamondsuit$

Recall from Section 1.3.1.1 that these updating operators could be used to describe Cournot learning or fictitious play models.

2.6.4 Preserving both the MLRP and FOSD

Both the MLRP and FOSD are preserved in the case of only one updating operator—the trivial operator where new information is ignored entirely.

Lemma 17 *The only updating operator which preserves both the MLRP and FOSD is:*

$$op(f(\cdot), g(\cdot), x) = g(x).$$

Proof: This is the only operator which satisfies both Lemmas 11 and 15.$\diamondsuit$

There is no function which symmetrically preserves both the MLRP and FOSD. The definitions given here are strict, but can be weakened: the ratio in the definition of MLRP must be increasing, which can be weakened to nondecreasing; and the comparison in the definition of FOSD may be weakened from 'less than' to 'less than or equal to'. Then the class of updating operators which symmetrically preserve both the weak MLRP and weak FOSD becomes a large one:

$$op(f(\cdot), g(\cdot), x) = h(x),$$

where $h(x)$ is any arbitrary PDF which depends on neither $f(x)$ nor $g(x)$.

2.7 Conclusion

This chapter discussed the result that Bayesian updating preserves the monotone likelihood ratio property. This means that if a pair of priors satisfy the MLRP, and both are updated with the same likelihood function and the Bayesian updating operation, then the two posteriors will also satisfy the MLRP. It is necessary and sufficient that the priors satisfy the MLRP for the posteriors to satisfy first-order stochastic dominance. Using a simple redefinition of the axes, we also find that if the priors are translations of one another, and are updated using a distribution which is appropriately log-concave, then the posterior medians (or any other order statistic) will be ordered.

But (weighted) Bayesian updating is necessary and sufficient for all of the above results. If a modeler believes that the MLRP is always preserved when people update their priors, or that posterior medians should be ordered as discussed above, then the modeler must describe people as using Bayesian updating. If a modeler believes that people do not use Bayesian updating, then he must also believe that the ordering among people's beliefs can be destroyed by certain pieces of new information.

Similarly, the class of averaging operators preserves first-order stochastic dominance. A researcher who desires that FOSD be preserved must assume an averaging operator, and a researcher who does not assume an averaging operator must accept that the FOSD ordering will not always be preserved.

Chapter 3

Information provision by one biased advocate

3.1 Introduction

This chapter discusses models of information revelation by biased advocates, such as lobbyists or lawyers. It takes two approaches. The first is to make the typical assumptions of Game Theory, in which judges can predict how advocates will react to their demands, and advocates take into account what information opposing advocates provide in deciding what they themselves should present. This approach requires minor assumptions to find that the outcome will always be that of full information revelation.

The second is to assume that the judge is an automaton, and look at what information the advocate will present that puts his cause in the best light. There are a few types of information that an advocate could give—experimental results, statistics, or facts about a legislator's constituency—and so there are a variety of models which will be considered here.

Having assumed that the judge is an automaton, unless there exists a monotonic transformation from the judge's utility from an outcome to that of the advocate (that is, the advocate is in agreement with the judge), it is easy to construct cases where only partial information will be revealed.

3.1.1 Prior models and types of advocates

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 [17], by Dewatripont and Tirole [18], and in the literature based on these papers. However, they are not discussed here.

This chapter 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 [36], Austin-Smith [3], and Calvert [10] model this type of advocate. These models will be discussed below (although Calvert is left to the footnote on page 40).

Glazer and Rubenstein [23] 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 chapter and those cited above includes an honesty rule of some sort, stipulating that advocates may obscure the facts or 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 General framework

This chapter will present a number of models, based on a number of specific assumptions and underlying spaces. In this section, I will present the notation and framework with which all of these models will be described. Since this is a general description, most of the spaces involved ($\mathcal{W}$, $\mathcal{S}$, and $\mathcal{A}$) will not be specified until the examples that follow. The power of Theorem 18 (page 33 is that it works with virtually any specification of these spaces.

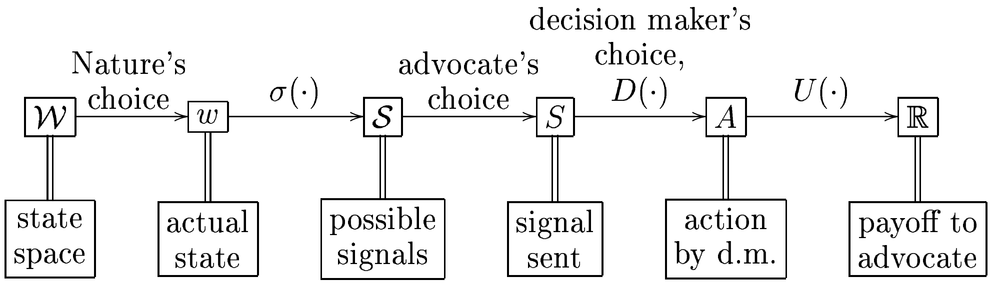

Figure 3.1: From state of the world to advocate's payoff

The process is a series of mappings, from states of the world to signals to actions to utilities. The full process is depicted in Figure 3.1. First, Nature picks one state of the world, w, from all true states, $\mathcal{W}$, and reveals that state to the advocate but not the decision maker. Given a state of the world, there is some set of signals $\mathcal{S}$ which the advocate may send.

Assumption 5 *Assume that there exists a mapping $\sigma : \mathcal{W} \to \mathcal{S}$, which is common knowledge. The inverse mapping $\sigma^{-1} : \mathcal{S} \to \mathcal{W}$ is also common knowledge. Also assume that $\sigma^{-1}(S)$ is closed for all $S \in \mathcal{S}$.*

So $\sigma(w)$ is the set of signals which may be sent given that w is the true state of the world, and given a single signal S, $\sigma^{-1}(S)$ is the set of states in which the advocate could have sent the signal S. Notice that σ may embody an honesty rule, but I only assume that it is some fixed, common knowledge mapping.

From these possible signals, the decision maker may present any one signal S to the decision maker, who chooses some action A for any given signal S, from the space of possible actions $\mathcal{A}$. The decision function which makes this mapping, $D(\cdot) : \mathcal{S} \to \mathcal{A}$, is selected by the judge, and may include strategic considerations.

The advocate has some utility associated with each action, $U(A) : \mathcal{A} \to \mathbb{R}$, which he hopes to maximize.

3.2.0.1 Full information

For the question, 'when will full information be revealed?' to be a coherent one, we need to ensure that something like full information can always be presented. The following definition and assumption allow this.

Definition 6 *A signal S is* fully informative *iff $\sigma^{-1}(S)$ is a set with only one element.*

Assumption 7 *For any given w, there exists a signal S^w such that $\sigma^{-1}(S) = \{w\}$.*

That is, if the signal could only have been sent in one state of the world, then that signal tells the decision maker the state with certainty. For any state of the world, the advocate is capable of sending such a decisive signal.

I assume that the decision maker can do no better from the advocate than being told the state of the world with certainty.

Assumption 8 *The decision maker's utility is maximized when the advocate reveals S^w.*

This is the only assumption about the decision maker's goals that we will need.

From here, an example is in order.

3.3 Milgrom and Roberts's model

This section describes a model best suited for the presentation of evidence, such as facts at a trial or statistics about a proposed policy. Barring a few modifications made for expositional clarity, the model is from Milgrom and Roberts [36].

The state of the world, $w \in [0,1]$, is known to the lobbyist but not the decision maker. The lobbyist may send to the decision maker a signal consisting of a subset of $[0,1]$, $S = [a,b]$, and is bound by honesty to reveal only a subset S such that $w \in S$. That is, σ maps from w to all closed subsets of $[0,1]$ which contain w. The inverse, σ^{-1}, therefore maps from any closed interval to itself.

$\mathcal{W}$	$\equiv$	$[0,1]$
σ	$\equiv$	$w \rightarrow$ sets of $[0,1]$ which include w
$\mathcal{S}$	$\equiv$	sets of $[0,1]$ which include w
$\mathcal{A}$	$\equiv$	$\mathcal{W}$
$U(A)$	$\equiv$	A

Figure 3.2: Unidimensional model based on that of Milgrom and Roberts

The decision maker's action will be to select a value of w, based on her beliefs. That is, $\mathcal{A} = \mathcal{W}$. Define her loss function to be $-|A - w|$, so she wants to select A as close to w as possible.

The lobbyist's utility is simply $U(A) = A$, meaning that he hopes for as large a decision as possible.

The model, within the framework presented here, is summarized in Table 3.2. All of the elements from Figure 3.1 have now been defined, save for the strategic choices of the actors.

Milgrom and Roberts show that there is an attractive equilibrium in which the decision maker uses the decision function $D(S) = \min(S)$, meaning that she assumes the worst possible state as measured by the advocate's utility.

Given this action, the advocate will send a signal of the form $[w, y]$, where y is any value greater than w. If he sends a signal $[x, y]$ with $x < w$, then he will get a lower payoff than if he had revealed $[w, y]$. He is barred from giving a signal $[x, y]$ with $x > w$, since in this case $w \notin [x, y]$. Therefore, the decision and payoff will be equivalent to the decision and payoff that would be reached if the advocate had given full information.

3.3.1 How robust is this equilibrium?

3.3.1.1 Robust to changes in the model

Milgrom and Roberts's equilibrium is entirely independent of the information, signal, and action spaces. The decision maker can always comport herself as if she has received the most pessimistic full information signal consistent with the data. Formally:

Theorem 18 *Assume that the conditions of Assumption 5, Assumption 7, and Assumption 8 hold. Then there exists a Nash equilibrium whose payoff to the advocate is equal to the payoff he would get from revealing full information.*

Assumption 5 assumes that the decision maker has some way of mapping from signals back to states of the world, which is what we would expect any decision maker to be doing anyway after she receives her signal. Assumption 7 says that there is always full information to be revealed, and Assumption 8 says that full information is a good thing. These are relatively uncontroversial statements, set in a framework which is general enough to describe a variety of models from the literature, so we would expect this to be a rather robust result. If we hope to write a model where full information is not revealed, then we will not get there by creative specifications of the signal, state, or action spaces.

3.3.1.2 Not robust to changes in equilibrium concepts

Say that instead of the state of the world being a point $w \in [0, 1]$, the state is a point $w \in \mathbb{R}$. The decision maker could still insist on the assume-the-worst strategy given a report from the advocate that $w \in [a, b]$, for any real a and b, or even a report $w \in [a, \infty)$. As above, she may choose the minimum, $w = a$.

But what should the decision maker do if the advocate gives the no-information report $w \in (-\infty, \infty)$? The assume-the-worst strategy dictates that the decision maker should choose $-\infty$. This is indeed an equilibrium: if the advocate knows that the decision maker will choose $-\infty$ given a report of $(-\infty, \infty)$, then he will never dare make that report. But by choosing $-\infty$, the decision maker is giving herself an infinite loss, while deciding on any real number would give her a finite loss.

On a technical level, this means that the equilibrium is subgame perfect, but is not a sequential Nash equilibrium.

On a colloquial level, the gist of these refinements is that an advocate may 'call the decision maker's bluff' and present partial information, or may simply forget to hand in a report. If the advocate commits such a failure, then the decision maker would be throwing away useful priors and any information the advocate does provide by continuing to follow her proclaimed strategy.

3.3.1.3 Possible responses

From here, there are three directions we may take: the first is to presume that the decision maker can make good on her threat to select $A = -\infty$ if need be, in which case our models may ignore the information revelation process entirely. A biased information provider can always be coerced to provide full information, and so the advocate and decision maker may be treated as a unit. For some cases, this is appropriate, but for others this is a little too optimistic.

The second option, chosen by some subset of the literature, is to assume that the decision maker is not strategic: she calculates her posterior beliefs given whatever information the advocate is willing to give, and takes appropriate action. This direction will be followed in Section 3.4

The third option is to make fundamental changes to the standard framework. This approach is taken in Chapter 4.

3.4 A survey of models of information provision to automata

This section gives a survey of some selected models of information provision by one biased advocate. All of them assume that the information provider is a non-strategic automaton, which avoids the full information result of Theorem 18, even though it could be applied to all of the models below if the decision maker were allowed to act as a strategic agent. The focus will be on lobbying applications.

The result is that without judges who are capable of strategically interpreting information, partial information is possible in almost all cases. The exception is when there is a monotonic mapping from the utility from a decision to the information provider to the utility from the same decision to the decision maker. In this case, the advocate becomes an extension of the decision maker, and so will comply with the decision maker's desire for full information.

Since partial information revelation is now possible, let $S_1 \succeq_I S_2$ signify that signal one is more informative than signal two. The exact meaning of this term will be given on a case-by-case basis. Also, let S_F signify full information and $S_\emptyset$ signify no information.

3.4.1 Revelation of a PDF

Rosenthal [48] and Goldstein [24] report that many lobbyists describe their jobs as that of producing 'traceability'. Government policy is far too complex for the average individual to understand, and it is easy to lose the connection between the passage of a law and its effects. The role of the lobbyist is to make clear that connection, so that voters know that a change can be traced to a bill and the bill can be traced back to their representative, and representatives know that voters know.

In abstract terms, a representative is interested in placing herself at the median of (or some other statistic based on) voter preferences along some single dimension. Let the true density of preferences be $p(x) : \mathbb{R} \to \mathbb{R}$. The lobbyist has the option of approaching the representative and presenting the value of $p(x)$ for some range $x \in (a, b)$.

In a world where signals may be density functions, the question of ordering becomes nontrivial, so before discussing this model, I will first digress to the subject of Blackwell's ordering of information.

3.4.1.1 Blackwell's garbling criterion

Blackwell's [5] definition of 'more informative' was to simply say that if a decision maker's *ex ante* utility was higher given experiment A than given experiment B, then experiment A was more informative to the decision maker. He defines 'garbling transformations', and shows that in the case where there is a garbling transformation from one set of signals to another, then the pre-transformation signal set is more

$\mathcal{S}$	$\equiv$	Ω
$S_1 \succeq_I S_2$	$\equiv$	$\text{argmax}_D EU_d(D\|S_1) > \text{argmax}_D EU_d(D\|S_2)$ (or $S_2 = G_{12} \cdot S_1$)
$D(S)$	$\equiv$	$\text{argmax}_D EU_d(D\|S)$

Figure 3.3: Information is trivially more valuable if Blackwell's garbling definition applies.

valuable to the decision maker than the post-transformation set regardless of the utility function involved.

If a garbling matrix exists, then a model based on this definition of informativeness would be trivial, since 'more informative' corresponds to 'more useful to the decision maker'. The question of whether the advocate will reveal full information is then simply a question of whether the advocate's interests align with the decision maker's. However, for an arbitrary pair of information sets, a garbling matrix is unlikely to exist.

Take Ω to be any set of signals such that any pair of signals S_i and S_j are related by some garbling matrix G_{ij} (that is, $S_i = G_{ij} \cdot S_j$), and $U_d(D)$ to be the decision maker's utility from decision D. The setup is summarized in Figure 3.3.

If the advocate's utility, $U_A(D)$, is a monotonically increasing transformation of the decision maker's utility function $U_d(D)$, then the advocate will reveal as informative a signal as possible; if $U_A(D)$ is a monotonically decreasing transformation of $U_d(D)$, then the advocate will reveal as uninformative a signal as possible; and if there is no monotonic transformation, then it is unclear what the advocate will choose.

3.4.1.2 Traceability

Return to the traceability question discussed at the beginning of this section. Let the density of voter tastes over some policy dimension be $p(x)$, $x \in [0, 1]$. The legislator's only source of information about $p(x)$ is a lobbyist who may choose to reveal some but not all of the density function. There are right-leaning lobbyists, for whom $U(A) = A$,

$\mathcal{W}$	$\equiv$	prob. densities over $x \in [0,1]$
decision maker's prior	$\equiv$	$U[0,1]$
$\mathcal{S}$	$\equiv$	$2^{[0,1]}$
$S_1 \succeq_I S_2$	$\equiv$	$\|S_1\| > \|S_2\|$
$D(S)$	$\equiv$	median or mean of posterior
$U(D)$	$\equiv$	D

Figure 3.4: Traceability model

and left-leaning lobbyists, for whom $U(A) = -A$.

Garbling provides only a very partial ordering of the space of possible signals. For example, say that a left-leaning lobbyist is only willing to reveal $p(x)$ for $x < \frac{1}{3}$, while a right-leaning lobbyist is only willing to reveal $p(x)$ for $x > \frac{1}{3}$. There is no garbling that would transform the left-leaning information into the right-leaning information, so Blackwell's Theorem can not help us in ordering the distributions. Instead, I offer an alternate model here.

Let $\mathcal{W}$, the space of states of the world, be the set of PDFs over $x \in [0,1]$. The lobbyist can send information about any subset of points in $[0,1]$, or $\mathcal{S} = 2^{[0,1]}$. One possible ordering of $\mathcal{S}$ which allows any two elements to be compared would be based on the magnitude of the sets: $S_1 \succeq_I S_2$ iff $|S_1| > |S_2|$.

The posterior beliefs of the decision maker depend on her prior and the information proffered by the lobbyist. The legislator will assimilate the new information using Bayesian updating (despite the caveats of Chapter 1). For simplicity, I will assume that the decision maker has the uniform prior $p(x) = 1$, $\forall x \in [0,1]$. The lobbyist is bound to tell the truth when he does say anything, so the decision maker's posterior beliefs about those points reported on is equal to the report given. If $\neg S$ is the complement of S, then for any $x \in \neg S$, $p(x) = 1/|1 - p(S)|$; that is, the decision maker is left with a uniform posterior belief, whose likelihood equals one minus the total probability reported by the lobbyist. With the posterior in mind, the decision maker can then act by choosing the median or mean of the distribution. See Figure 3.4 for a summary of the model.

3.4.1.3 Median model

It is easy to come up with examples where information is not fully revealed by a lobbyist of either extreme. For example, say that the distribution of tastes is

$$f(x) = \begin{cases} 8x & x \in [0, \frac{1}{2}] \\ 0 & x \notin [0, \frac{1}{2}] \end{cases}.$$

This is a well-behaved distribution: it is weakly single-peaked, and the discontinuity may easily be smoothed out without changing the discussion below. The true median of this distribution is at $\frac{\sqrt{2}}{4} \approx .35$. Say that the decision maker asks a left-leaning lobbyist for information. Then he would only reveal that $f(x) = 0$ for $x \in [\frac{1}{2}, 1]$. The decision maker would conclude with the posterior

$$post(x) = \begin{cases} 2 & x \in [0, \frac{1}{2}] \\ 0 & x \notin [0, \frac{1}{2}] \end{cases},$$

whose median is at $\frac{1}{4}$.

A right-leaning lobbyist would reveal the part of the distribution in $[0, .1255]$, leading to the following posterior beliefs:

$$post(x) = \begin{cases} 8x & x \in [0, .1255] \\ 1.004 & x \notin [0, .1255] \end{cases}.$$

Notice that at the point .1255, $f(x) = 1.004$. The median of this posterior is at .502.

In both cases, the advocates chose to reveal only partial information, and the decision maker choses a location some distance from the true median. The following proposition gives the general result.

Proposition 19 *If the true distribution of $x \in [0, 1]$ is $f(x)$, the decision maker's prior is $p(x)$, and the decision maker will choose the median of her perceived posterior distribution, then a lobbyist who prefers smaller outcomes to larger will not reveal full information if there exists some interval (a, b) above the median of $f(x)$ such that $f(y) > p(y)$ for all $y \in (a, b)$, and there also exists some interval (c, d) below the* *median of $f(x)$ such that $f(y) < p(y)$ for all $y \in (c, d)$.*

The same lobbyist will reveal some quantity of information if there exists some interval (a, b) above the median of $p(x)$ such that $p(y) > f(y)$ for all $y \in (a, b)$, and there also exists some interval (c, d) below than the median of $p(x)$ such that $p(y) < f(y)$ for all $y \in (c, d)$.

The inequalities are reversed for a lobbyist who prefers larger outcomes to smaller outcomes.

3.4.1.4 Mean model

If the decision maker chooses the mean of the posterior distribution, then:

Proposition 20 *A right-leaning lobbyist will reveal full information iff $p(x)$ is nondecreasing. The lobbyist will reveal no information iff $p(x)$ is nonincreasing.*

The reverse is true for a left-leaning lobbyist.

So in both the case of selecting a mean or a median, full information is revealed only under specific conditions, and it is possible that a sole lobbyist won't reveal full information regardless of whether he wants a high or low D.

3.4.2 Austin-Smith's two-dimensional model

To conclude the section on one-advocate models, I will discuss Austin-Smith's model [3], fitting it into the framework discussed above. As with Blackwell's method of ordering information, we find that with no monotone mapping from informatveness to advocate's utility, there is no guarantee of full information revelation.

Austin-Smith's world has two states, A and B, and the decision maker has two options, A and B. The value to the decision maker of taking the correct action takes two arguments, the state and the decision. When the decision matches the state, $U(A, A) = U(B, B) \equiv 1$; when they do not, $U(B, A) = U(A, B) \equiv 0$. Prior to any information from the lobbyist, the decision maker believes the likelihood of state A is $\pi_d > \frac{1}{2}$. The decision maker maximizes his expected utility, and therefore chooses A

with certainty if the posterior likelihood of state A is greater than a half, and similarly for B.[1]

There is one lobbyist, who wishes that either A is always chosen, or that B is always chosen, depending on his persuasion. Since there is no monotonic transformation from the lobbyist's goals to those of the decision maker, Blackwell's Theorem (discussed in section 3.4.1.1) is not useful here.

The signals the lobbyist chooses from are experiments, which turn up either outcome O_1 or O_2. [2] The experiment is described by two numbers: $p \in [0,1]$ gives the likelihood that the outcome is O_2 in state A, while $q \in [0,1]$ gives the likelihood that the outcome is O_2 in state B. The signal space is thus the set of ordered pairs $\mathcal{S} \equiv (p,q) \ni p \leq q$, and can be partially ordered by $S_1 \succeq_I S_2$ iff $q_1 - p_1 > q_2 - p_2$. The fully informative signal in this context is $S_F = (0,1)$: the decision maker knows that the state is A if the experiment turns up O_1 and the state is B given outcome O_2. The fully uninformative signal here is $S_\emptyset = (\frac{1}{2}, \frac{1}{2})$, or any other signal where $q - p = 0$.

Austin-Smith shows that an experiment can induce two types of behavior in the decision maker: it can convince her to choose A with certainty, or it can induce her to choose A given the signal O_1 and B given signal O_2. Given the parameters here, notably $\pi > \frac{1}{2}$, the decision maker will always choose A given outcome O_1. The decision maker will choose B given outcome O_2 iff $P(B|O_2) > \frac{1}{2}$, which occurs iff

$$p\pi \leq q(1-\pi). \tag{3.1}$$

[1] Austin-Smith's setup bears a resemblance to that of Calvert [10], though neither paper cites the other. But Calvert's definition of bias is different. His decision maker is considering two projects, whose value is in $[0,1]$, and must choose the better of the two. An advisor will give project A a 'good' rating only if the utility from A is over a certain threshold. If that threshold is near one, then the advisor is considered to be biased against A, since A has to be almost flawless to receive a 'good' rating. But unlike Austin-Smith's model, after receiving a 'good' rating, the decision maker is guaranteed that the project is indeed a good one. Advisors therefore represent different information partitions, and it is tenuous to say that an advocate who reliably reports where a project's value is located in the partition $\{[0,.5],[.5,1]\}$ is 'less biased' than an advocate who reliably reports where a project's value is located within $\{[0,.95],[.95,1]\}$. It is also difficult to predict *a priori* which of these two partitions will be more informative to a decision maker. I therefore stick to Austin-Smith's variant of this model, where bias and informativeness are easier to define.

[2] Austin-Smith used the notation S_0 and S_1.

Assuming $p \leq q$ ensures that the signals can not be reversed, and the outcome where B is chosen with certainty can be shown to be impossible given a prior belief that the likelihood of A is greater than $\frac{1}{2}$. The *ex ante* probability of outcome A is one if Condition 3.1 does not hold, and is equal to the probability of O_2 if it does hold. The probability of O_2, from the point of view of the lobbyist with prior belief that outcome A will occur with probability π^A is

$$P(O_2|A) \cdot P(A) + P(O_2|B) \cdot P(B) = p\pi^A + q(1 - \pi^A).$$

Figure 3.5 shows that for $q = .5$ and $\pi = .6$, the probability of B is an increasing function, until $p = \frac{1}{3}$, at which point A will be chosen with certainty.

The model is summarized in Figure 3.6.

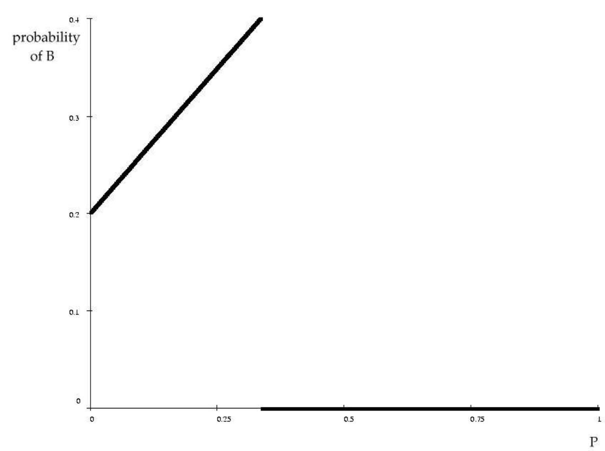

Figure 3.5: The probability of B increases with p, but then falls to zero.

The *ex ante* utility to a B-lobbyist is increasing as $P(O_2)$ increases, but only so long as condition 3.1 is met. Fixing $q = 1$, for example, as p increases from zero to one, the signal goes from being fully informative to being fully uninformative. But the B-lobbyist's utility is monotonically increasing until $p = (1-\pi)/\pi$, at which point utility drops to zero. As shown in Figure 3.5, a monotonic change in informativeness (by any measure, not just $p - q$) does not imply a monotonic change in utility. This allows the optimal solution to be neither full nor partial revelation; for the B-lobbyist,

$$
\begin{array}{ccl}
\mathcal{W} & \equiv & [A, B] \\
\mathcal{S} & \equiv & (p, q) \\
S_1 \succeq_I S_2 & \equiv & |q_1 - p_1| > |q_2 - p_2|, \text{ or } \frac{q_1}{p_1} > \frac{q_2}{p_2} \\
D(S) & \equiv & \begin{cases} A & p\pi > q(1-\pi) \\ A & p\pi \leq q(1-\pi) \ \& \ O_1 \\ B & p\pi \leq q(1-\pi) \ \& \ O_2 \end{cases} \\
U(D) & \equiv & \begin{cases} 0 & \text{if } D \neq \text{lobbyist type} \\ 1 & \text{if } D = \text{lobbyist type} \end{cases}
\end{array}
$$

Figure 3.6: Austin-Smith's model

the optimum is in fact $S = (\frac{(1-\pi)}{\pi}, 1)$.[3]

The A-lobbyist will choose p and q to fail to satisfy condition 3.1, so that A will be chosen with certainty. $S_\emptyset = (\frac{1}{2}, \frac{1}{2})$ achieves this goal: since neither signal O_1 nor O_2 imparts any information, the decision maker is left with her prior that $\pi > \frac{1}{2}$, and so will choose A with certainty.

3.5 Conclusion

This section discussed the amount of information revealed by biased advocates. We would like advocates to consistently reveal unbiased information, or enough information that an unbiased decision is possible. To achieve this, it is sufficient that the decision maker can design a monotonically increasing mapping from information provided to the advocate's utility. But it is not always possible to design such a mapping—especially if one wishes to ensure robustness properties such as equilibria being sequential Nash—and so some situations will lead to full information revelation as desired, while in others, the advocate's bias will influence the final policy outcome.

[3] In equilibrium, the decision maker chooses B when indifferent. Otherwise B would signal $((1-\pi)/\pi) - \epsilon$, and no equilibrium would exist given continuous signals.

Chapter 4

Information provision by two biased advocates: A model of the perception of causation

4.1 Introduction

This chapter presents a model of a trial. Two advocates present arguments to a decision maker,[1] who then decides which arguments are correct, and the severity of the punishment (if any).

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.

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

[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 chapter, 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.

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 and theoretically robust: opposing advocates will always reveal full information. A full discussion is in Section 4.3.2.

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 4.3.2 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 explanations. 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, 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 4.2 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. As in the one-advocate case of Chapter 3, Section 4.3.2 proves that under the simplest of assumptions, full information is always revealed by two advocates in a data-based framework. Section 4.4 gives an overview of an alternate story-based model and its elements, and Section 4.5 describes the causal weights and probabilities alluded to above. Section 4.6 then asks which stories the advocates will choose to reveal.

4.2 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.

4.2.1 Prior work on causality

4.2.1.1 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 [53] 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 [26] concept of causality, also known as 'screening off', is the basis for the statistical causality literature (e.g., Perl [44]). 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 $t-1$" ([26], 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.

4.2.1.2 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 ([41], [42], [43]). 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.

4.2.1.3 Lawyers

Legal philosophers are familiar with the idea of separating necessary causes from sufficient causes, e.g. Honoré [30]. 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.

4.2.2 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.

4.2.2.1 Appellate lawyers exist

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

The distinction is articulated in the voluminous rules of evidence. Consider the situation in which a defendant gave a clearly sincere confession to a crime before having his Miranda rights explained to him. The fact that he confessed may easily reach the judge's ears, but for a number of reasons, it is not admissible evidence. A Platonic truth-seeking judge would be unconcerned: she knows beyond a reasonable doubt that the true state is that the defendant is guilty, and could ignore any additional information the prosecutor may present. But in reality, the prosecutor must present a case using admissible evidence, and the judge must decide whether the case proves guilt. If there is evidence that the judge based her decision on knowledge of the truth rather than on the prosecutor's case, then the decision may be overturned by an appellate court.

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.

4.2.2.2 The story told by the law texts

For a look at how lawyers argue, I found the two books on trials from USC Law's Fall 2002 reading list, by Mauet [35] and Murray [39]. Mauet also appears on what other reading lists I could find (those of the law schools at NYU and the University of Chicago), and is in its sixth edition, implying that it has come recommended by many law professors in the past.

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.

As any good storyteller (or anyone with a long-winded friend) knows, the story can be ruined if too many extraneous or tangential facts are included. Mauet [p 516] seems to have this in mind when he advises the law student:

> 1. Do not overprove your case. [...] In general, calling a primary witness and one or two corroboration witnesses on any key point is enough. It's usually best to make your case in chief simple, fast, and then quit while ahead.

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."

These passages are representative of what actual law students are advised to do in selecting a trial strategy. If we wish to model a trial process as it is actually executed, then we must take into account the fact that the lawyer is a storyteller, and that revealing full information is anathematic to the process of telling a convincing story.

4.2.2.3 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 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.

4.3 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 Chapter 3. It provides motiviation for the main results, which appear later in this chapter.

4.3.1 Definitions

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

Definition 9 *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* $U_2(D_2) > U_2(D_1)$*.*

So 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 [33]. 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 signal of the other player can be ignored and questions of game theory become irrelevant.

Definition 10 *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 the sake of concreteness, I present the following game between two opposing advocates. 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 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 $G_1 = I$.

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.

4.3.2 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 21 *In a strictly competitive game with a unilateral strategy, players will always receive the unilateral payoff.*

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 [33]) 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, or 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. The only assumption inherent to this functional form is that she bases her decisions only on the information set $\mathcal{G}$.

Full information revelation is a unilateral strategy by most definitions of the term (including Definition 6). 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 21, 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 that is the outcome that will prevail if the conditions of Lemma 21 hold.[2]

[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.

It can be shown that every single example given in Chapter 3 for one player fits the description here.

Robustness, and alternatives Lemma 21 has only two premises: that the game is strictly competitive and that a unilateral strategy exists. We needed no assumptions about the structure of information that may be presented, the order in which players move, the decision maker's method of processing information, or other typical details of the game form.

So the theorem has few moving parts that could break, and should apply in a wide range of situations. Yet, as will be discussed in Section 4.2.2.2, advocates typically tell only one story each, revealing only the subset of information apropos to the story they tell. What this means is that in these situations of partial information revelation, the judge takes into account more than just the list of facts, but the context of those facts, such as who presents which fact.

One approach we may take is to presume that the judge calculates the likelihood of certain events or stories depending on who presents them. Phrasing the main question in the terminology of Chapter 1, what updating operators would admit the revelation of only partial information? Recall the definition of a unilateral strategy: that if one player plays that strategy, the outcome is the same regardless of what the other player does, and this is true for both players. In other words, if $\mathcal{I}$ is a unilateral strategy, then $op(a, \mathcal{I}) = op(\mathcal{I}, b)$ for all a and all b.

For a partial information revelation outcome to prevail in a strictly competitive game, it must be that no unilateral strategy exists. In other words, $op(a, \mathcal{I}) \neq op(\mathcal{I}, b)$ for some a and b. All symmetric updating operators therefore have a unilateral strategy, and some subset of the asymmetric updating operators do not.

This is a problem because any data-oriented method of determining the state of the world will be symmetric: the fact is the same regardless of the teller, so the decision maker shouldn't care who told her which fact. The model here reconciles this by presenting a natural model of the decisionmaking process which allows the judge to decide the state of the world using a symmetric operator such as Bayesian updating, but to make asymmetric decisions within certain states of the world, after she has her posteriors in hand.

4.4 The model

As discussed in Section 4.3.2, 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 4.2.2.2. Therefore, this section proposes a model which adds causal machinery to the estimation of subjective probabilities, resulting in theoretical equilibria that match the actual stable state much more closely.

The focus of this model will be upon a civil trial. Unlike other types of trial, judgments are usually in dollars, which can easily be represented by a continuous variable, and there is no requirement that judgments be 'beyond a reasonable doubt'. The additional considerations of a criminal, appellate, or Supreme Court trial may easily be added to the model here.

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.

4.4.1 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 $\mathcal{SC}$.

4.4.1.1 Stories as legal arguments

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 [52]) 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.

4.4.2 The judge's beliefs

The judge will develop subjective beliefs about the probability that a given story or cause is true. Here, I lay the framework for describing the judge's beliefs, and over the course of this section, will describe the spaces and objects on which these beliefs act.

Definition 11 *Let $J : \mathcal{S} \times \mathcal{SC} \to [0,1]$ 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.

A reasonable judge will have beliefs which have some structure to them, meaning that there will be certain restrictions on the function $J(\cdot,\cdot)$. The first is that the judge should have true beliefs about the stories:

Assumption 12 $J(S) = P(S)$ *for all* $S \in \mathcal{S}$.

4.4.3 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. All parties have perfect priors over the judge's subjective probabilities. That is, any impartial observer could predict the judge's subjective opinions regarding the probability that S_i is both true and sufficient to cause E; and the verdict, $v(S_i) \in \mathbb{R}$, which will be selected if S_i is the only story which is true and sufficient to cause E.

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 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 17 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,

[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.

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.

4.5 Causal weights and probabilities

This section will discuss the calculation of the subjective probability that $s(S)$ is true, the weight the judge places on a causation, and the final verdict given that some number of stories are true.

This section will conclude by describing the expected final verdict.

4.5.1 The probability that a sufficiency claim is true

I will begin by describing the subjective probability $J(s(S))$ for syllogisms. Then, I will generalize to real-world histories. A legal argument will include elements of both logical classifications or proofs, and of storytelling about the history of an event.

4.5.1.1 Syllogisms

Consider the following three-cause story to explain the event $E =$ Socrates is a mortal: $S_1 = \{C_1 =$ Socrates is a man, $C_2 =$ all men are mammals, $C_3 =$ all mammals are mortals$\}$. Assume for now that S_1 is the only mechanism by which E could be true. It is clearly sufficient to explain E, in that if all its elements are given to be true, then E is true with certainty. Now consider $S_2 = \{C_1 =$ Socrates is a man, $C_2 =$ all men are mammals$\}$. Story S_2 is sufficient to explain E if and only if C_3 is true, and so the probability that S_2 is sufficient to explain E equals the probability that S_1 is true given S_2, i.e., $J(s(S_2)) = J(S_1|S_2)$.

Now say that in addition to story S_1, E could also be true via story S_3: $\{C_1 =$ Socrates is a man, $C_2 =$ all men are mammals, $C_4 =$ all mammals are living things, $C_5 =$ all living things are mortal$\}$. Now S_2 is sufficient to explain E if either S_1 or

S_3 are true given S_2, and so the probability that S_2 is sufficient to explain E is now $J(s(S_2)) = J(S_1 \cup S_3 | S_2)$.

Finally, consider $S_4 = \{C_2$ all men are mammals, $C_3 =$ all mammals are mortals$\}$. One manner of thinking says that since $S_4 \not\subset S_3$, the probability that S_4 is sufficient to explain E should not be based on S_3. After all, if an advocate presented an encyclopedia, facts C_1 through C_4 would all be in there somewhere, but we would not say that the advocate has provided an explanation for E. Therefore, $J(s(S_4)) = J(S_1 | S_2)$.

Alternatively, we could allow oversufficient evidence, meaning that S_3 is apropos even though $S_4 \not\subset S_3$. If we allow this, then $J(s(S_4)) = J(S_1 \cup S_3 | S_4)$.

4.5.1.2 A full information definition of the probability of sufficient causation

Outside of syllogisms, the statement 'S is sufficient for E to occur' is impossible to quantify. Even if we have a firm grasp on the causal mechanism, there are still contingencies where S occurs but E does not, e.g., a meteor may fall from the sky and crush all involved.

Let U be a universal story, listing *all* of the events which are necessary for us to completely guarantee that an event will be true. Such a set could only be known by an omniscient observer. But if we allow that such a set exists, then the discussion regarding syllogisms above applies directly. Let $\mathcal{U}^E$ be the set of universal stories which explain E.

For example, a lawyer may tell a short story S_5 to explain a driver hitting a pedestrian, saying only that the driver hit a patch of ice and slid into the pedestrian. But this simple story is part of an infinitely larger sequence of events, U_1, which includes such causes as: the pedestrian was out shopping because she had just won the lottery, the driver was not skilled with handling ice, the driver was going home late that day because the copier at work broke, and of course, a meteor did not fall from the sky and crush all involved. The probability that S_5 is sufficient to explain the accident requires that all of the 'helper events', $U_1 \backslash S_5$ are all true as well. That is, $J(s(S_5)) = J(U_1|S_5)$. There are a multitude of other sets of helper events which would be sufficient to cause the accident with probability one, and which are each a superset of S_5; let these be $U_1, \dots U_{100}$. If any one of them is true, then the accident would have occurred, so $J(s(S_5)) = J\left(\bigcup_{i=1\dots100} U_i|S_5\right)$

The following definitions derive directly from the discussion above.

Definition 13 *Allowing for oversufficiency,* $J(s(S)) \equiv J(\bigcup_{U\in\mathcal{U}^E} U|S)$

Definition 14 *Disallowing oversufficiency,* $J(s(S)) \equiv J\left(\bigcup_{\{U|U\in\mathcal{U}^E, S\subseteq U\}} U|S\right)$

4.5.1.3 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 15 *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,* $1 - J(S, s(S))$.

If the judge is omniscient and knows U^E, then in the respective cases of allowing and disallowing oversufficiency,

$$PA(S) = P\left(S \cup \bigcup_{U\in U^E} U\right) \text{ or } PA(S) = P\left(\bigcup_{U|U\in U^E, S\subseteq U} U\right).$$

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

4.5.2 Weights on sufficient causation

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 16 *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 $S_i \in \mathcal{S}$:*
(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 16 it is possible that only one story is accepted as an explanation for E. In this case, $W(S_1, \{S_1\}) = 1$.

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:

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

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.

4.5.2.1 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:

$$V(\mathcal{S}) = \sum_{i=1}^{n} W(S_i, \mathcal{S}) \cdot v(S_i). \tag{4.1}$$

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 accepted stories.

4.5.2.2 Opinion formally defined

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

Definition 17 *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 subset 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.

If a judge seems biased, we need to test that claim using cases which minimize the probability that competing stories are simultaneously true, so that the ruling will be a question of evaluating the likely state of the world instead of the judge's opinion.

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

$$\sum_{\mathcal{S}_i \in 2^S} \left[\left(\prod_{S \in \mathcal{S}_i} PA(S) \right) \left(\sum_{S \in \mathcal{S}_i} v(S) W(S, \mathcal{S}_i) \right) \right]. \tag{4.2}$$

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 4.1.

The defendant wants to minimize Expression 4.2, while the plaintiff wants to maximize it.

4.5.3 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.

4.5.4 An example from advertising

Having introduced all of the concepts, I present an example from the realm of advertising.

Apple wishes to convince the listener of the event $E_a =$ 'an Apple computer will make you happy'; while Dell wishes to convince the listener of $E_d =$ 'a Dell computer will make you happy'. Take the universe of possible claims to be as follows:

$S_a =$ Apple computers are cute.

$S_b =$ Apple computers have a superior bus architecture.

S_d = Dell computers are inexpensive.

Dell and Apple may each present as many stories as they wish, each consisting of some subset of the above causes. Let us take the status quo to be Situation 4.3 from Figure 4.1. Let us presume the status quo is that Apple tells story S_a and Dell tells S_d.

Since there are two events to be accounted for, I specify the event the story is said to be sufficient to explain, e.g., $s(S_a, E_a)$ should be read, 'S_a is sufficient to explain E_a'. Similarly, $W(()S_a, \{\mathcal{S}\})$ will be extended to $W(()S_a, E_a, \{\mathcal{S}\})$.

Let us take all of these stories to be true with certainty, i.e., $P(S_i) = 1$, for all i. This means that $PA(S_i) = J(S_i, E_{\{a,d\}})$.

There are two pairs of arbitrary elements which the decision maker chooses. The first pair is $J(s(S_a, E_a))$ and $J(s(S_d, E_d))$: the probability that the story presented by the advertiser is sufficient to lead to the computer making the buyer happy. For artists, $J(s(S_a, E_a))$ tends to be high and $J(s(S_d, E_d))$ tends to be low; the reverse is typically true for frugal businessmen.

The second pair of arbitrary elements are the weights $W(S_a, E_a, \{S_a, S_d\})$ and $W(S_d, E_d, \{S_a, S_d\})$. That is, given that both stories are accepted as true and sufficient, what weight does the consumer place on each?

The verdict given that S_a is accepted is that the consumer buys an Apple computer and not a Dell with certainty; the verdict given S_d is similarly defined. Let W_a signify $W(s(S_a, E_a, \{S_a, S_d\})$ (and similarly for W_d). Then a linear combination such as $W_a v(S_a) + W_d v(S_d)$ results in the purchase of an Apple with probability W_a and a Dell with probability W_d. The model here assumes that $W_a + W_d = 1$, but if that assumption is dropped and $W_d + W_a < 1$, then the decision maker buys no computer with probability $1 - W_d - W_a$; if $W_d + W_a > 1$, then she buys both computers with probability $W_d + W_a - 1$.

It is difficult to specify what W_a or W_d will be, but a few reasonable guesses will allow comparative statics to be made.

Apple's stories	Dell's stories	P(Apple)	
S_a	S_d	.8	(4.3)
S_a, S_b	S_d	.7	(4.4)
S_a	S_b, S_d	.9	(4.5)

Figure 4.1: Some stories and the likelihood of buying an Apple induced therefrom

4.5.4.1 Comparative statics

Figure 4.1 shows some sample outcomes given that the advertisers told some set of stories. Situation 4.3 is the status quo: Apple tells its 'user-friendly' story, while Dell tells its 'inexpensive commodity' story. Alternatively, Apple could tell its user-friendly story and the additional story about its bus architecture, as in Situation 4.4. The technical story may be detrimental to the user-friendly story, creating confusion. The sum of weights placed on both stories $[W(S_a, E_a, \{S_a, S_b, S_d\}) + W(S_a, E_a, \{S_a, S_b, S_d\})]$ may be less than the weight originally allocated to the user-friendly story by itself $[W(S_a, E_a, \{S_a, S_d\})]$, so the probability of purchasing an Apple falls.

It would be still more confusing if Dell were to advertise the superiority of its competitor's product, which would cause the probability of buying a Dell to fall.

In short, it may be detrimental for either side to present S_b, and so both sides prefer Situation 4.3 over the alternatives which reveal all the facts.

The key is that the consumer evaluates a story based on who tells it. In a model which does not take the teller of the story into account, it is impossible that the outcome in Situation 4.4 is not equal to the outcome in Situation 4.5, since the same three stories are told in both cases. If both situations result in a probability of buying an Apple of .7 and a probability of buying a Dell of .3, then Dell will tell $\{S_b\}$; if both outcomes give P(A)=.9 and P(D)=.1, then Apple will tell $\{S_b\}$; full information will always be revealed.

The following sections describe the general conditions under which partial information revelation will occur.

4.6 How many stories to present

The intent of this section is to explain why lawyers present only one story, as per Mauet's and Murphy's comments above.

First, note that if the judge's opinion is appropriately restricted, then it will be irrelevant to the advocate's information revelation decision: all available stories will be told in all situations.

Lemma 22 *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 4.3.2: revealing all of $\mathcal{S}$ is a unilateral strategy, as per Definition 10, and the game among the lawyers is strictly competitive, as per Definition 9. Apply Lemma 21.$\diamondsuit$

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 23 *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; there exist conditions and a judge's opinion such that the defendant would not present S_m; and all of these conditions may simultaneously be true.

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.

Risk aversion I have not defined the likelihood that the judge will select an advocate's worst outcome, nor have I defined the advocates' levels of risk aversion. In the case where the conditions alluded to in Theorem 23 and explicitly stated in Equations 7.17 and 7.18 are not satisfied, then questions of risk aversion are obviated: full information will always be revealed. In the case where these conditions are satisfied, then it is possible that we will not have full information revelation. As the worst case receives more weight, or as the advocates become more risk averse, the probability of partial information revelation rises. In the case of absolute risk aversion, the advocates will reveal full information if and only if the conditions are not satisfied.

4.6.1 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

[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".

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. Then the judge 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, 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.

Advocates v inquisitors In some ways, the claim advocates against the advocate model. The fact that the judge or jury may be inclined to use the presenter of the information in reaching a verdict means that the advocates may be reticent to present all of the information they could present. Notably, the more moderate explanations get lost. If the trial were instead conducted by one inquisitor who is advised by the prosecution and defense, but who will not reveal the source of any piece of information, then the only equilibria are those outcomes with verdicts equal to the full information verdict.

4.6.2 Comparative statics

The limits given in Theorem 23 move as we would expect them to.

Corollary 24 *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 $PA(S_m)$.

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.

4.6.3 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 23 reduces to the following:

Corollary 25 *Assume the same situation as Theorem 23, 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*

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

4.7 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 basically 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.

Chapter 5

Conviviality

5.1 Introduction

People find benefit in behaving like others. There are a variety of explanations for this fact, depending on the situation: others' actions can signal which choice is better; the value of joining a network is greater when it has more members; people will pay more for a good if they know they can resell it to others who also want the item; or people may simply enjoy the feeling of commonality produced by imitating other members of a group. In this chapter, I develop a model to describe any situation where the value from acting or consuming is an increasing function of the proportion of other people who are also acting or consuming. People in this model will use their prior information and any available public information to develop expectations about the final number of consumers, and behave accordingly.

Models of 'herding' behavior exist in the literature, but all are sequential models. This is the first model describing people who make a simultaneous decision. Since actors move simultaneously, they can not simply observe others who acted before them, as the extant models describe people as doing. Instead, they must base their actions on their prior beliefs about what people will do.

This chapter then focuses on the question: how do *ex ante* priors on the distribution of tastes predict who will act *ex post*?

5.1.1 Comparison with previous scholarship, with focus on applications

The body of this chapter takes conformity effects as given, and from that starting point, derives implications about possible outcomes. As such, I will not delve into the question of *why* it is that these conformity effects exist in certain situations.

The purpose of this introduction is therefore to discuss some of the situations where one would expect conformity effects to be important. I do not intend to prove that such effects must exist based on other underlying features, since these underlying features would be irrelevant to the reduced-form model which follows. I will explain for each of the situations below why it is plausible that actors would have utility which is increasing in the number of other people who act like them, and then, having discussed why it is plausible that the percentage of actors is included in the utility function, I will simply take this as given when the model itself is described.

5.1.1.1 The restaurant problem

The most common models of the emulation of others are the 'herding' or 'information cascade' models, e.g. Banerjee [4] or Bikhchandani, Hirshleifer, and Welch [7].

The main feature of the model is that there is a sequence of moves, which lead to a path-dependent outcome.[1]

The model here does away with the assumption of sequential information. It also uses information about the distribution of tastes within society, allowing discussion of how prior information predicts who will act. There is no analog to such predictions in the herding model.

The herding model remains untested in the literature. Tests searching for herding effects in various situations abound (e.g., Weinberg on employment [56], Evans et al on racial attitudes [20], Campbell on high school dropout and pregnancy rates [13], Kennedy on prime-time television [32], Anderson and Holt in the lab [2], et cetera),

[1] Orléan [40] almost addresses this, by asking how people will change after everyone has already acted, given only the aggregate number of people acting at the time. In this respect, his model is more akin to that of Brock and Durlauf, discussed below.

but none of these tests are informed by the models from the herding literature which they cite. The problem is not with the empirical tests, it is with the theoretical model, which does not make testable predictions beyond the basic claim that people herd. The model of this chapter makes specific, testable claims about herding based on publicly available information and prior beliefs.

5.1.1.2 Network externalities

This is a property of goods where consumption by others increases the utility of the good, such as computer equipment. The typical analysis (e.g., that of Choi [15]) is similar to that of the restaurant problem, except expected utility is gained directly instead of through information; the discussion above applies.

5.1.1.3 Finance

Many financial instruments fit this model, for a number of reasons:

- Pricing is partly based on the value of the underlying asset and partly on what others are willing to pay for the asset. At the extreme, people will buy a stock which pays zero dividends only if they are confident that there are other people who will also buy the stock; as more people are willing to buy, the value of the stock to any individual rises.

- It has long been a lament of the fund manager that if the herd does badly but he does well, he sees little benefit; but if the herd does well and he does badly, then he gets fired. Therefore, 'behave like others' may very explicitly appear in a fund manager's utility function.

- Since an undercapitalized company is likely to fail, the success of a public offering may depend on how well-subscribed it is, providing another justification for putting the behavior of others in the fund manager's utility function.

- If a large number of banks take simultaneous large losses, then they may be bailed out; since a bail-out is unlikely if only one bank takes a loss, this may also serve as an incentive for financiers to take risks together.

- Simply following the herd: "[...] elements such as fashion and sense of honour affected the banks' decision to take part in a syndicated loan. Banks are certainly not insensitive to prevailing trends, and if it is 'the in thing' to take part in syndicated loans[...], people sometimes consent too readily." ([31], p 337)

As noted above, the model here is a reduced form model which simply assumes that a financier's expected utility from an action is increasing in the percentage of other people acting. I make no effort to explain why the above effects can happen, but posit that given these effects, the model below is applicable.

Empirically, there is evidence that analysts do indeed herd. For example, Graham [25] finds evidence of herding among investment newsletter recommendations, and finds that the *more* reputable ones are more likely to herd. Meanwhile, Hong [29] finds evidence of herding among investment analysts, and finds that inexperienced analysts are "more likely to be terminated for bold forecasts that deviate from consensus", and therefore *less* reputable analysts are more likely to herd. Welch [57] find that an analyst recommendation has a strong impact on the next two recommendations for the same security by other analysts, and that this effect is uncorrelated with whether the recommendations prove to be correct or not. Although these papers disagree in the details, they all find empirical evidence that analysts are inclined to behave like other analysts (and therefore the people who listen to analysts will also tend to behave alike), so the model below is apropos.

Within the theoretical finance literature, papers abound regarding herding behavior (Grossman [27, 28], Radner [46], Choi [15], Minehart and Scotchmer [38]), yet all concern themselves with equilibrium after a sequence of events, when information about who chose what has been disseminated (usually in the form of prices). The model here is more applicable to the case of an initial public offering (IPO) of a security, where there is not yet information about others' choices available for dissemination.

It is commonly known that log equity returns have a kurtosis larger than that of a Normal distribution. Theorem 35 shows that if we assume that private equity values are Normally distributed, but people value equities based on a mix of private value and public value, then equity demand must be leptokurtotic. Under appropriate conditions, this will imply leptokurtotic returns as well.

5.1.1.4 Brock and Durlauf's model

Brock and Durlauf [9] specify a model which is similar to the one presented here. In the first round, a prior percentage of actors is given, and people act iff that percentage would be large enough to give them a positive utility from acting. In subsequent rounds, individuals use the percentage of people who chose to act in the prior round to decide whether to act or not.

Brock and Durlauf make a number of strong assumptions, including quasilinear utility, a logistic distribution of preferences, and a cutoff-type equilibrium, where everyone with a taste parameter over a certain point consumes and everyone below that point does not.[2] Their model is comparable to the setup of examples A and B of Section 5.4.2, and arrives at a comparable equilibrium in those specific situations. But since they assume such a specific form of distribution and utility function, their paper gives a limited answer to the question of how prior information affects the number of people who will choose to act.

Timing and information The timing of Brock and Durlauf's model is also worth considering: they have a *tatônnement* process toward a Nash equilibrium, while I calculate a Nash equilibrium directly. This is basically equivalent, except that implicit to Brock and Durlauf's setup is the assumption that people have full information about the distribution of tastes in the society. This paper also discusses the cases of uninformative or only partially informative prior information.

[2]Theorem 28 shows that by assuming both a logistic distribution and a cutoff equilibrium, they made one assumption more than they needed—a cutoff equilibrium can be derived from their other assumptions.

5.1.1.5 Consumer goods

Fashion good models such as Frijters [21], Pesendorfer [45], or Bernheim [6] are based on goods as a signal of status. They claim that people want to send signals to others indicating their quality—in Pesendorfer's language, they buy a "design commodity" for the sake of attracting a high-quality "date". The strength of that signal declines as more people send it.

An important feature of such items is that they are often Giffen goods, since an expensive item may be in demand simply because it is expensive. As such, this chapter will have something to say about status-oriented models, in that it places restrictions on the sort of assumptions that they may be founded on. Most notably, Theorem 34 shows that such goods can not exist at equilibrium if the distribution of tastes is sufficiently diffuse and there is full information about tastes and prices.

5.1.2 Some conclusions

Following are the main conclusions of the analysis of the model presented in this chapter:

Equilibria A unique equilibrium exists in many specifications. Conditions where there may be multiple equilibria are specified and discussed. An equilibrium where there is a cutoff, such that everybody whose taste parameter is above the cutoff acts and everybody whose taste is below the cutoff does not act, will always exist. In some cases all equilibria will be cutoff equilibria.

That said, we begin with a distribution of tastes, and end with a distribution of cutoffs. Assume for the sake of this summary that tastes are normally distributed, meaning that they are symmetric with a single peak at zero. A single peak at zero means that indifference is the most likely opinion.

No information In the case here, if there are no public signals, then there is only one cutoff, such that everyone acts as if they gained no utility from the behavior of others. So in many situations where there *should* be network externalities, peer pressure, et cetera, these forces have no effect.

Full information Given full information, the distribution of tastes will be more leptokurtotic (spread out toward the extremes) relative to the distribution of tastes. For example, movie quality may be drawn from a bell curve centered around a moderate quality, but movie returns tend to be either in the 'flop' or 'blockbuster' category. There is reason to believe that equity valuations should be normally distributed, but equity returns tend to be leptokurtotic; the claim here is that these effects are due to the conformity effect inherent in movie or equity demand.

The distribution of cutoffs may be so extreme that virtually everyone acts or does not act—herding. Such equilibria are guaranteed only under certain conditions. First, people have to weight the actions of others very heavily, or the society's tastes must be known to be very homogenous. Second, there must be prior information about the mode of the distribution of tastes.

So as the weight of other's actions rise (or the variance of the society's tastes become more concentrated), the distribution of cutoffs become more and more spread out, until moderate outcomes are impossible.

Giffen goods The hallmark of the status-driven fashion good is an upward sloping demand curve. This is shown to be impossible under certain conditions, such as when people have full information about prices and the distribution of tastes is sufficiently diffuse.

Advertising 'Better' advertising will be defined. Not only will better advertising induce more consumption, but better *expectations* about advertising can increase consumption.

An empirical test The model is applied to TV specials, which give evidence of multimodality in demand.

5.1.3 How to use this chapter

This paper is a catalog of results, most of which take some informational structure as given and conclude something about equilibria: that they are cutoff-form, that they are unique, that they are distributed leptokurtotically, that they must involve heavy conviviality effects, et cetera. These descriptions of equilibrium may easily be translated into statements about demand, which will be much more detailed than those possible using the basic herding model discussed above.

But demand is at most half of any microeconomic model, requiring a supply curve or competing goods for context. So this paper is not a complete model of anything by itself, but an input into the production of microeconomic models of situations where people gain utility from emulating others. The assumptions are thus intended to cover enough ground that the model designer may select those assumptions that best describe the situation he is modeling, and may then couch the corresponding results about demand within a complete microeconomic model.

5.1.4 Outline

Section 5.2 defines the model. Section 5.3 defines a Bayesian Nash equilibrium in this context. Section 5.4.2 discusses equilibria given complete information, Section 5.4.3 discusses what the equilibria will look like in the case of uninformative prior information, and Section 5.4.4 gives a comparative statics result for the case of informative but incomplete information. Section 5.5 shows two ways in which the model could be extended to include advertising, Section 5.6 offers some extensions, and Section 5.7 shows some empirical evidence that the viewership of TV specials shows simultaneous conviviality effects of the type discussed in this chapter .

5.2 The model

Actors are faced with a binary action. For example: stand in line at the popular restaurant or go straight in to the unpopular one, install either a DOS-based operating system or a UNIX-based system, revolt or don't revolt, buy a good or don't buy it. For the sake of consistent terminology, I will describe the choice as being between 'acting' and 'not acting'.

A countably infinite number of individuals will simultaneously choose to act or not act. The decision is based on two factors. The first factor is the percentage[3] of others acting, $k \in [0, 1]$, which will be endogenously determined. The second factor is the net individual utility to consumer i, denoted by a real number t_i. The distribution of t within the society depends on a parameter m, which is discussed below. Given m, the distribution of tastes is described by a PDF $f(t|m)$ which has support $(-\infty, \infty)$ for any given m. The family of distributions $f(t|m)$ is common knowledge, even though m may not be.

5.2.1 Time line

The sequence of events is as follows:

1. Everyone shares a common prior belief about the distribution of the parameter m, in the form of a probability density function $a(m)$, with full support over all $m \in (-\infty, \infty)$, derived from either previous experience or (credible) advertising. This prior may range anywhere from fully informative to entirely uninformative. As a common special case, the parameter m will often be taken to be the mode.

2. Nature draws a fixed value μ for m, but may or may not reveal μ to actors.

3. Individuals draw a private utility t_i.

4. Individuals use t_i to update their prior beliefs $a(m)$ to a new private likelihood distribution over the distribution of tastes.

[3] This is an approximation, since it is impossible to define a percentage of a countably infinite number of people. More precisely, we will find that there exists a consistent estimate of the likelihood that a randomly drawn individual will act, $k \in [0, 1]$. I will stick to the description 'percentage of actors', however, because it makes more intuitive sense, and is formally correct in the real world, where there are a large but finite number of actors.

5. Person i now has enough information to calculate the expected utility from acting and from not acting, and makes a decision accordingly.

6. All individuals act simultaneously.

7. Person i receives a payoff based on t_i and the actual proportion of people who acted.

8. Section 5.4.3.4 considers the case where individuals have a second chance to act and receive a second round of payoffs.

5.2.2 Assumptions about the utility function

The total value from acting is a function $V(t_i, k) : \mathbb{R} \times [0,1] \rightarrow \mathbb{R}$, where $t_i \in \mathbb{R}$ is a level of tastes drawn from the distribution discussed above, and $k \in [0,1]$ is the percentage of other actors, *ex post*, after everyone has decided whether to act or not.

Assumption 18 *The function $V(t,k)$ is the same for all actors, is continuous at all but a finite number of discontinuities, and is strictly monotonically increasing in both t and k.*

The utility from not acting is defined to be $V(0, 1-k)$.

Notice that when not acting, people care about the percentage of others who are not acting, just as they care about the percentage of others who act when taking the action themselves.

We could have described the situation with three variables: t being the private utility from acting, k being the percent of actors, and s the private utility from not acting. But s would be redundant:

Lemma 26 *For any function $U(\cdot,\cdot)$ which is monotonically increasing in both arguments, there exists a function $V(\cdot,\cdot)$ which is also monotonically increasing in both arguments, such that if $U(t,k) > U(s, 1-k)$, then $V(t-s, k) > V(0, 1-k)$.*

Normalizing the private utility from not acting to zero is therefore a convenient and unrestrictive assumption.

Finally:

Assumption 19 *There is some sufficiently low value $t_{\min}$ such that $V(t_{\min}, 1) < V(0,0)$.*

This means that if if i draws $t_i < t_{\min}$ then t_i will not act, regardless of expectations about k. In other words, tastes do matter somewhat, so that even if i knows that everyone will act with certainty ($k = 1$), he still won't act if he finds doing so sufficiently distasteful. This assumes away some perverse equilibria.

Having described the value from acting and not acting, the game is simply a question of comparing the expected values of V: all consumers observe their private utility t_i, and simultaneously decide whether to consume or not, based on whether $EV(t_i, k) > EV(0, 1-k)$ or vice versa.

5.3 Equilibria defined

This section defines pure strategy symmetric Bayesian Nash equilibria in the context of the model, along with the closely related expected demand function.

The manner in which Bayesian Nash equilibria will be defined in this chapter requires two new sets of notation: one to describe sets of actors and non-actors, and one to describe the subjective distribution of k given t.

In the equilibria below, a person will be able to decide whether to act or not based entirely upon his or her draw of t_i. Let those t which prompt the consumer to act be T^A, and those that cause the consumer to not act as T^N. I will assume that consumers who are indifferent will not act; let the ts which induce indifference be T^*, meaning that $T^* \subseteq T^N$.

Equilibria defined The actors in the model are completely described by the parameter t, and the set of pure-strategy actions available to them are to act or not act.

Therefore, a pure strategy symmetric Bayesian Nash equilibrium (herein referred to as 'an equilibrium') is completely described by a set T^A. For linguistic simplicity, I will often say 'T^A is an equilibrium' to mean 'in equilibrium, only members of the set T^A act'.

A set T^A is an equilibrium iff any individual who draws $t_i \in T^A$ chooses to act, given the prior information held by the individual and given that all other players are playing the strategy 'act iff I draw tastes $t_i \in T^A$'; and no individual who draws $t_i \notin T^A$ acts, given the same information.

Conviviality This notation allows the following definitions: active conviviality is when someone whose draw of t_i is less than zero chooses to act. That is, $T^A \cap (-\infty, 0) \neq \{\emptyset\}$. Inactive conviviality is when someone whose draw of t_i is greater than zero chooses not to act. That is, $T^N \cap (0, \infty) \neq \{\emptyset\}$.

The gist of these definitions is that if utility were purely a function of tastes, then everyone with $t_i < 0$ wouldn't act while everyone who draws $t_i > 0$ would. Situations of active or inactive conviviality are those where someone has changed their behavior based on the behavior of others.

5.3.1 Cutoff equilibria and expected demand curves

Without saying anything about the priors, the following is true for any Bayesian Nash equilibrium:

Lemma 27 *Define $f(t|m)$ to be a translation family*[4] *based on a PDF with a finite number of discontinuities. Then in equilibrium, T^A is a (possibly empty) set of open intervals.*

A desirable refinement would be to have a cutoff equilibrium:

Definition 20 *A cutoff equilibrium is where T^A is of the form (T^*, ∞) for some point T^*.*

[4] Translation families were defined in Section 2.2.1.

In the analysis of the three cases which follow, all equilibria in the full information case will be cutoffs, reasonable assumptions will induce cutoff equilibria in the case of an uninformative prior, and a cutoff will simply be assumed for the case of an informative public prior.

Bayesian updating The decisions made by the actors are made using posterior probability distributions, arrived at after absorbing all available information. Within the context of this chapter, Bayesian updating (Equation 2.1) takes the likelihood function $f(t|m)$ and the prior density $a(m)$ and produces a posterior density of m given t defined to be

$$post(m|t) \equiv \frac{f(t|m)a(m)}{\int_{-\infty}^{\infty} f(t|\nu)a(\nu)d\nu}.$$

Demand as a function of tastes This section defines the expected demand curve associated with each equilibrium T^A. Discussion of the implications of cutoff equilibria is best done in the context of the expected demand functions they imply: Section 5.4.2.4 will show that cutoff equilibria are related to monotonically upward-sloping expected demand, while non-cutoff equilibria will have nonmonotonic expected demand.

Taking as given the set T^A and the value of m, the total number of people acting is

$$k(T^A, m) \equiv \int_{T^A} f(x|m)dx.$$

A belief about $k(T^A, \cdot)$ may be thought of as the expected demand given equilibrium T^A, mapping m to the percentage of actors.

Actor i's value from acting given the above information would be $V\left(t_i, k(T^A, m)\right)$. However, if m is not known, i must rely on the posterior likelihood of a given value of m: $post(m|t_i)$. Taking this into account, the expected utility from acting is the sum of [(the payoff given m) $\times$ (the likelihood of m)] over all possible values of m, that is:

$$\int_{-\infty}^{\infty} V\left(t_i, k(T^A, m)\right) post(m|t_i)dm. \tag{5.1}$$

If $k(T^A, m)$ were an invertible function of m (see below), so that there exists a function $m(T^A, k)$, then we could write i's perceived probability that k percent of the people will consume, given T^A, as:

$$p_i(k) = post(m(T^A, k)|t_i),$$

and Equation 5.1, the expected utility of acting, could be written as

$$\int_0^1 V(t_i, \kappa)\, p_i(\kappa) d\kappa.$$

5.3.2 Cutoff equilibria and monotonic demand

We don't yet know what T^A or T^N look like in equilibrium; we don't yet know how the expected value of k is derived from t_i; but we can still state the following relation:

Theorem 28 *Assume that $f(\cdot|m)$ and $post(\cdot|m)$ satisfy FOSD.*[5] *Then: $k(T^A, m)$ is monotonically increasing in $m \Leftrightarrow T^A$ is a cutoff equilibrium.*

One would expect that under normal conditions, as the median of consumer tastes increases, the percentage of people acting would also increase. This theorem says that such 'normal conditions' imply a cutoff equilibrium. Or, inherent to a non-cutoff equilibrium is the claim that there is some region in which consumers liking the good more would imply fewer consuming.

For the remainder of the chapter, I will consider only cutoff equilibria. In some of the cases below, this will be proven to be the only possibility, meaning that goods with upward-sloping expected demand curves can not exist. In other cases, it will be assumed; the commentary above shows that this is a plausible but not trivial assumption.

[5]FOSD was defined in Section 2.2.1.

5.4 Equilibria given different types of public information

I will describe equilibria in three situations: the first is the case where the true parameter m is known with certainty. The second is the case where there is no prior information about m. The third is the case where there is commonly known prior information (but still some uncertainty) about m.

The main conclusions of this section are as follows:

- Given full information, all equilibria are cutoff equilibria. There can be multiple equilibria, but only if the distribution of tastes is sufficiently concentrated at certain points or if the utility from conformity is high enough. [Theorem 34]

- A prior distribution of medians, $a(m)$, induces a distribution of cutoffs; under appropriate assumptions (notably full information and that $f(\cdot|m)$ is single-peaked), the distribution of cutoffs is less concentrated around the center than the distribution of medians. Given some parameters, all equilibria are extreme. [Theorem 35]

- If m is taken to represent prices (see Section 5.4.2.4 for details), then demand curves can slope upward only if people are sufficiently concerned with how others behave, or if tastes are sufficiently homogenous, as defined by Theorem 35. Since fashion goods are often used as examples of Giffen goods, this places restrictions on how one could model them. [Proposition 36]

- Given full information about m, a symmetric distribution assures us that there is an equilibrium with no conviviality effects; an upward-leaning distribution assures us that there exists an equilibrium with active conviviality; a downward-leaning distribution assures us that there is an equilibrium with inactive conviviality. [Theorem 41]

- Given an uninformative prior, there can be only one cutoff equilibrium. [Proposition 39]

- Given an uninformative prior and a symmetric distribution, there are no conviviality effects; given a distribution leaning toward the positive, there are active conviviality effects; and given a distribution leaning toward the negative, there are inactive conviviality effects. [Theorem 40]

- Given one distribution of prior information $a_1(\cdot)$ which satisfies the monotone likelihood ratio property (MLRP) with respect to another, $a_2(\cdot)$. Then any equilibria induced by $a_1(\cdot)$ must involve strictly fewer actors than any equilibria induced by $a_2(\cdot)$, meaning that more people will act. In other words, better advertising makes more people want to buy, where 'better' is defined using the MLRP. [Theorem 42]

5.4.1 Some preliminary results

The following lemmas will used in all three informational cases below. A cutoff equilibrium will be either proven or assumed for each case. As such, a single point T^* defines an area $T^A \equiv (T^*, \infty)$, so all we need to find a pure strategy Bayesian Nash equilibrium of a cutoff form is the single point T^*.

The consumer is indifferent if the expected utility from acting is the same as the expected utility from not acting. For each t_i, define $t^*(t_i) : \mathbb{R} \to \mathbb{R}$ to be the point such that if the cutoff T^* were to equal $t^*(t_i)$, then one who drew t_i would be indifferent. It solves:

$$\int_{-\infty}^{\infty} V\left(t_i, k([t^*(t_i), \infty), \tau)\right) post(\tau|t_i) d\tau = \int_{-\infty}^{\infty} V\left(0, 1 - k([t^*(t_i), \infty), \tau)\right) post(\tau|t_i) d\tau. \tag{5.2}$$

Proposition 29 *If the posteriors $post(\cdot|m)$ satisfy FOSD, and the likelihood function $f(\cdot)$ is continuous, then $t^*(t_i)$ is a continuous, increasing function of t_i.*

This is the tool we need to find the equilibria:

Corollary 30 *Given that the posteriors* $post(\cdot|m)$ *satisfy FOSD and the likelihood function* $f(\cdot)$ *is continuous; then the cutoff-type pure strategy Bayesian Nash equilibrium* T^* *is at the value(s) of* t_i *for which* $t^*(t_i) = t_i$*. There is always at least one cutoff equilibrium, either at some finite value or at* $T^* = \infty$*.*

We can also transform this result to the space of k. Theorem 28 stated that if there is a cutoff equilibrium, then $k(T^A, t)$ is a monotonically decreasing function of t, so this is a trivial transformation. Define $k^*(t_i) : \mathbb{R} \to [0, 1]$ as the percentage of people who would need to act to make one who drew t_i indifferent between action and inaction. If t_i has density $f(t)$, then let $F(t)$ be the cumulative distribution function (CDF) of t. Also, note that if t_i were the cutoff in a cutoff equilibrium, and if i believes that the t_is are distributed $\sim F_i(\cdot|m)$, then $1 - F_i(t_i|m)$ percent will act. Restating Corollary 30 in the space of k:

Corollary 31 *The cutoffs for cutoff-type pure strategy Bayesian Nash equilibria are at the value(s) of* t_i *for which* $k^*(t_i) = 1 - F_i(t_i)$*. If* $k^*(t_i) > 1 - F_i(t_i)$ *for all* t_i*, then* $K^* = 1$ *is an equilibrium (where no one acts).*

In other words, we need only find a person who would be indifferent if they happened to be the cutoff. Such a person may consistently be a cutoff in equilibrium, and a person may only be a cutoff in equilibrium if he is indifferent given that he is the cutoff.

5.4.2 Full information

This section discusses the full information case, defined to mean that it is common knowledge that the parameter m equals some fixed value μ with certainty.

The caveats about when we are assured a cutoff equilibrium (and therefore monotonic expected demand) become moot in the full information case:

Lemma 32 *Given full information, all perfect strategy Bayesian Nash equilibria are cutoff equilibria.*

This result did not make any assumptions about the form of the distribution of tastes $f(\cdot)$, and requires only the monotonicity from Assumption 18 with regards to the function $V(\cdot,\cdot)$.

From here, complications ensue; some examples may be helpful.

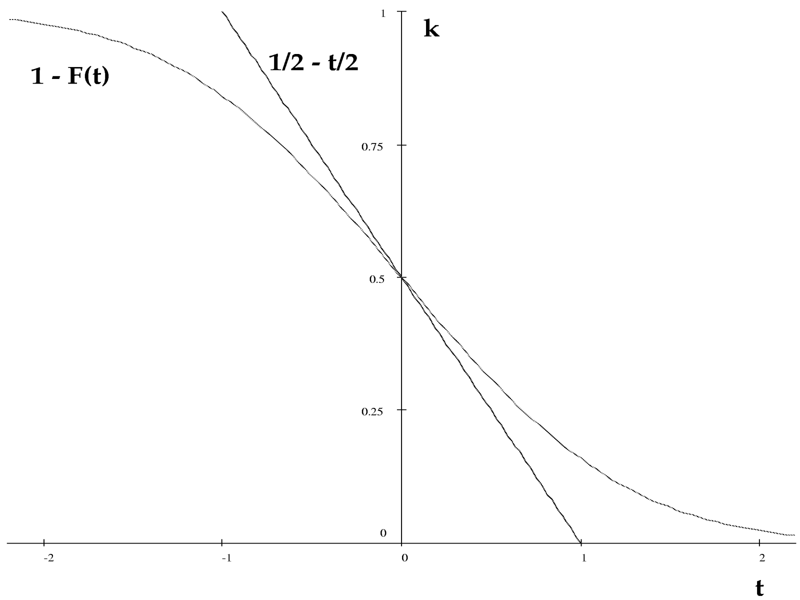

Figure 5.1: the 'normal' case

5.4.2.1 Example A: the normal case

Take $f(\cdot|m) = \mathcal{N}(m,1)$, and take the value function to be linear and separable: $V(t,k) = t + nk$.

If $n = 1$, then we have the case in Figure 5.1: $k^*(t_i)$ is the straight line; $1 - F(t_i|0)$ is the curve. The equation $k^*(t_i) = 1 - F(t_i|m)$ is satisfied at the crossing of these two curves, meaning that this is the location of any equilibria. Clearly, there is only one crossing point, and therefore a unique equilibrium.

Comparative statics are simple in this case: if m rises from zero to a positive value, the curve translates to the right, and therefore the cutoff falls. Since the cutoff falls, more people act. In other words, if it is revealed that more people like a good, then more people will consume.

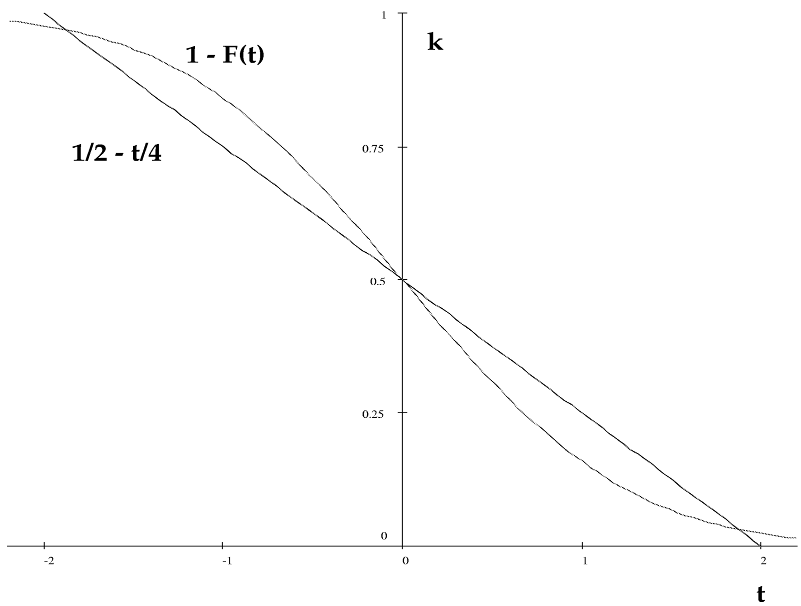

Figure 5.2: If people care enough about k, perverse equilibria appear.

5.4.2.2 Example B: the abnormal case

The case $n = 2$, so $V(t, k) = t + 2k$, is much more interesting. Now consumers are twice as interested in the percentage of actors than in the good itself, and the result is as in Figure 5.2. There are three possible cutoffs, and the society agreeing on any one of them is a Nash equilibrium. One way to resolve this is to make a pessimism assumption, that given a few equilibria, everyone agrees on the one with the lowest number of people acting (and therefore the highest value of t).

The comparative statics for the pessimist (or the optimist) case is very simple: as the mean of the distribution rises, T^* falls and more people act. There is a jump in the value of the cutoff at a certain value (as the 'S' crosses the line).

Much more interesting is the 'moderation' case: if consumers agree upon the middle of the three equilibria, then as μ_A increases, T^* also increases, meaning that *fewer* people consume. People may coordinate on this middle equilibrium, but the equilibrium is unstable: if people think the cutoff is perturbed to the left of T^*, this would only lead to a greater shift to the left, eventually reaching the leftmost crossing point, and similarly with a shift to the right. Therefore the equilibrium point can be thought of as the dividing line between points which lead by *tatônnement* to a low T^* and points which lead to a high T^*; since this is increasing, the number of points which lead to high consumption rates increases as the mean taste for the action or good increases, consistent with intuition.

General results These examples show that if the distribution of tastes is diffuse enough or (equivalently) the desire for conformity is low enough, then there will be only one equilibrium. The following results explain sufficient conditions for one equilibrium.

Proposition 33 *The PDF $f(\cdot|m)$ may take any form, and the parameter m will be known to have the value μ with certainty. If for any equilibrium point $t = \tau$, $f(t|\mu) < -\frac{\partial k^*}{\partial t}(t)$ for all $t > \tau$, then there will be only one equilibrium for the given μ.*

Theorem 34 *The PDF $f(\cdot|m)$ may take any form, and the parameter m will be known to have the value μ with certainty. If $f(t) < -\frac{\partial k^*}{\partial t}(t')$ for all t, t', then there will be only one equilibrium for any μ.*

The left-hand side of the inequality in this statement is the derivative of the CDF, while the right-hand side is the derivative of $k^*(t)$, so this statement holds when the CDF is expanding more slowly than $k^*(t)$ is contracting [Notice that $\partial k^*/\partial t$ is always negative]. In words, this means that the value from conforming expands quickly relative to the most concentrated parts of the distribution. If the PDF has low peaks, or if $k^*(t)$ falls very quickly, then there will be only one equilibrium. This condition is sufficient for a unique equilibrium, but not necessary.

5.4.2.3 The distribution of cutoffs

Even though m was taken as fixed above, it may have been drawn from a distribution $a(m)$ before being fixed and made common knowledge. For example, the median of movie enjoyability may be observed to have distribution $a(m)$. When a new movie comes out, the median of its quality is drawn from $a(m)$, and reviewers reveal the value of m to the viewing public. The model is then as above: m is a fixed, known scalar, and Theorem 34 showed that if certain conditions hold, then $T^*(m)$ is a function producing only one equilibrium for each fixed value of m.

Let $T^*(\mu)$ be the function mapping values of μ to the equilibrium T^* that they induce.

Given the original *ex ante* distribution of $a(m)$, then, what is the *ex ante* distribution of T^*? The following theorem then says that under appropriate conditions, the distribution of T^* is a fat-tailed one relative to $a(m)$.

Theorem 35 *The ex ante distribution of the median m is $a(m)$, which can take any form. After drawing μ from $a(m)$, μ is known with certainty. Assume the PDF $f(x)$ is single-peaked, with a maximum value of $f(x) = \frac{1}{n}$, and the value function is linear: $V(t,k) = nk + t$. Let $d(T^*)$ be the ex ante distribution of T^* (before μ is known).*

Given these assumptions, the ratio $a(m)/d(T^*(m))$ is single-peaked with a peak at the point where $\mu = T^*(\mu)$. That is, $d(T^*)$ is less concentrated toward the center than $a(m)$.

Given the assumptions about $f(\cdot)$ and $V(t,k)$ the conditions of Theorem 34 now apply, so there will be only one cutoff for any given value of m, and $T^*(m)$ is truly a function.

Returning to the example, if movie quality is normally distributed, then movie returns will be fat-tailed. De Vaney and Walls ([55], p 1512) studied movie returns and found exactly this: "The end-of-run or total revenue distribution for motion pictures ... never quite reaches log normality; it has fatter tails than the log normal and mass points at the far right, where the superstars are located."

Continuing the example in the context of the theorem, $a(m)/d(T^*(m))$ is single-peaked, so its inverse $[d(T^*(m))/a(m)]$ can be described as 'single-troughed'. Also, Theorem 41 below will show that $T^*(0) = 0$, so this is where the nadir of the trough will be located. These facts imply that the distribution $d(T^*)$ must have a lower center and fatter tails than the Normal distribution. That is, actual movie quality which is distributed log-normally leads to log movie returns which have a leptokurtotic distribution, as found by De Vaney and Walls.

An example For further intuition on the meaning of this theorem, consider the case where the distribution $a(m)$ is the Uniform$[-1, 1]$ distribution, and where $f(\cdot|m)$ is a Normal$(m, 1)$ distribution. Continue to assume that the value function is linear: $V(t, k) = nk + t$. Then the proof shows that the distribution of $d(T^*)$ is proportional to the ratio

$$\frac{\frac{1}{2n}}{\frac{1}{2n} - f(T^*(m)|m)}.$$

The premises of Theorem 34 assure us that the denominator is always positive, which would make this a continuous function of n for the range where these premises hold. When the distribution $f(t|m)$ is $\mathcal{N}(0, 1)$, the peak of the distribution is at zero, where $f(0|0) \approx .4$. If $\frac{1}{2n} >> .4$ [i.e., $n << 1.25$], this ratio as a function of m is approximately constant. If $\frac{1}{2n}$ is only a little larger than .4, the ratio will be larger at $m = \pm 1$ and dip to a minimum at $m = 0$. As n gets larger, the trough gets deeper, until $\frac{1}{2n} = .4$, at which point the function becomes discontinuous, multiple equilibria are possible, and only extreme values of T^* will be stable equilibria.

5.4.2.4 Giffen goods

Fashion goods are often used as an example of Giffen goods, so it is worth considering when Giffen goods would arise in the model here. It is easy to modify the model to include prices: take the parameter m to be the negation of the market price, so as the price rises, then the median of the consumers' value for purchasing the good falls. For example, if the good were free, then the taste for the action of purchasing the good may be distributed $\sim \mathcal{N}(0, 1)$, but if the good cost two dollars, then tastes for the action of purchasing the good would be distributed $\sim \mathcal{N}(-2, 1)$, with most people not wanting to consume the good, while those who strongly preferred the good when it was free still gaining positive utility from buying the good for \$2.

The question of drawing the demand curve them becomes: as the final price μ

increases, does the cutoff T^* rise or fall?

The setup above gave conditions where $T^*(m)$ is a one-to-one function. Returning to the original interpretation where $f(t|m)$ is increasing in m,

Proposition 36 *Given the premises of Theorem 34, $T^*(\mu)$ is a strictly decreasing function.*

This reverses if m represents prices, so $f(t|m)$ is decreasing in m: as the fixed price μ rises, the cutoff T^* rises, meaning that the percentage of consumers falls. In other words, given the premises of Theorem 34, the demand curve always slopes down.

The premises of Theorem 34 are not trivial, and all of them may be plausibly dropped to describe a society where Giffen goods could exist. First, it is assumed that either consumers are not too concerned with the behavior of others, or that tastes are not too concentrated. Fashion goods are universally those where people are concerned with emulating the behavior of others. Also, if there are multiple equilibria, then a shift in prices can both cause a change in the location of the equilibria and a change in which equilibrium the consumers coordinate upon.

Second, it is assumed that the value of acting is monotonically increasing in the percentage of other actors. But consider fashion good models such as Frijters [21], Pesendorfer [45], or Bernheim [6], based on goods as a signal of status. Each of these models feature a utility function which is increasing in the percentage of high-type consumers of a good, and is decreasing in the percentage of low-type consumers. Therefore the value of consuming may increase or decrease as a function of the total number of consumers. The moral is not that fashion goods can't be Giffen goods—there is evidence that such goods do exist—but that any model of fashion goods with Giffen demand curves can not be based on a homogenous society with full information and purely self-interested actors.

5.4.3 Uninformative prior information

The uninformative prior I will use in this chapter is constructed as follows: Define the function

$$U_d(m) \equiv \begin{cases} 2/d & m \in [-d, d] \\ 0 & m \notin [-d, d] \end{cases}.$$

This is a Uniform distribution over the range $[-d, d]$. The uninformative prior is then defined as the function $\lim_{d \to \infty} U_d(m)$. This type of distribution is discussed extensively in texts on Bayesian inference, e.g. Zellner [59], pp 41*ff*.

Defining $a(m)$ to be an uninformative prior as defined here greatly simplifies the equation for Bayesian updating:

$$\begin{aligned} post(m|t) &= \frac{f(t|m)a(m)}{\int_{-\infty}^{\infty} f(t|\mu)a(\mu)d\mu} \\ &= \lim_{d \to \infty} \begin{cases} \frac{f(t|m)[\frac{2}{d}]}{\int_{-d}^{d} f(t|\mu)[\frac{2}{d}]d\mu} & m \in [-d, d] \\ 0 & m \notin [-d, d] \end{cases} && (5.3) \\ &= \lim_{d \to \infty} \begin{cases} \frac{f(t|m)}{\int_{-d}^{d} f(t|\mu)d\mu} & m \in [-d, d] \\ 0 & m \notin [-d, d] \end{cases} && (5.4) \\ &= f(t|m) && (5.5) \\ &= f(t - m). && (5.6) \end{aligned}$$

Equation 5.3 breaks the statement down into the case where $U_d(m) = [\frac{2}{d}]$, and where $U_d(m)$—and thus the whole expression—is zero. Equation 5.4 follows from the fact that $[\frac{2}{d}]$ is constant with respect to μ, so it comes out of the integral in the denominator and cancels out. Equation 5.5 uses the fact that $\int_{-\infty}^{\infty} f(t - \mu)d\mu = 1$, and that the case where $m \notin [-d, d]$ is obviated when $d \to \infty$. Equation 5.6 uses the assumption that the family $f(t|m)$ is a translation family. This assumption provides the structure needed to arrive at the results below.

Notice that as m rises, the point at which $f(\cdot)$ is evaluated declines, so that $post(\cdot|t)$ is the mirror image of $f(\cdot|t)$ around t.

Here are a few useful definitions for any generic PDF $p(\cdot)$:

Definition 21 *A distribution is* symmetric *iff* $p(d) = p(-d)\ \forall d$.
A distribution is upward-leaning *iff* $p(d) > p(-d)\ \forall d > 0$.
A distribution is downward-leaning *iff* $p(d) < p(-d)\ \forall d > 0$.

So if a symmetric prior $f(\cdot|t)$ is updated using an uninformative prior, the posterior will be symmetric; if an upward-leaning prior is updated using an uninformative prior, then the posterior will be downward-leaning; and vice versa for a downward-leaning prior.

5.4.3.1 Uninformative priors with symmetric, single-peaked priors imply cutoff equilibria

A distribution $f(x|0)$ is single-peaked iff: for all $x < 0$, $f(x)$ is an increasing function; for all $x > 0$, $f(x)$ is a decreasing function; and $f(x)$ attains its maximum at $x = 0$.

This conditions and a few more are sufficient to guarantee that equilibria will be of a cutoff form:

Proposition 37 *Assume uninformative prior information; that $f(t|0)$ is single-peaked and either symmetric or upward-leaning; and that $f(t|m)$ is a translation family. Then the only pure strategy Bayesian Nash equilibria are of a cutoff form.*

Some examples Here are a few examples to give some intuition as to why upward-leaning or symmetric priors imply a cutoff equilibrium, while downward-leaning priors allow relatively perverse equilibria.

Consider the following case:

$$V(t,k) = \frac{\arctan(t)}{\pi/2} + k$$

$$post(x|t) = \begin{cases} 3/2 & x \in [t - \frac{1}{6}, t + \frac{1}{2}] \\ 0 & x \notin [t - \frac{1}{6}, t + \frac{1}{2}] \end{cases}$$

$$T^A = (-.295, .295).$$

The function $\arctan(t)/(\pi/2)$ is monotonically increasing in t, but asymptotically approaches one; this means that no one would be willing to act if no one else does. The posterior is upward-leaning, so the priors must have been downward-leaning.

The person who drew $t = -.295$ believes that a total density of $k = \frac{3}{4}$ is within T^A and will act, and $\arctan(-.295)/(\pi/2) = -\frac{1}{2}$, so $V(t,k) = V(0, 1-k) = \frac{1}{4}$. The person who drew $t = \frac{1}{2}$ believes $k = \frac{1}{4}$, and $\arctan(.295)/(\pi/2) = \frac{1}{2}$, so $V(t,k) = V(0, 1-k) = \frac{3}{4}$. Thus, both ends are valid cutoffs, and the reader may verify that all individuals with draws in T^A will prefer to act and all those with draws outside of T^A will not act.

At the same time, $T^A = (-.295, \infty)$ is also an equilibrium.

Now, consider the same setup as above but a downward-leaning posterior:

$$post(x|t) = \begin{cases} 3/2 & x \in [t - \frac{1}{2}, t + \frac{1}{6}] \\ 0 & x \notin [t - \frac{1}{2}, t + \frac{1}{6}] \end{cases}.$$

Consider any two points t_i and t_2. Given the posterior here, $k(T^A, t_1) \leq k(T^A, t_2)$. This means that $V(t_1, k(T^A, t_1)) < V(t_2, k(T^A, t_2))$, while $V(0, 1 - k(T^A, t_1)) > V(0, 1 - k(T^A, t_2))$, so it could never be the case that both t_1 and t_2 are cutoffs, so we could not see an equilibrium of the form $T^A = (t_1, t_2)$.

An immediate implication of Proposition 37:

Corollary 38 *For a fixed set T^A; if $f(t|m)$ is a single-peaked and either symmetric or upward-leaning function and people have an uninformative prior; then $k(T^A, t)$ is monotonically increasing in t.*

Proof: This is simply extending the result of Proposition 37 via Theorem 28.

$\diamondsuit$

The question of whether it is plausible to assume that the expected demand is an increasing function of t is therefore obviated in the case of an uninformative prior: it is a direct result of the model's setup and assumptions.

5.4.3.2 Uniqueness of a cutoff equilibrium

Regardless of whether Proposition 37 applies, the following can be said about the class of cutoff equilibria:

Proposition 39 *Define $f(t|m)$ to be a translation family. If individuals have uninformative prior information about the true center of the distribution, then there is a unique cutoff equilibrium.*

If Proposition 37 holds, then this implies a unique equilibrium, but if it does not hold, there may be non-cutoff equilibria in addition to the unique equilibrium of a cutoff form.

Further, the shape of the distribution says something about where the unique cutoff equilibrium is located:

Theorem 40 *Assume $f(t|m)$ is a translation family of a PDF that is continuous at all but a finite number of points, and that the prior is uninformative.*

(i: no conviviality) In the symmetric case $[f(d) = f(-d)]$, the unique equilibrium satisfies $T^* = 0$, meaning that there is neither active nor inactive conviviality. That is, if the distribution is symmetric, people behave as if they only cared about their private signals.

(ii: active conviviality) If the distribution is upward-leaning, meaning that $f(d) = [f(-d)/A]$ for all $d > 0$ and some $A < 1$, then the unique cutoff equilibrium satisfies $T^* < 0$, meaning that everyone with a private signal to act will do so, but some people with negative private signals will ignore them and act anyway.

(iii: inactive conviviality) If the distribution is downward-leaning, so $f(d) = [f(-d)/A]$ for all $d > 0$ and some $A > 1$, then the unique cutoff equilibrium satisfies $T^* > 0$, meaning that everyone with a private signal not to act will not act, but some people with positive private signals will ignore them and not act as well.

This theorem underscores the importance of the distribution of tastes—ignored in herding models—in determining the outcome of the game. If I think that the distribution of outcomes is skewed upward, then I am more likely to ignore a small negative signal and follow the crowd. If I think the distribution is symmetric, then I have no information to go on either way, and so I can do no better than just follow my own signal.

5.4.3.3 Comparing posteriors

A theorem similar to Theorem 40 would be useful, comparing outcomes given different shapes of distributions. In this case there is the additional information $m = \mu$. Thus the definition of 'leaning' used below is less stringent, but depends upon μ. A $\mathcal{N}(0,1)$ is symmetric by this definition, for example, but the $\mathcal{N}(1,1)$ distribution is upward-leaning, since $p(i) > p(-i)$ for all $i > 0$.

Definition 22 *A distribution $f(\cdot|\mu)$ is* symmetric *iff $f(d|\mu) = f(-d|\mu)$ $\forall d$.*

A distribution $f(\cdot|\mu)$ is upward-leaning *iff $f(d|\mu) > f(-d|\mu)$ for all $d > 0$.*

A distribution $f(\cdot|\mu)$ is downward-leaning *iff $f(d|\mu) < f(-d|\mu)$ for all $d > 0$.*

Theorem 41 *Given that $m = \mu$ is known.*

(i: no conviviality) In the symmetric case, there is an equilibrium at $T^* = 0$, meaning that there is neither active nor inactive conviviality.

(ii: active conviviality) If the distribution is upward-leaning, then there is an equilibrium at some point $T^* < 0$.

(iii: inactive conviviality) If the distribution is downward-leaning, then there is an equilibrium at some point $T^* > 0$.

None of these cases preclude other equilibria existing elsewhere, as shown by example B above, which was symmetric but had equilibria greater than, less than, and equal to zero. However, in the case where there is only one equilibrium, this demonstrates the points also be made by Theorems 40 (above) and 42 (below): the shape of the distribution matters in determining the final level of consumption, where upward-leaning distributions lead to more consumption; and the news of some ms are 'better' in the sense of causing more people to act.

5.4.3.4 A two-period model

Consider the case where the model has been executed (with an informative or uninformative prior), and actors are given a second opportunity to decide. Total utility will now be a monotonically increasing function of the payoffs in both periods. With a countably infinite number of actors, each actor has measure zero, and therefore can not behave strategically in the first period to change the information available in the second period. Therefore all actors will try to maximize their utility in the first round, and the model above applies in that round. If it is common knowledge that the equilibrium T^A was chosen, then both inputs into the function $m(T^A, k)$ as defined in Section 5.3.1 may be observed, and therefore m is known with certainty. For example, if it is common knowledge that a cutoff strategy was chosen; that the cutoff was at zero, so $T^A = (0, \infty)$; that tastes are distributed $\mathcal{N}(m, 1)$; and that $k = 90\%$ of the actors acted; then anyone can check the back of any statistics textbook to verify that the center of the distribution must be at $m = 1.28$.

Therefore, in the second round, given all of the information assumed to be available, actors will have full information, and the results of Section 5.4.2 apply in the second round.

5.4.4 The case of prior information

Changes in prior information shift the equilibria in the same manner as a similar change in m in the full information case. Formally,

Theorem 42 *Assume prior distribution $a_1(m)$ satisfies the monotone likelihood ratio property with respect to $a_2(m)$, and assume cutoff equilibria. Assume nothing about the distribution $f(t)$. Then someone who is indifferent between action and inaction given prior $a_2(m)$ will strictly prefer acting given $a_1(m)$.*

This means that if public information $a_1(m)$ MLRPs $a_2(m)$, then a_1 is 'better': it induces a cutoff further to the left, and thus causes more people to act. This will have an immediate application in the model of Section 5.5.2.

The reader may wonder why the MLRP was needed for this theorem. The short answer is that this is what is necessary to ensure posteriors which satisfy FOSD using Theorem 5, but Chapter 2 gives a complete discussion of this question.

5.5 Application: advertising

5.5.1 Advertising model one

One interpretation of advertising is that it directly raises the expected utility from consuming a product, by revealing facts about the product or its quality.

Let $A \in [0,1]$ be the percentage of people who see an ad and let $\nu \in \mathbb{R}^+$ be the added value from seeing the ad. For the sake of simplicity, assume that ν's effect is additive and the same for all, so after viewing an ad the taste for the product or action rises from t to $t+\nu$.

Seeing and ad imparts both A and ν, since ν is directly observable, and A is observable in the medium chosen—commercials on prime time TV imply a wide viewership, while a mass mailing implies that everyone threw out the ad before looking at it. Those who don't see the ad need to have prior expectations, A_0 and ν_0, in order to form expectations about how others will behave. In the sequel, we take the advertiser's strategic choices for A and ν as given, leaving these considerations (and the question of what is credible) to another paper.

The functions used to evaluate expected utility, and the information used to make the evaluation, are therefore:

	consume	don't consume
see ad	$EV(t_i+\nu, k \mid A, \nu, A_0, \nu_0, t_i)$	$EV(0, 1-k \mid A, \nu, A_0, \nu_0, t_i)$
don't see ad	$EV(t_i, k \mid A_0, \nu_0, t_i)$	$EV(0, 1-k \mid A_0, \nu_0, t_i)$

There are now two cutoffs, since those who see ads have more information and a different value function from who don't. Label T_A^* as the cutoff after viewing advertising and T_0^* as the cutoff without advertising.

Theorem 43 *The no-advertising cutoff T_0^* is decreasing in A_0 and ν_0. The cutoff with advertising, T_A^*, is decreasing in A, ν, A_0, and ν_0.*

This means that more people consume as the As and νs rise. The observation that an effective ad campaign (that is, a high value of A and ν) can expand consumption is not at all surprising. What makes this statement distinct from the results of a model based on purely private utility is that *expectations* about advertising (A_0 and ν_0) affect the actions of both those who do and don't see advertising.

5.5.2 Advertising model two

The second manner of modelling advertising focuses on the advertiser's goal of convincing the consumer that members of an imagined community[6] all wish to consume the product. As such, the advertiser presents a distribution $a(m)$, which shows the likelihood that the mode of the distribution is at a given point m. Rarely do we see histograms of tastes in advertising, but a consumer who is familiar with his cohort will know how many of his fellows will benefit from an advertised feature, and thus impute a distribution of medians $a(m)$. The consumer then acts on that distribution, rather than on the desirable characteristic itself.

From here, the results above apply directly. Notably, Theorem 42 defines when one advertisement is 'better' than another: when the distribution of tastes implied by one MLRPs the other.

Theorem 43 can also be applied in this setup, where advertising offers an entire distribution instead of a scalar (so we define the ordering $a_1(m) \succ a_2(m)$ iff $a_1(m)$ MLRPs $a_2(m)$), and the key conclusion will be the same: both better advertising and better expectations about advertising will induce more people to consume or act.

[6]Imagined Communities is the title of a book by Benedict Anderson, but my usage has only passing relation to his.

5.6 Extensions

5.6.1 Politics and spatial models

If voter turnout is a convivial activity (see Chapter, 6), then this directly contradicts median-voter type results, which assume that a candidate will do his best to maximize the number of people who sympathize with his position. A candidate who claims to have a small following, all of whom will turn out to vote, may do better than one who has a larger, more tepid, following.

Given this claim, the question of where candidates should place themselves to maximize their share of votes cast becomes a nontrivial one. Such a model of candidate placement may have equilibria where some candidates place themselves at extremes and some place themselves in the center.

5.6.2 Finance

Anywhere increased demand increases the expected payoff for an asset, this model is applicable. Given any of the situations enumerated on page 72, the results here may be used to produce more accurate forecasts of expected payoffs.

The reader will note that the model presented so far falls just short of proving that leptokurtotic returns are caused by conviviality effects: it showed that demand would be leptokurtotic, but this is only have of the supply and demand model which would have to be written out to completely explain prices. Fully specifying the model is not a difficult extension, but is beyond the scope of this chapter.

5.6.3 Competing groups and fashion goods

One possible extension to the model is to take into account the size of the opposing group. Schuessler [49] points out that if 50% of the population is in a group the actor wishes to join, then joining makes a statement about the actor and his beliefs. But if 99% of the population is in that group, then joining is a meaningless default action. This implies that one would enjoy joining a group more if the opposition is large but not too large. Since the percentage of people in the opposition is $1 - k$, this means that the value function $V(t, k)$ would be nonmonotonic in k.

The proofs in this chapter relied heavily on monotonicity, and for good reason: if we allow $V(t, k)$ to be nonmonotonic in k, then virtually any behavior can be explained by some admissible value function, leaving us with a vacuous theory. Therefore, one must be cautious in designing a theory that takes this into account.

The same missing facet that Schuessler discusses also makes this model inappropriate for studying fashion goods *per sé*. The story told by many of the models of fashion goods (and many fashion magazines) is that there are 'fashion leaders' who try to distinguish themselves from others, and the hoi polloi who try to emulate the fashion leaders, but not each other. We would again need a model that includes multiple groups to describe the situation.

5.7 Is TV viewership bimodal?

This section analyzes some data for viewership of TV specials. It finds evidence that the distribution of the number of viewers across programs is not unimodal. This supports the hypothesis that TV viewing is a convivial good.

The 'herding' literature, as discussed on page 71, can not explain this: it says that some viewers know that the show will be good (or not), and they inform others, who then herd along with the information provider. But TV specials are one-time broadcasts, meaning that no one has more information about the show's quality than anyone else. Everyone sees the same advertisements beforehand, and knows that everyone else has as much information as they do. There is no information to be cascaded.

Therefore, the viewership of TV specials, if it is to be modelled as including a conviviality component, requires a simultaneous model such as the one presented here.

5.7.1 Data

I use a set of ratings data furnished to me by Michael Chwe, who obtained the data directly from AC Nielsen. They cover viwership for October 1988, February 1989, and July 1989.

The data specified which programs were one-time specials. Within specials, I excluded three large classes of program, based on the fact that information about the quality of such programs may exist and be unevenly distributed among potential viewers, allowing for information cascades. These classes are TV broadcasts of movies, since those who saw the movie in theatres could initiate an information cascade; sports events, if only because they include many repeating events such as the World Series or Olympics; and spin-offs of existing television programs, such as special episodes of a sitcom broadcast at a special hour or retrospectives of long-running programs.

This paring-down left 86 specials, with data in all of the demographic categories listed below.

5.7.2 The unimodality assumption

The key assumption of the model here is that there is a unimodal distribution of viewers' tastes for any given program. Deviations from this could be due to two possible causes: the distribution of program quality is fundamentally multimodal, or the taste for any given program is multimodal.

A program's quality is an amalgamation of the work of dozens or hundreds of individuals: writers, producers, technicians, set designers, and on-screen talent. Being such an amalgamation, it seems the Central Limit Theorem (CLT) would be ideal: a series of reasonably independent draws are made from dozens of different distributions (the pool of writing talent, set design talent, et cetera), and the program's overall quality is a weighted mean of these individual draws. The CLT then says that the distribution of program quality should be Normal (and therefore unimodal).

5.7.2.1 Why assuming unimodality in tastes furthers the study

One may explain the result that viewership is multimodal by claiming that viewer tastes are multimodal, but this simply restates the question on a different axis: why are tastes multimodal?

Without reference to some desire to emulate others in one's cohort, explanations for multimodal tastes are impossible to make clear.

For example, one common argument is that certain racial or demograpic groups have a preference for certain types of dress or music, which immediately engenders the question: why do some groups prefer one form over the other? This question is especially difficult given that many seemingly disparate forms, such as sephardic oud music and surfer rock, bear a close resemblance in spite of a very disparate following.[7]

Some argue that members of certain groups all share a common upbringing, meaning that all consumers of this generation all had parents and teachers who consumed the same music and wore clothing in the same manner. Of course, this does not answer the question from the previous paragraph, but rephrases it in terms of the previous generation.

A related argument is that search is difficult, and so members of a cohort look to each other to determine what media are of high quality. This is difficult to support because search in the context here, broadcast television in the USA of the 1980s, is as cheap as pushing a button on a remote control. In the context of oud vs. surfer music, the search claim is that people require help from their cohort to determine the quality not of individual performers, but of *entire genres*; this places an undue level of stupidity upon the consumer.

In short, the arguments as to why tastes are multimodal, or that certain demographic groups tend toward certain programs, can not be satisfactorily described without some reference to a simple desire to emulate others—which is the alternate hypothesis tested here. So by assuming a unimodal distribution of tastes, we give the null hypothesis (no desire to emulate others) a chance to be true.

[7]Dick Dale, whose music defined the surfer genre, is Lebanese. His music consists of playing a Middle Eastern melody on an electric guitar, sometimes doubling the rhythm.

The identical consumer story One final suggestion with which the author has often been confronted is as follows: assume that every potential viewer has exactly the same tastes, and exactly the same value of their time. Then everyone will gauge the quality of a program equally: if the program is not worth one's time, it will not be worth anybody's time; and similarly if the program is worthy. Thus, a Normal distribution of program quality can quickly lead to a sharply bimodal distribution of turnouts. However, this model is based on the clearly false assumptions that everyone is identical, and fails to be robust to weakening of that assumption. If people's tastes or values for time are Normally distributed, we have the model of Section 5.4.2, with $k = 0$. As such, a unimodal distribution of turnout will result.

5.7.3 Method

I use Silverman's test of bimodality [51]. It is based on this function: $\hat{f}(t, X) = \frac{\sum_{i=1}^{n} \mathcal{N}((t-X_i)/h)}{n \cdot h}$, where $X_1, X_2, \ldots X_n$ are the data observed, $\mathcal{N}(y)$ is a Normal$(0,1)$ density function evaluated at y, and h is a smoothing parameter. Figure 5.3 shows the effect of raising h on the shape of the fitted curve, using viewership data for all men. When h is very small, the Normal distributions around each data point approach spikes, so there is a mode at every data point. As h rises, the spike around each point spreads out and merges with other spikes, until eventualy the entire data set is subsumed under one single-peaked curve. Thus, there is a monotonic relationship between h and the number of modes in the density function induced by h.

Silverman offers a bootstrap method of testing a null hypothesis of the form 'the distribution has k or fewer modes'. The procedure begins by finding the smallest value of h such that $\hat{f}(t, X)$ has k modes; the null hypothesis is that this value of h (herein h_0) is consistent with the data. Then, we draw a bootstrap sample, X', from the data, write down $\hat{f}(t, X')$ based on h_0, and count its modes. Thanks to the monotonic relationship between h and the number of modes, the percentage of such bootstrap distributions with k or fewer modes matches the likelihood that $\hat{f}(t, X)$ given h_0 is consistent with the data. A full proof is in Silverman[51]. If we reject

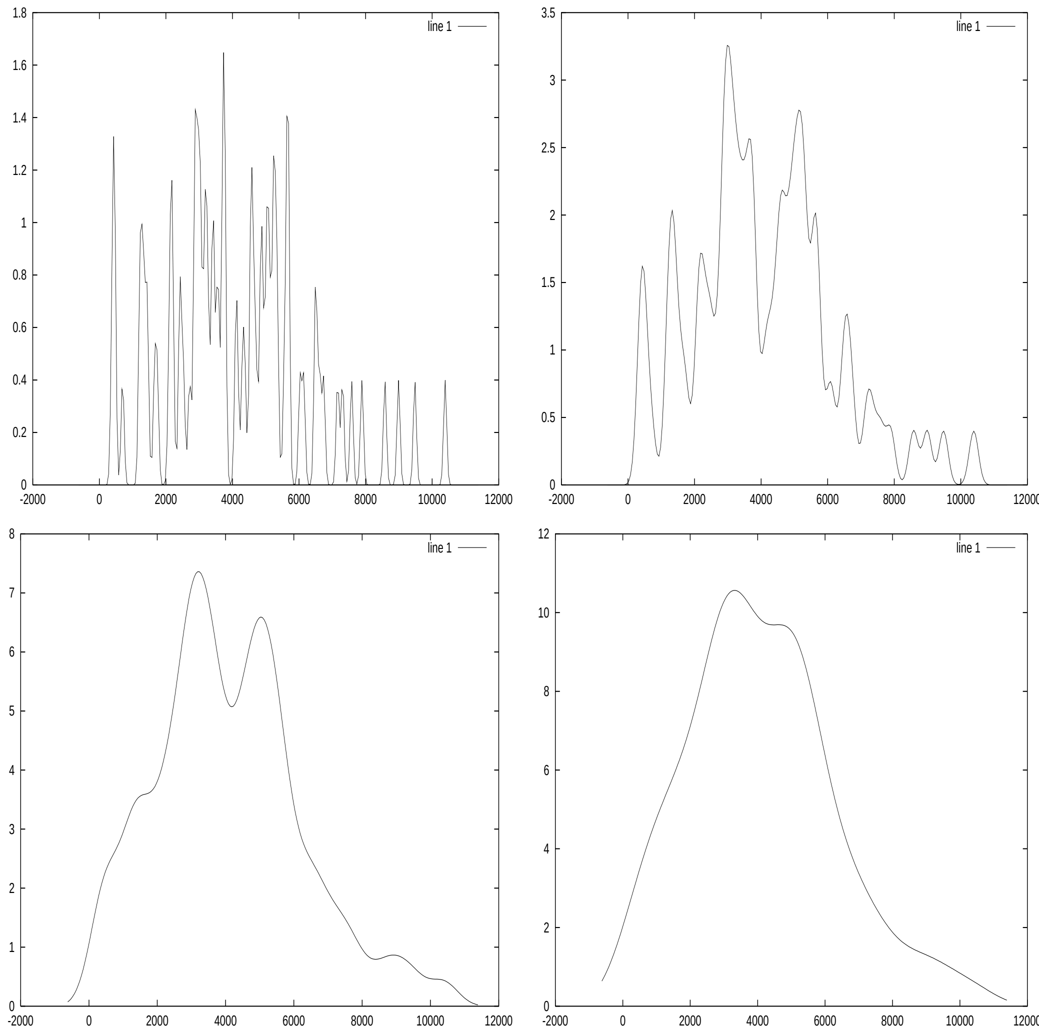

Figure 5.3: As h rises, the kernel density smooths out and has fewer modes.

this hypothesis, then we would need a lower value of h, and therefore more modes, to explain the data.

5.7.4 Results

Figure 5.7.4 lists p-values for tests of k modes, run over a number of different demographic subcategories. The tests were run over 200 bootstrap intervals.

In all but one subcategory, we reject the hypothesis that there is only one mode, and in all but one subcategory, we fail to reject a hypothesis that there are two or more modes. In the one case where we failed to reject the possibility of one mode

max # of modes	All women	Women 18-34	Women 18-49	Women 25-54
1	.385	.69	.91*	.87
2	.875	.985**	.835	.995**
3	.995**	.54	.985**	.78
4	.61	.605	1.00**	.93*
5	.89	.89	.96**	.95**
	All men	Men 18-34	Men 18-49	Men 25-54
1	.705	.465	.66	.73
2	.65	.995**	.94*	.58
3	.47	.95**	.85	.81
4	.365	.60	.61	.66
5	.725	.365	.79	.92*
	All households	East Coast	West Coast	South
1	.35	.18	.28	.64
2	.84	.33	.66	.96**
3	.865	.48	.94*	.62
4	.90*	.85	.92*	.34
5	.645	.94*	.68	.53

Figure 5.4: The p-values for modality tests. Stars indicate significance.

(women 18-49), the test proved especially unpowerful, and failed to reject four out of six of the proposed hypotheses.

The test thus gives evidence that the distribution of viewership is multimodal among many cohorts. Given that there are two or more modes, and making the assumptions discussed above, the model of Section 5.4.2 tells us that k—the desire to emulate others—is large, and clearly nonzero.

5.7.5 Test conclusion

The test fails to reject the hypothesis that the structure of demand for TV specials demonstrates a desire to emulate others. However, any such emulation could not have been derived from an information cascade, since no viewers had more information about the programs prior to their one-time broadcast. The analysis assumed that underlying tastes were unimodal, but it is difficult to explain why that assumption would not be true if people have no interest in emulating other members of their group.

5.8 Conclusion

People gain utility from behaving like others in a multitude of situations. If all individuals act simultaneously, however, the problem of predicting how others will behave typically requires more information than any one person has.

There are two classes of equilibria: one is described by a cutoff level of tastes, above which people act; and the other is characterized by disjoint sets of actors. The first category includes normal distributions in the uninformative prior case and *all* distributions in the full information case, while the second class includes equilibria where the expected demand is somewhere a decreasing function of the median of tastes. In the case where tastes are distributed normally, adding the fact that people like to behave like others doesn't change behavior from the entirely private valuation case, where information is limited. Without additional information, individuals are forced to assume that the average is like them, and the no-change result follows. However, they may switch to a new equilibrium after they know how many others actually are consuming. The new equilibrium will be more likely to be extreme the more consumers care about how others act (or equivalently, if the distribution of tastes has a tall peak); if consumers care enough, then *only* extreme equilibria are possible.

Advertising and prior expectations are self-fulfilling: given a constant personal utility, low expectations of the number of consumers may still prevent a consumer from acting, because he expects that nobody else will act. This is a testable conclusion entirely distinct from a random utility maximization model with no social component; in such a model, after a consumer draws t_i, he doesn't care about the distribution from which it was drawn.

Better advertising can cause more people to consume, as one would expect, but so can better *expectations* about advertising.

Chapter 6

Appendix: Voting as convivial act

Siding with Riker & Ordeshook [47], I will describe the desire to vote as stemming from 'civic duty.' The question is then: where does civic duty come from? One plausible answer is: emulation or agreement with the behavior of one's cohorts. Symbolically, let t be a real number representing a person's private taste for voting and k be the percentage of one's cohort who votes. Then the null hypothesis of Chapter 5 is that we can write the utility from voting as $u(t)$, while the alternate is that we need to write it as $u(t, k)$, where utility is monotonically increasing in both terms. Schuessler [49] describes, but does not detail or test, a similar model.

The way to test this statement is by checking the variance and kurtosis of the distribution of turnouts among different districts. If k is relevant, then the distribution of turnout will be wider, with fatter tails, than if it is not. This is a restatement of Theorem 35.

This appendix tests the empirical claim that actual turnout has a higher variance and a leptokurtotic return relative to a model of self-interested turnout. I will first describe the self-interested model, based on civic duty being normally distributed, then discuss the data used to test this model, and finally ask whether the self-interesed model could plausibly have led to the variance and kurtosis we see in the data.

The conclusion is that a non-convivial model can indeed explain the data. Although it may or may not describe how people actually think, a model which does not explicitly take into account the emulation of others in actions or opinions may be adequate for answering empirical questions.

6.1 Theoretical model

As a baseline, consider the model where utility is entirely private. Let the sense of civic duty, or the taste for the act of voting, be represented by $t \in \mathbb{R}$. This may be negative, for example, if the cost of going to the polls is high. Let the utility from voting be $u(t) = t + n \cdot k$, where n represents the intensity of the desire to emulate others. Within district d, let the mean of t be m_d, and let tastes be distributed $t \sim \mathcal{N}(m_d, 1)$.

Let $erf(x|m_d)$ be the percentage of people below some level x given m_d. The percentage of the population which votes will be $erf(m_d|0)$. This is because everyone with $t > 0$ votes, meaning that $1 - erf(0|m_d)$ people are voting, and the symmetry of the normal distribution allows us to write this as $erf(m_d|0)$. This function can be inverted: if p_d percent voted, then it must be that the district's mean is located at $m_d = erf^{-1}(p_d)$.

But the taste for voting changes from area to area. To take this into account say that $m_d \sim \mathcal{N}(\mu, 1)$, where μ is a parameter to be specified below, and the variance is fixed at one for scaling. Notating the probability density of means m_d as $N_{\mu,\sigma}(m_d)$, the probability of outcome p_d is $N_{\mu,1}\left(erf^{-1}(p_d)\right)$.

So the input into the model is the parameter μ describing the means of tastes among districts. The output is a density function of turnout, with support between zero and one. Statistics about this density function, such as its mean, variance, and kurtosis, may be found by direct calculation. If these statistics describe a Normal distribution, then we may conclude that $n = 0$; if the statistics describe something wider than a Normal, then Theorem 35 tells us that $n \neq 0$.

6.2 Data

I used the National Election Study's cumulative data file from 1948-2000. For certain years, the researchers went to registrars' offices and checked for the name of survey respondents on the rolls, confirming whether or not he or she had voted. This is the measure I used as my data, which gave me 6,836 confirmed votes/nonvotes.[1] I divided these by their district and year. After throwing out district/years which claimed 100% turnout or had 10 or fewer respondents, I had 346 observations of district/years and their corresponding percent turnout. A histogram of the turnouts is shown in the left-hand histogram of Figure 6.1.

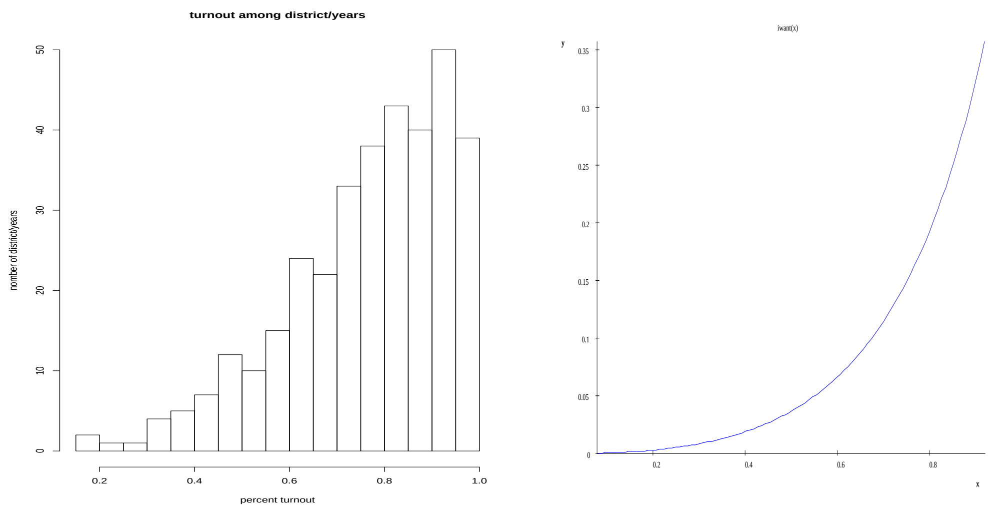

Figure 6.1: NES turnout by district/year

Figure 6.2: A private-utility model of turnout by cohort

6.2.1 Calibrating the model

Recall that the mean of district means, μ, needs calibration from the model above. We need only one value of μ which leads to a reasonable posterior density to fail to reject the hypothesis; I selected $\mu = 2.4$, resulting in the posterior density of turnout drawn in the right-hand graph of Figure 6.2, and the model parameters in the table below:

[1] Another measure, which directly asked people if they voted or not, was hopelessly flawed and revealed 100% turnout in a large plurality of districts.

	model	data	data std. dev
mean	.757	.753	.0092
variance	.0217	.0247	.0024
skew	-0.00419	-0.00393	.00076
kurtosis	.00212	.00222	.00046

The standard deviations were produced by the bootstrap method; the other data is from direct calcuation of the appropriate moments. Notice that these moments are the average of certain pieces of data (such as $(x-\mu)^4$, in the case of kurtosis), and so the Central Limit Theorem dictates that these statistics will be normally distributed. The usual interpretations of the standard deviation and confidence intervals therefore apply.

The theoretical mean, skew and kurtosis are all within a standard deviaton of the data's actual values, and the variance is within two standard deviations. Since there was only one parameter to the model, the fact that four statistics match the data is a nontrivial result. Therefore, I fail to reject the null hypothesis: the data can be explained by a model with no conviviality effects and normally distributed random tastes for voting.

In conclusion, although it seems reasonable that there are cohorts of people who tend to influence each other in voting or not voting together, the data can be explained using a simple model which does not explicitly include this effect.[2] If the desire to emulate others were a sufficiently strong independent factor, then the variance and kurtosis of the distribution would be larger than the theoretical distribution with the same mean, but this is not the case.

[2]With cohorts significantly smaller than districts, we may be able to find more evidence for conviviality effects, but the data is not available.

Chapter 7

Appendix: Proofs

I first present some auxiliary results which I feel have little value by themselves, but which are necessary for proving some of the results used in the body of this work; and then prove those results not proven in the body of the dissertation.

7.1 Auxiliary results

7.1.1 For Chapter 2

Proposition 44 *Let $op(f, g)$ be monotonically increasing in g, let $f(x)$ be any PDF, and let $g_1(x)$ and $g_2(x)$ be any pair of PDFs which satisfy single-crossing. Then $post(op, f, g_1, x)$ FOSDs $post(op, f, g_2, x)$ for any such functions iff*

$$\int_{-\infty}^{\infty} op(f, g_1, x)dx = \int_{-\infty}^{\infty} op(f, g_2, x)dx.$$

Proof: First, I will show that this equality condition is sufficient for the posteriors to satisfy FOSD. Since $op(f, g)$ is monotonically increasing in g, it preserves single-crossing, meaning that $op(f, g_1)$ single-crosses $op(f, g_2)$.

Since $\int_{-\infty}^{\infty} op(f(x), g_1(x))dx = \int_{-\infty}^{\infty} op(f(x), g_2(x))dx$, this means that the posteriors

$$\frac{op(f, g_1)}{\int_{-\infty}^{\infty} op(f(x), g_1(x))dx} \text{ and } \frac{op(f, g_2)}{\int_{-\infty}^{\infty} op(f(x), g_2(x))dx}$$

also satisfy single-crossing, and therefore FOSD is satisfied (by Lemma 2).

Say that, contrary to the above premise, there exist three functions $g_1(\cdot)$, $g_2(\cdot)$, and $f(\cdot)$ such that

$$\int_{-\infty}^{\infty} op(f(x), g_1(x))dx > \int_{-\infty}^{\infty} op(f(x), g_2(x))dx. \tag{7.1}$$

FOSD is satisfied when

$$\frac{\int_k^{\infty} op(f(x), g_1(x))dx}{\int_{-\infty}^{\infty} op(f(x), g_1(x))dx} > \frac{\int_k^{\infty} op(f(x), g_2(x))dx}{\int_{-\infty}^{\infty} op(f(x), g_2(x))dx},$$

which we can rewrite as

$$\frac{\int_k^{\infty} op(f(x), g_1(x))dx}{\int_k^{\infty} op(f(x), g_2(x))dx} > \frac{\int_{-\infty}^{\infty} op(f(x), g_1(x))dx}{\int_{-\infty}^{\infty} op(f(x), g_2(x))dx} \equiv 1 + \delta. \tag{7.2}$$

If the premise of the theorem held, then δ would always equal zero. Assuming Inequality 7.1 means that $\delta > 0$, but is constant with respect to k, so if the ratio of $\int_k^{\infty} op(f(x), g_1(x))$ to $\int_k^{\infty} op(f(x), g_2(x))$ approaches one for some sequence of ks, then FOSD will not be not satisfied.

If it is not the case that this ratio approaches one, then we may easily construct a set of distributions where it does. Recall that one of the conditions on $op(f(x), g_1(x))$ was that its integral over all $x \in \mathbb{R}$ be finite. For this to be true, it must be that $\int_c^{\infty} op(f(x), g_1(x))$ approaches zero as c approaches ∞. Therefore, there is some c such that

$$\frac{\int_c^{\infty} op(f(x), g_1(x))dx}{\int_{-\infty}^{\infty} op(f(x), g_2(x))dx} < \frac{\delta}{2}.$$

Now define

$$g_1'(x) = \begin{cases} g_1(x) & x < c \\ \text{any } y \text{ such that } op(f(x), y) = \left(1 + \frac{\delta}{4}\right) op(f(x), g_2(x)) & x \geq c \end{cases}.$$

That is, $g_1'(x)$ is the same as $g_1(x)$ for all x up to c, and then takes on any value such that we are assured that $op(f(x), g_1'(x))$ is slightly larger than $op(f(x), g_2(x))$ for all

$x \geq c$. Single-crossing is still satisfied, and any offending discontinuity may be easily smoothed out.

Notice that

$$\begin{aligned}\frac{\int_{-\infty}^{\infty} op(f(x), g_1(x))dx}{\int_{-\infty}^{\infty} op(f(x), g_2(x))dx} &> \frac{\int_{-\infty}^{c} op(f(x), g_1(x))dx}{\int_{-\infty}^{c} op(f(x), g_2(x))dx} \\ &> 1 + \frac{\delta}{4}, \end{aligned} \tag{7.3}$$

since $g_1'(x)$ is at its smallest if it were zero above c, and c was defined so that Inequality 7.3 is true.

We now have

$$\frac{\int_{c}^{\infty} op(f(x), g_1'(x))dx}{\int_{c}^{\infty} op(f(x), g_2(x))dx} < 1 + \frac{\delta}{4},$$

while

$$\frac{\int_{-\infty}^{\infty} op(f(x), g_1'(x))dx}{\int_{-\infty}^{\infty} op(f(x), g_2(x))dx} = 1 + \frac{\delta}{2}.$$

Inequality 7.2 is not satisfied for $g_1'(x)$, $g_2(x)$, and $k = c$, meaning that FOSD is not satisfied for these functions.

If the reverse of Inequality 7.1 is true, then there are two ways we can construct single-crossing functions which do not lead to FOSD posteriors. One is to repeat the above procedure, instead modifying the left-hand tail of $g_2(x)$. A neater way is to let $f^m(x) = f(-x)$, $g_1^m(x) = g_2(-x)$, and $g_2^m(x) = g_1(-x)$. By taking the mirror image in this way, we still have $g_1^m(x)$ single-crossing $g_2^m(x)$, and

$$\int_{-\infty}^{\infty} op(f^m(x), g_1^m(x))dx = \int_{-\infty}^{\infty} op(f(x), g_2(x))dx$$

and

$$\int_{-\infty}^{\infty} op(f^m(x), g_2^m(x))dx = \int_{-\infty}^{\infty} op(f(x), g_1(x))dx.$$

meaning that either Inequality 7.1 is true or if it is false, then

$$\int_{-\infty}^{\infty} op(f^m(x), g_1^m(x))dx > \int_{-\infty}^{\infty} op(f^m(x), g_2^m(x))dx.$$

The conditions up to Inequality 7.1 are now satisfied, and we may apply the procedure which followed that inequality to show that there exist functions $f^m(x)$, $g_1^{m\prime}(x)$, and $g_2(x)$ which do not satisfy FOSD.

In conclusion, if there exist three distributions such that

$$\int_{-\infty}^{\infty} op(f(x), g_1(x))dx \neq \int_{-\infty}^{\infty} op(f(x), g_2(x))dx,$$

then either FOSD is not satisfied for the distributions as given, or new PDFs can be constructed for which FOSD does not hold. $\diamondsuit$

7.1.2 For Chapter 5

Lemma 45 *Define $f(t|m)$ to be a translation family based on a continuous PDF, where m is the translation parameter. Define the difference between the utility from acting and not acting (given a fixed T^A) as:*

$$D(t_i, T^A) \equiv \int_{-\infty}^{\infty} V\left(t_i, k(T^A, \tau)\right) post(\tau|t_i)d\tau - \int_{-\infty}^{\infty} V\left(0, 1 - k(T^A, \tau)\right) post(\tau|t_i)d\tau.$$

$D(t_i, T^A)$ is a continuous function of t_i for any fixed T^A.

Proof: We need to show that for all ϵ, there is some δ such that:

$$\begin{aligned} \epsilon &> D(t, T^A) - D(t - \delta, T^A) \\ &= \int_{-\infty}^{\infty} V\left(t_i, k(T^A, \tau)\right) post(\tau|t_i) - V\left(0, 1 - k(T^A, \tau)\right) post(\tau|t_i) - \\ &\quad \left[V\left(t_i - \delta, k(T^A, \tau)\right) post(\tau|t_i - \delta) - V\left(0, 1 - k(T^A, \tau)\right) post(\tau|t_i - \delta)\right] d\tau. \end{aligned}$$

I will show this by dividing the space into the interior, $[U, L]$; and the tails, $(-\infty, L)$ and (U, ∞); where L and U are functions of δ. Then, I will find an appropriate δ, such that both the interior and the tails are bounded, even given the worst-case value function.

The center Within any given range $\tau \in [L, U]$, both $V(t, k(T^A, \tau))$ and $V(0, 1 - k(T^A, \tau))$ are bounded and continuous, as is $post(m|t)$, so

$$\int_L^U V\left(t_i, k(T^A, \tau)\right) post(\tau|t_i) d\tau, \quad \int_L^U V\left(0, 1 - k(T^A, \tau)\right) post(\tau|t_i) d\tau,$$

and their difference are continuous in t_i. Therefore, given L, U, and ϵ, we can find a δ such that

$$\begin{aligned} \epsilon/2 \;\; > \;\; & \int_L^U V\left(t_i, k(T^A, \tau)\right) post(\tau|t_i) - V\left(0, 1 - k(T^A, \tau)\right) post(\tau|t_i) - \\ & \left[V\left(t_i - \delta, k(T^A, \tau)\right) post(\tau|t_i - \delta) - V\left(0, 1 - k(T^A, \tau)\right) post(\tau|t_i - \delta)\right] d\tau. \end{aligned}$$

The worst case Let $\Delta(t, \delta) \equiv D(t, T^A) - D(t - \delta, T^A)$. For a given condtional PDF $post(m|t)$ and fixed δ, $\Delta(t, \delta)$ would be at its largest for a $k(T^A, \cdot)$ such that

$$\begin{cases} k(T^A, \tau) = 1, k(T^A, \tau - \delta) = 0 & \forall \tau \ni post(\tau|t_i) > post(\tau|t_i - \delta) \\ k(T^A, \tau) = 0, (T^A, \tau - \delta) = 1 & \forall \tau \ni post(\tau|t_i) < post(\tau|t_i - \delta) \end{cases}$$

Let $M(t_i, \delta)$ be the maximum of

$$|V(t_i, 1) - V(0, 0) - [V(t_i - \delta, 0) - V(0, 1)]|$$

or

$$|V(t_i, 0) - V(0, 1) - [V(t_i - \delta, 1) - V(0, 0)]|.$$

Since $M(t_i, \delta)$[1] is not a function of τ, it comes out of the integral:

$$\Delta(t, \delta) \leq M(t_i, \delta) \int_{-\infty}^{\infty} |post(\tau|t_i) - post(\tau|t_i - \delta)| d\tau. \tag{7.4}$$

The tails Since $post(\tau)$ is a PDF with only a finite number of discontinuities, $post(\tau) \to 0$ as $\tau \to \infty$ or $\tau \to -\infty$, meaning that for any ϵ and any δ, there must

[1]The reader may be uncomfortable with the fact that as we search for an appropriate value of δ, $M(t_i, \delta)$ changes. But the reader may then pick an arbitrary value, say seven, and is assured by the monotonicity of $V(t, k)$ that $M(t_i, \delta) \in [M(t_i, 0), M(t_i, 7)]$ for all $\delta \leq 7$. The integral in Equation 7.4 is therefore bounded above by a similar expression with $M(t_i, \delta)$ replaced by $\max[M(t_i, 0), M(t_i, 7)]$

be some interval $[L(\delta), U(\delta)]$ such that outside of that interval, both

$$\int_{-\infty}^{L(\delta)} post(\tau|t_i)d\tau + \int_{U(\delta)}^{\infty} post(\tau|t_i)d\tau < \frac{\epsilon}{2M(t_i,\delta)},$$

and

$$\int_{-\infty}^{L(\delta)} post(\tau|t_i - \delta)d\tau + \int_{U(\delta)}^{\infty} post(\tau|t_i - \delta)d\tau < \frac{\epsilon}{2M(t_i-\delta,\delta)}.$$

Since the sum in Inequality 7.5 is bounded above for both $post(\tau|t_i)$ and $post(\tau|t_i - \delta)$, their difference (as in the integral of Inequality 7.4) is also bounded above by $\max(\frac{\epsilon}{2M(t_i,\delta)}, \frac{\epsilon}{2M(t_i-\delta,\delta)})$. This means that the part of $\Delta(t,\delta)$ pertaining to $(-\infty, L) \cup (U, \infty)$ is bounded by $\epsilon/2$. Above, I showed that we can also find an appropriate δ such that the part of $\Delta(t,\delta)$ pertaining to $[L, U]$ is also bounded by $\epsilon/2$. So there exists a δ such that the sum of the two parts is bounded by ϵ, which was to be shown. $\diamondsuit$

Lemma 46 *If $m_1 \neq m_2$, then $T^*(m_1) \neq T^*(m_2)$.*

Proof: Assume $m_1 > m_2$, and that $T^*(m_1) = T^*(m_2)$. Since $T^*(m_1)$ and $T^*(m_2)$ are assumed to be the same, let T^* signify both. One who drew $t_i = T^*$ would be indifferent between action and inaction given either m_1 or m_2, meaning that

$$V(T^*, 1 - F(T^*|m_1)) = V(0, F(T^*|m_1)) \quad (7.5)$$

and

$$V(T^*, 1 - F(T^*|m_2)) = V(0, F(T^*|m_2)) \quad (7.6)$$

But $m_1 > m_2$ and the monotonicity of $V(t,k)$ immediately implies that

$$V(T^*, 1 - F(T^*|m_1)) > V(T^*, 1 - F(T^*|m_2))$$

and

$$V(0, F(T^*|m_2)) > V(0, F(T^*|m_1))$$

meaning that only one of Equations 7.5 or 7.6 can be true. $\diamondsuit$

Lemma 47 *Say two lines, defined by $y = ax + k_a$ and $y = bx + k_b$, intersect at the point (x_1, y_1). Then a one unit horizontal shift in the first line, to $y = a(x-1) + k_a$ leads to a horizontal shift in the point of intersection of $\frac{a}{b-a}$, so the x-coordinate of the new point of intersection is $x_1 + \frac{a}{b-a}$.*

Proof: For the entertainment of the reader, I will present two proofs, an algebraic calculation and a geometric demonstration.

Algebraically, we can solve the initial system of equations to find that the point of intersection for $y = ax + k_a$ and $y = bx + k_b$ is at the point

$$x_1 \equiv \frac{k_b - k_a}{a - b},$$

while the intersection for the lines $y = a(x-1) + k_a$ and $y = bx + k_b$ is at the point

$$\begin{aligned} x_2 &\equiv \frac{k_b - k_a + a}{a - b} \\ &= x_1 + \frac{a}{a - b}, \end{aligned} \tag{7.7}$$

which was to be shown.

Geometrically, the situation is illustrated in Figure 7.1: the line of slope b is shifted laterally by one unit, so the bottom line is given to have length one. The slope of the hypotenuse of the bottom triangle is b, so the ratio of the vertical length to the horizontal length is b. Since the horizontal length is one, the vertical length must be b.

For the next triangle up, we now know that the vertical length is b, and the ratio of vertical to horizontal is a, so the horizontal leg must have length b/a. We can continue going from lower right to upper left, calculating the length of one leg at a time, giving the lengths as shown in the diagram.

So given that the line has shifted by one unit, the intercept between the line of slope a and the lines of slope b has shifted laterally by:

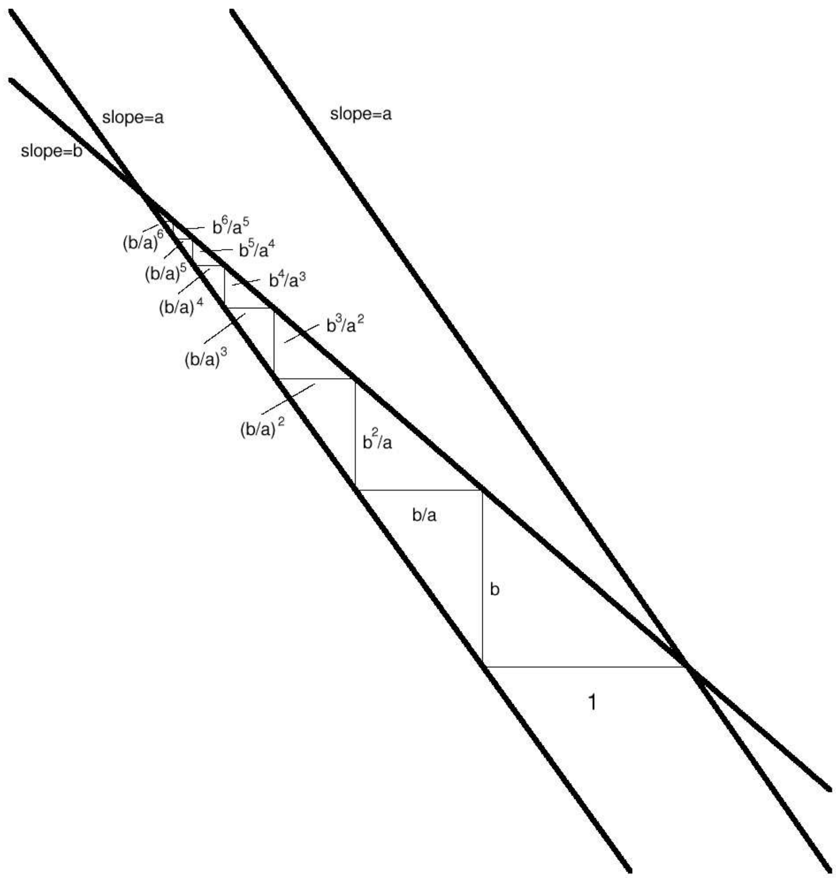

Figure 7.1: Shifting the line of slope a one unit to the left shifts the point of intersection by more than one unit.

$$\begin{aligned} 1+\left(\frac{b}{a}\right)+\left(\frac{b}{a}\right)^2+\left(\frac{b}{a}\right)^3+\ldots &= \sum_{i=0}^{\infty}\left(\frac{b}{a}\right)^i \\ &= \frac{1}{1-\frac{b}{a}} \\ &= \frac{a}{a-b}, \end{aligned}$$

as above. $\diamondsuit$

7.2 Proofs for Chapter 2

Lemma 1 *Let $t \in \mathbb{R}$, and $p_1(t)$ and $p_2(t)$ be continuous PDFs. Define K to be the set of ts such that $p_1(t)/p_2(t) = 1$. If $p_1(t)$ MLRPs $p_2(t)$, then K is a single point. Also, $p_1(t)/p_2(t) < 1$ for all $t < K$ and $p_1(t)/p_2(t) > 1$ for all $t > K$. In other words, MLRP $\Rightarrow$single-crossing.*

Proof: If $K = \{\emptyset\}$, then it must be that either $p_1(t) > p_2(t)$ for all t or $p_1(t) < p_2(t)$ for all t. The first case implies that

$$\int_{-\infty}^{\infty} p_1(t)dt > \int_{-\infty}^{\infty} p_2(t)dt,$$

but since both integrals must equal one, this is a contradiction. Similarly for the second ($<$) case. Therefore $K \neq \{\emptyset\}$. [Note that this is true for all continuous PDFs, without regard to the MLRP.]

If there are two points $t_1, t_2 \in K$, meaning $p_1(t_1)/p_2(t_1) = p_1(t_2)/p_2(t_2)$, then the MLRP is violated.

By the MLRP, if $t < K$, then $p_1(t)/p_2(t) < p_1(K)/p_2(K)$; since $p_1(K)/p_2(K) = 1$, this means that $p_1(t)/p_2(t) < 1$. Similarly for $t > K$. Thus, $p_1(t)$ and $p_2(t)$ satisfy single-crossing. $\diamondsuit$

Lemma 2 *If $p_i(t)$ is a family of single-crossing distributions then the family satisfies FOSD.*

Proof: Take K defined as in Lemma 1, and consider two members of the family, $p_1(t)$ and $p_2(t)$. Then for any point $k > K$, $p_1(t) > p_2(t)$ for all $t > k$, so

$$\int_{k}^{\infty} p_1(t)dt > \int_{k}^{\infty} p_2(t)dt$$

and FOSD is demonstrated. Now considering $k \leq K$,

$$\int_{-\infty}^{k} p_1(t)dt < \int_{-\infty}^{k} p_2(t)dt \tag{7.8}$$

Since these are PDFs, they integrate to one, and so Inequality 7.8 is equivalent to

$$1 - \int_k^\infty p_1(t)dt < 1 - \int_k^\infty p_2(t)dt$$

Subtracting one and negating both sides demonstrates FOSD again. $\diamondsuit$

Lemma 3 *Given $x \in \mathbb{R}$ and $t \in \mathbb{R}$, and a functions $f(x,t)$ with full support over the ranges of x and t. Then if $f(x,t)$ satisfies the MLRP with respect to t [that is, $f(x,2)$ MLRPs $f(x,1)$], then it also satisfies the MLRP with respect to x [that is, $f(2,t)$ MLRPs $f(1,t)$].*

Proof: Let $t_1 > t_2$ and $x_1 > x_2$. We assume that

$$\frac{f(x_1,t_1)}{f(x_2,t_1)} > \frac{f(x_1,t_2)}{f(x_2,t_2)}.$$

Cross multiplying gives

$$\frac{f(x_1,t_1)}{f(x_1,t_2)} > \frac{f(x_2,t_1)}{f(x_2,t_2)},$$

proving our result. $\diamondsuit$

Proposition 10 *The posterior distribution $post(f, g_1, op, x)$ MLRPs $post(f, g_2, op, x)$ iff the function $op(f, g_1, x)$ MLRPs the function $op(f, g_2), x$.*

Proof: MLRP means that the ratio

$$\frac{op(f(x), g_1(x))}{op(f(x), g_2(x))}$$

is increasing in x. Of course, multiplying the ratio by a constant (in terms of x) won't change this, so

$$\frac{op(f(x), g_1(x))}{op(f(x), g_2(x))} \cdot \frac{\int_{-\infty}^{\infty} op(f(y), g_2(y))dy}{\int_{-\infty}^{\infty} op(f(y), g_1(y))dy}$$

is also increasing in x. But this is the ratio of the posteriors, so the posteriors satisfy the MLRP.

The steps reverse to show the 'only if' part of the statement. ◇

Lemma 11 *A continuous operator preserves the MLRP iff it is of the form*

$$op(a,b) = t(a) \cdot b^p,$$

with $p > 0$, and $t(\cdot)$ any transformation.

Proof: First, consider four positive real numbers, a, b, c, and d, and a continuous function $\phi : \mathbb{R} \to \mathbb{R}$ such that

$$\frac{a}{b} > \frac{c}{d} \Rightarrow \frac{\phi(a)}{\phi(b)} > \frac{\phi(c)}{\phi(d)} \tag{7.9}$$

for all positive a, b, c, and d. What can be said about the function $\phi(\cdot)$?

Setting $b \equiv (ad/c) + \epsilon$, $ad > cb$, so $\phi(a)\phi(d) > \phi(c)\phi(b)$; setting $b \equiv (ad/c) - \epsilon$, $ad < cb$, so $\phi(a)\phi(d) < \phi(c)\phi(b)$; so by continuity, it must be the case that $\phi(a)\phi(d) = \phi(c)\phi(b)$ whenever $ad = cb$. This means that $\phi(a)\phi(d)$ is a function only of the product ad, and can be expressed as a new one-argument function $\pi(ad) \equiv \phi(a)\phi(d)$.

Taking the derivative of this identity gives two new equations:

$$\phi'(a)\phi(d) = \pi'(ad) \cdot d$$

and

$$\phi(a)\phi'(d) = \pi'(ad) \cdot a$$

Rearranging:

$$\frac{\phi'(a)}{\phi(a)} \cdot a = \frac{\phi'(d)}{\phi(d)} \cdot d.$$

Rewrite $\phi'(a)/\phi(a)$ as $d\ln(\phi(a))/da$, and notice that this equation holds for any a and d. So for some constant p and for any $x \in \mathbb{R}$,

$$\frac{d\ln(\phi(x))}{dx} \cdot x = p.$$

Integrating p/x and exponentiating gives:

$$\phi(x) = Cx^p,$$

where $\ln C$ is the constant of integration.

Let $\alpha > \beta$ and take $a = g_1(\alpha)$, $b = g_2(\alpha)$, $c = g_1(\beta)$, $d = g_2(\beta)$, and $\phi(x) = op(f, g, x)$. Then making these substitutions into Implication 7.9 gives:

$$\frac{g_1(\alpha)}{g_2(\alpha)} > \frac{g_1(\beta)}{g_2(\beta)} \Rightarrow \frac{op(f, g_1, \alpha)}{op(f, g_2, \alpha)} > \frac{op(f, g_1, \beta)}{op(f, g_2, \beta)}. \tag{7.10}$$

That is: if $g_1(\cdot)$ MLRPs $g_2(\cdot)$, then $op(f(\cdot), g_1(\cdot), \cdot)$ MLRPs $op(f(\cdot), g_2(\cdot), \cdot)$. The proof shows that this can only be the case when $op(f, g)$ is of the form Cg^p. The constant C may be a function of f but not of g.

This proves that any function which preserves the MLRP must be a monomial.

To show that any monomial updating operator preserves the MLRP, let $g_1(x)/g_2(x)$ be an increasing function of x. Then

$$\frac{Cg_1^p(x)}{Cg_2^p(x)} = \left(\frac{g_1(x)}{g_2(x)}\right)^p$$

must also be increasing, so long as $p > 0$. $\diamondsuit$

Proposition 13 *Given any pair of priors g_1 and g_2 which satisfy single-crossing, and any function f, and an updating operator $op(f, g)$ which is monotonically increasing in f and g, the operator $op(\cdot, \cdot)$ provides posteriors ordered by FOSD iff the function is of the form*

$$op(f, g) = t(f) + qg,$$

where $t(\cdot)$ is a transformation function of any form, and q is any positive constant.

Proof: We need only prove that

$$\int_{-\infty}^{\infty} op(f, g_1, x)dx = \int_{-\infty}^{\infty} op(f, g_2, x)dx \tag{7.11}$$

if and only if $op(f, g, x)$ can be written in the form $t(f) + qg$; Proposition 44 (page 113) proves the rest.

To show that the integral of $t(f(x)) + qg(x)$ is constant in changes in $g(\cdot)$, we need only break down the integral:

$$\begin{aligned}\int_{-\infty}^{\infty} t(f(x)) + qg(x)dx &= \int_{-\infty}^{\infty} t(f(x))dx + \int_{-\infty}^{\infty} qg(x)dx \\ &= \int_{-\infty}^{\infty} t(f(x))dx + q.\end{aligned}$$

This is constant with respect to $g(\cdot)$.

To prove the other direction, I use the following three sample functions to show that if the updating operator is not of the form given here, then Equation 7.11 does not hold (and so posteriors will not satisfy FOSD, by Proposition 44).

$$\begin{aligned}f(x) &= \begin{cases} \frac{1}{\epsilon} & x \in [0, \epsilon] \\ 0 & x \notin [0, \epsilon] \end{cases} \\ g_1(x) &= \begin{cases} \frac{1}{\epsilon} & x \in [0, \epsilon] \\ 0 & x \notin [0, \epsilon] \end{cases} \\ g_2(x) &= \begin{cases} \frac{1}{\epsilon} & x \in [1, 1+\epsilon] \\ 0 & x \notin [1, 1+\epsilon] \end{cases}\end{aligned}$$

For simplicity, $g_1(x)$ and $g_2(x)$ are weakly single-crossing, but they may be smoothed out to functions which are strictly single-crossing.

Here are some useful integrals, which hold for all $n, m > 0$:

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} g_1(x)dx = \int_{-\infty}^{\infty} g_2(x)dx &= 1 \\ \int_{-\infty}^{\infty} f^n(x)g_1^m(x)dx &= \left(\frac{1}{\epsilon}\right)^{(n+m-1)}\end{aligned}$$

$$\int_{-\infty}^{\infty} f^n(x) g_2^m(x) dx = 0$$

Now consider any function $op(a, b)$. Weirstrass's polynomial approximation theorem tells us that any continuous function can be approximated arbitrarily well by a polynomial of the form

$$op(a, b) = \sum_{\substack{i=0,1,2,\ldots \\ j=0,1,2,\ldots}} k_{ij} a^i b^j$$

When will $\int_{-\infty}^{\infty} op(f(x), g_1(x))dx = \int_{-\infty}^{\infty} op(f(x), g_2(x))dx$? If the expansion includes terms of the form $k_{i0} f^i(x)$, where $g_1(x)$ or $g_2(x)$ do not appear, these will be equal between both integrals. A term of the form $k_{01} g(x)$ will integrate to one in both cases. But any other term will differ from one integral to the other. Letting ϵ get arbitrarily large will make this difference as large as wished, outstripping the error term in the polynomial approximation. So if $op(a, b)$ is not of the form given in the theorem, there exist functions $f(x)$, $g_1(x)$, and $g_2(x)$ such that $g_1(x)$ and $g_2(x)$ are single-crossing, but Equation 7.11 does not hold, and so the posteriors do not satisfy FOSD. $\diamondsuit$

Lemma 14 *An updating operator $op(f, g)$ which is monotonically increasing in f and g preserves single-crossing iff the function is of the form*

$$op(f, g) = t(f) + qg,$$

where $t(\cdot)$ is a transformation function of any form, and q is any positive constant.

Proof: Any pair of single-crossing priors satisfies FOSD (Lemma 2). Therefore, if $op(a, b_1)$ and $op(a, b_2)$ take any pair of functions which satisfies FOSD and return a pair of functions which satisfy FOSD, then this updating operator takes any pair of functions which satisfies single-crossing, and returns a pair of functions which satisfy FOSD. By Proposition 13, it must be the case that this operator is of the linear form given in that proposition.

If $g_1(x) > g_2(x)$, then since

$$\begin{aligned}\int_{-\infty}^{\infty} t(f(x)) + g_1(x)dx &= \int_{-\infty}^{\infty} t(f(x)) + g_2(x)dx, \\ \frac{t(f(x)) + g_1(x)}{\int_{-\infty}^{\infty} t(f(x)) + g_1(x)dx} &> \frac{t(f(x)) + g_2(x)}{\int_{-\infty}^{\infty} t(f(x)) + g_2(x)dx}.\end{aligned}$$

So if $g_1(x) > g_2(x)$ for all $x < K$, where K is the crossing point of the priors, then $post_1(x) > post_2(x)$ for all $x < K$; conversely for all $x \geq K$, meaning that if the priors satisfy single-crossing, so do the posteriors. $\diamondsuit$

Theorem 15 *Within the class of operators $op(f, g)$ which are monotonically increasing in both arguments, an operator preserves FOSD iff it is of the form $op(f, g) = t(f) + qg$, where $t(\cdot)$ is a transformation function of any form and q is a positive constant.*

Proof: Necessity: Single-crossing priors are a subset of FOSD priors, so it must be the case that this operator is in the class of operators described by Proposition 13 (or equivalently, Proposition 44).

Sufficiency is easily checked: if $g_1(x)$ FOSDs $g_2(x)$, then Inequality 7.12 holds for any k; the rest follows from algebra (and the fact that the integral in the denominator of Inequality 7.13 is the same for both g_1 and g_2).

$$\int_k^{\infty} g_1(x)dx > \int_k^{\infty} g_2(x)dx \tag{7.12}$$

$$\int_k^{\infty} t(f(x)) + q\int_k^{\infty} g_1(x)dx > \int_k^{\infty} t(f(x))dx + q\int_k^{\infty} g_2(x)dx$$

$$\int_k^{\infty} t(f(x)) + qg_1(x)dx > \int_k^{\infty} t(f(x)) + qg_2(x)dx$$

$$\frac{\int_k^{\infty} t(f(x)) + qg_1(x)dx}{\int_{-\infty}^{\infty}(f(x)) + qg_1(x)dx} > \frac{\int_k^{\infty} t(f(x)) + qg_2(x)dx}{\int_{-\infty}^{\infty} t(f(x)) + qg_2(x)dx} \tag{7.13}$$

$\diamondsuit$

7.3 Proofs for Chapter 3

Theorem 18 *Assume that the conditions of Assumption 5, Assumption 7, and Assumption 8 hold. Then there exists a Nash equilibrium whose payoff to the advocate is equal to the payoff he would get from revealing full information.*

Proof: The decision maker may set

$$D(S) = \text{argmin}_{w \in \sigma^{-1}(S)} U(D(S^w)). \tag{7.14}$$

Say that the advocate reveals S^2, where $S^w \subset S^2$. Since the decisionmaker is now doing the minimization in Equation 7.14 over a larger set, she will arrive at a value such that $U(D(S^2)) \leq U(D(S^w))$. So the decision maker would rather defect from revealing S^2 to the strategy of revealing S^w, so S^2 is not an equilibrium. So the only equilibrium possible for the advocate has a payoff equal to S^w. Assumption 8 tells us that if the advocate reveals S^w, then the decisionmaker can do no better, and therefore will not defect from this equilibrium. $\diamondsuit$

Proposition 19 *If the true distribution of $x \in [0,1]$ is $f(x)$, the decision maker's prior is $p(x)$, and the decision maker will choose the median of her perceived posterior distribution, then a lobbyist who prefers smaller outcomes to larger will not reveal full information if there exists some interval (a,b) above the median of $f(x)$ such that $f(y) > p(y)$ for all $y \in (a,b)$, and there also exists some interval (c,d) below the median of $f(x)$ such that $f(y) < p(y)$ for all $y \in (c,d)$.*

The same lobbyist will reveal some quantity of information if there exists some interval (a,b) above the median of $p(x)$ such that $p(y) > f(y)$ for all $y \in (a,b)$, and there also exists some interval (c,d) below than the median of $p(x)$ such that $p(y) < f(y)$ for all $y \in (c,d)$.

The inequalities are reversed for a lobbyist who prefers larger outcomes to smaller outcomes.

Proof: If the decision maker has full information, then her posterior matches the true distribution $f(x)$, and the median of the posterior is therefore the median of $f(x)$. Now say the advocate fails to reveal the value of $f(x)$ within (a,b) and (c,d).

Then the posterior given this lesser information matches $f(x)$ outside these regions but has more weight in (c, d) and less in (a, b), so the posterior median favors the advocate more, and the advocate will therefore not reveal full information.

If the decision maker has no information, then her posterior matches her prior, $p(x)$, and the median of the posterior is therefore the median of $p(x)$. Now say the advocate reveals the value of $f(x)$ within (a, b) and (c, d). Then the posterior given this greater information matches $p(x)$ outside these regions but has more weight in (c, d) and less in (a, b), so the posterior median favors the advocate more, and the advocate will therefore prefer to reveal the information about $f(x)$ inside (a, b) and (c, d). $\diamondsuit$

Proposition 20 *A right-leaning lobbyist will reveal full information iff $p(x)$ is non-decreasing. The lobbyist will reveal no information iff $p(x)$ is nonincreasing.*

The reverse is true for a left-leaning lobbyist.

Proof: If there is some region $[a, b]$ where the density function is decreasing, then the mean of $p(x)$ for $x \in [a, b]$ is less than the mean of the uniform distribution over the same area. Therefore the lobbyist is better off not reporting this region. Conversely, for any area over which $p(x)$ is increasing, the mean is larger if the lobbyist reveals $p(x)$ than if he doesn't. $\diamondsuit$

7.4 Proofs for Chapter 4

Theorem 23 *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; there exist conditions and a judge's opinion such that the defendant would not present S_m;

and all of these conditions may simultaneously be true.

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

$$V(\{S_d, S_p\}) = \frac{\alpha}{1 - PA(\{\neg S_d, \neg S_p\})}, \tag{7.15}$$

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 7.15.]

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:

$$V(\{S_d, S_m, S_p\}) = \frac{\alpha' + \beta}{1 - PA(\{\neg S_d, \neg S_p, \neg S_m\})}, \tag{7.16}$$

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 7.15, and the expected payoff after S_m is presented($EV(\{S_d, S_m, S_p\})$), as in Equation 7.16. 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 7.15 is at a maximum and Equation 7.16 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:

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

$$v(S_m) > \frac{PA(S_p)}{PA(\{\neg S_d,S_m\})}\frac{[1-PA(\{\neg S_d,\neg S_p,\neg S_m\})]}{[1-PA(\{\neg S_d,\neg S_p\})]} - \frac{PA(\{S_p,\neg S_m\})}{PA(\{\neg S_d,S_m\})}. \tag{7.17}$$

When Inequality 7.17 does not hold, then there exist circumstances such that the plaintiff prefers that S_m not be revealed. Notice that 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:

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

As above, we may now move the leftmost terms of the left side to the right side, giving:

$$
\begin{array}{rcl}
\frac{PA(\{\neg S_p, S_m\})v(S_m)}{[1 - PA(\{\neg S_d, \neg S_p, \neg S_m\})]} & < & \frac{PA(\{\neg S_d, S_p\})}{[1 - PA(\{\neg S_d, \neg S_p\})]} - \frac{PA(\{\neg S_d, S_p, \neg S_m\}) + PA(\{S_p, S_m\})}{[1 - PA(\{\neg S_d, \neg S_p, \neg S_m\})]} \\
v(S_m) & < & \frac{PA(\{\neg S_d, S_p\})}{[1 - PA(\{\neg S_d, \neg S_p\})]} \cdot \frac{[1 - PA(\{\neg S_d, \neg S_p, \neg S_m\})]}{PA(\{\neg S_p, S_m\})} \\
 & & - \frac{PA(\{\neg S_d, S_p, \neg S_m\}) + PA(\{S_p, S_m\})}{PA(\{\neg S_p, S_m\})}.
\end{array}
\tag{7.18}
$$

Under reasonable assumptions, neither Constraints 7.17 nor 7.18 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 7.17 reduces to $v(S_m) > \frac{4}{3}$, and Constraint 7.18 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 24 *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 7.17, which we can now write as:

$$L_p = \frac{PA(S_d)}{(PA(S_p) - PA(S_d) - PA(S_d)PA(S_p))(PA(S_d) - 1)}. \tag{7.19}$$

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

Taking the derivative of the plaintiff's limit in Equation 7.19 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 23. 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 25 *Assume the same situation as Theorem 23, 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:[2]

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

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

[2] For consistency with the proof in Theorem 23, 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 $(PA(S_p)(1 - PA(\{\neg S_d, \neg S_m\}))/PA(\{\neg S_d, S_m\}))$.

move.

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

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

$\diamondsuit$

7.5 Proofs for Chapter 5

Lemma 26 *For any function $U(\cdot,\cdot)$ which is monotonically increasing in both arguments, there exists a function $V(\cdot,\cdot)$ which is also monotonically increasing in both arguments, such that if $U(t,k) > U(s, 1-k)$, then $V(t-s, k) > V(0, 1-k)$.*

Proof: Let $k_1^*(t,s)$ be the value of k such that $U(t, k_1^*(t,s)) = U(s, 1 - k_1^*(t,s))$. Let $\phi(t,s) = 1 - 2k_1^*(t,s)$, and let $V(\tau, k) = \tau + k$. Substituting $\phi(t,s)$ for τ, we see that a decision maker using V is indifferent when $1 - 2k_1^*(t,s) + k = 1 - k$, or when $k = k_1^*(t,s)$. That is, a decision maker using V and the transformed taste parameter $\tau = \phi(t,s)$ is indifferent when he would have been indifferent using the original utility function.

The reader may verify that the decision maker would also prefer acting or not acting in the same manner using both functions.

Finally, notice that $f(t,s)$ is monotonically increasing in t and monotonically decreasing in s, and $V(\tau,k) = \tau + k$ is monotonically increasing in τ. $\diamondsuit$

Lemma 27 *Define $f(t|m)$ to be a translation family*[3] *based on a PDF with a finite number of discontinuities. Then in equilibrium, T^A is a (possibly empty) set of open*

[3] Translation families were defined in Section 2.2.1.

intervals.

Proof: The premises of Lemma 45 (page 116) are met, and so $D(t|T^A)$ (defined therein) is a continuous function of t for any given T^A.

T^A comprises those t_is such that $D(t_i, T^A) > 0$. Since $D(t, T^A)$ is continuous, it must be that if $D(t_i, T^A) > 0$, there is some neighborhood where $D(t_i \pm \epsilon, T^A)$ is also greater than zero, so all points in the neighborhood of t_i are also in T^A. $\diamondsuit$

Notice that if $D(t)$ were based on a set T^A which included singleton points or other closed intervals, T^A simply would not be consistent with itself—some neighborhood around those singletons would still want to act. Put another way, if it were announced to the society that some arbitrary set T was going to act, this would inspire to action only those who drew t from some open set of intervals.

Theorem 28 *Assume that $f(\cdot|m)$ and post$(\cdot|m)$ satisfy FOSD.*[4] *Then: $k(T^A, m)$ is monotonically increasing in $m \Leftrightarrow T^A$ is a cutoff equilibrium.*

Proof: $\Rightarrow$ Say that person i does act, and consider another person j, with $t_j > t_i$.

- Since $V(t,k)$ is monotonically increasing in t, $V(t_j,k) > V(t_i,k)$ for all k, so

$$\int_0^1 V(t_j,\kappa)\, p_j(\kappa) d\kappa > \int_0^1 V(t_i,\kappa)\, p_j(\kappa) d\kappa.$$

- Since (by assumption) there is a monotonic transformation $k(T^A, t)$ from t to k, and $post(\cdot|t_i)$ satisfies FOSD on the range of t, it also satisfies FOSD on the range of k, where its transformation is notated as $p_i(\cdot)$. Since $V(t_i,k)$ is an increasing function of k, FOSD dictates that

$$\int_0^1 V(t_i,\kappa)\, p_j(\kappa) d\kappa > \int_0^1 V(t_i,\kappa)\, p_i(\kappa) d\kappa.$$

[4]FOSD was defined in Section 2.2.1.

- Person i will act iff

$$\int_0^1 V\left(t_i, \kappa\right) p_i(\kappa) d\kappa > \int_0^1 V\left(0, 1-\kappa\right) p_i(\kappa) d\kappa.$$

- Since $V\left(0, 1-k\right)$ is monotonically decreasing in k, FOSD gives:

$$\int_0^1 V\left(0, 1-\kappa\right) p_i(\kappa) d\kappa > \int_0^1 V\left(0, 1-\kappa\right) p_j(\kappa) d\kappa.$$

Chaining all four of these inequalities together gives:

$$\int_0^1 V\left(t_j, \kappa\right) p_j(\kappa) d\kappa > \int_0^1 V\left(0, 1-\kappa\right) p_j(\kappa) d\kappa,$$

so that person j will also act. Since all that was assumed was that $t_j > t_i$, this proves that if $t_i \in T^A$, then $[t_i, \infty) \in T^A$. Similarly, if $t_n < t_i$, and i does not act, then neither will n, so $t_i \in T^N$ implies $(-\infty, t_i] \in T^N$. This leaves a single point for T^*.

$\Leftarrow$ If T^A is a range of the form (t^*, ∞), then $k(T^A, t)$ is simply

$$\int_{t^*}^{\infty} f(x|t) dx.$$

FOSD says that this integral is increasing in t for fixed t^*, which was to be shown. $\diamondsuit$

Proposition 29 *If the posteriors $post(\cdot|m)$ satisfy FOSD, and the likelihood function $f(\cdot)$ is continuous, then $t^*(t_i)$ is a continuous, increasing function of t_i.*

Proof: It must be shown that if $t_i > t_j$, then $t^*(t_i) > t^*(t_j)$. Given that $post(\cdot|t_i)$ FOSDs $post(\cdot|t_j)$ if $t_i > t_j$, Theorem 28 says that $k\left((t^*(t_i), \infty), \tau\right)$ is monotonically increasing in τ. Say that one who draws t_i is indifferent at the cutoff $t^*(t_i)$, so that Equation 5.2 holds, and now consider $t_j > t_i$. Since $V\left(t_i, k((t^*(t_i), \infty), \tau)\right)$ is monotonically increasing in τ, FOSD says that:

$$\int_{-\infty}^{\infty} V\left(t_i, k((t^*(t_i), \infty), \tau)\right) post(\tau|t_j) d\tau \geq \int_{-\infty}^{\infty} V\left(t_i, k((t^*(t_i), \infty), \tau)\right) post(\tau|t_i) d\tau \tag{7.20}$$

Similarly, $V\left(0, 1 - k((t^*(t_i), \infty), \tau)\right)$ is decreasing in τ, so FOSD gives:

$$\int_{-\infty}^{\infty} V\left(0, 1 - k((t^*(t_i), \infty), \tau)\right) post(\tau|t_i) d\tau \geq \int_{-\infty}^{\infty} V\left(0, 1 - k((t^*(t_i), \infty), \tau)\right) post(\tau|t_j) d\tau \tag{7.21}$$

Adding in monotonicity of $V(t, k)$ with respect to t, so $V(t_j, k) > V(t_i, k)$ for all k, and chaining together Equations 7.20, 5.2, and 7.21, we conclude that

$$\int_{-\infty}^{\infty} V\left(t_j, k((t^*(t_i), \infty), \tau)\right) post(\tau|t_j) d\tau > \int_{-\infty}^{\infty} V\left(0, 1 - k((t^*(t_i), \infty), \tau)\right) post(\tau|t_j) d\tau.$$

Therefore, where i was indifferent at the cutoff $t^*(t_i)$, j strictly prefers acting. So $t^*(t_j)$ must be greater than $t^*(t_i)$.

Notice that if $post(\cdot|t)$ has only a finite number of discontinuities and no point masses, then both of the integrals in Equation 5.2 are continuous, and $t^*(t_i)$ is a continuous function of t_i. $\diamondsuit$

Corollary 30 *Given that the posteriors $post(\cdot|m)$ satisfy FOSD and the likelihood function $f(\cdot)$ is continuous; then the cutoff-type pure strategy Bayesian Nash equilibrium T^* is at the value(s) of t_i for which $t^*(t_i) = t_i$. There is always at least one cutoff equilibrium, either at some finite value or at $T^* = \infty$.*

Proof: This simply states a necessary condition for a Bayesian Nash equilibrium. It is sufficient because everyone who draws $t_i > T^*$ will want to act and everyone who draws $t_i < T^*$ won't, by FOSD and monotonicity.

There will always be some value of $T^* \in [-\infty, \infty]$ which is an equilibrium. Proposition 29 showed that $t^*(t_i)$ is a decreasing function of t. This affords only two possibilities: $t^*(t_i) = t_i$ at some number of points, in which case these are all potential equilibria; or $t^*(t_i) > t_i$ for all values of t_i, in which case $T^* = \infty$ is a consistent cutoff (since it will inspire no one to act). The possibility that $t^*(t_i) < t_i$ for all values of

t_i, in which case everyone acts and T^* would equal $-\infty$, is precluded by Assumption 19. $\diamondsuit$

Corollary 31 *The cutoffs for cutoff-type pure strategy Bayesian Nash equilibria are at the value(s) of t_i for which $k^*(t_i) = 1 - F_i(t_i)$. If $k^*(t_i) > 1 - F_i(t_i)$ for all t_i, then $K^* = 1$ is an equilibrium (where no one acts).*

Proof: It is clearly necessary that the person at the cutoff be indifferent between acting and not acting, by definition, so at the cutoff, $t^*(t_i) = t_i$. This is sufficient because everyone who draws a t_i such that $k^*(t_i) > K^*$ will want to act and everyone who draws t_i such that $k^*(t_i) < K^*$ won't, by FOSD and monotonicity.

There will always be some value of $K^* \in (-\infty, \infty]$ which is an equilibrium. There are only two possibilities: $k^*(t_i) = 1 - F(t_i)$ at some number of points, in which case these are all potential equilibria; or $k^*(t_i) > 1 - F(t_i)$ for all values of t_i, in which case setting $K^* > 1$ is a consistent cutoff (since it will inspire no one to act). The possibility that $k^*(t_i) < 1 - F(t_i)$ for all values of t_i, in which case everyone acts and $K^* < 0$, is precluded by Assumption 19. $\diamondsuit$

Lemma 32 *Given full information, all perfect strategy Bayesian Nash equilibria are cutoff equilibria.*

Proof: In the full information case, it is known with certainty that the parameter m takes on the value μ, and so the percentage of actors given a set of intervals T^A (not necessarily a cutoff) is know to be $k(T^A, \mu)$. Now say that a person with a draw of t_1 is indifferent between action and inaction, i.e.:

$$V(t_1, k(T^A, \mu)) = V(0, 1 - k(T^A, \mu)).$$

Then by the monotonicity of $V(\cdot, \cdot)$, it must be that for all people who draw $t_i < t_1$,

$$V(t_i, k(T^A, \mu)) < V(t_1, k(T^A, \mu)) = V(0, 1 - k(T^A, \mu)),$$

so they therefore prefer not acting over acting. For all people who draw $t_j > t_1$,

$$V(t_j, k(T^A, \mu)) > V(t_1, k(T^A, \mu)) = V(0, 1 - k(T^A, \mu)),$$

so they prefer to act. Therefore t_1 is a cutoff, and T^A must be (t_i, ∞).

If the cutoff is $t^*(t_i) = \infty$, this is defined to mean that for all t_i, $V(t_i, k(T^A, \mu)) < V(0, 1 - k(T^A, \mu))$. $\diamondsuit$

Proposition 33 *The PDF $f(\cdot|m)$ may take any form, and the parameter m will be known to have the value μ with certainty. If for any equilibrium point $t = \tau$, $f(t|\mu) < -\frac{\partial k^*}{\partial t}(t)$ for all $t > \tau$, then there will be only one equilibrium for the given μ.*

Proof: For the sake of contradiction, say that there is another point $\tau' > \tau$ which could be an equilibrium cutoff, meaning that $1 - F(\tau'|\mu) = k^*(\tau')$. Writing this equation as the integrals of derivatives (and breaking the integrals into parts at τ),

$$1 - \left[\int_{-\infty}^{\tau} f(t|\mu)dt + \int_{\tau}^{\tau'} f(t|\mu)dt\right] = \int_{-\infty}^{\tau} \frac{\partial k^*}{\partial t}(t)dt + \int_{\tau}^{\tau'} \frac{\partial k^*}{\partial t}(t)dt \tag{7.22}$$

At τ, $1 - F(\tau|\mu) = k^*(\tau)$, or

$$1 - \int_{-\infty}^{\tau} f(t|\mu)dt = \int_{-\infty}^{\tau} \frac{\partial k^*}{\partial t}(t)dt$$

meaning that Equation 7.22 holds iff

$$\int_{\tau}^{\tau'} f(t|\mu)dt = -\int_{\tau}^{\tau'} \frac{\partial k^*}{\partial t}(t)dt$$

But if $f(t|\mu) < -\frac{\partial k^*}{\partial t}(t)$ for all $t > \tau$, then the left-hand side of this equation is always less than the right-hand side. Therefore τ can not be a cutoff. $\diamondsuit$

Theorem 34 *The PDF $f(\cdot|m)$ may take any form, and the parameter m will be known to have the value μ with certainty. If $f(t) < -\frac{\partial k^*}{\partial t}(t')$ for all t, t', then there* *will be only one equilibrium for any μ.*

Proof: Given the condition here, the conditions of Proposition 33 will hold for any draw of μ. $\diamondsuit$

Theorem 35 *The ex ante distribution of the median m is $a(m)$, which can take any form. After drawing μ from $a(m)$, μ is known with certainty. Assume the PDF $f(x)$ is single-peaked, with a maximum value of $f(x) = \frac{1}{n}$, and the value function is linear: $V(t,k) = nk + t$. Let $d(T^*)$ be the ex ante distribution of T^* (before μ is known).*

Given these assumptions, the ratio $a(m)/d(T^*(m))$ is single-peaked with a peak at the point where $\mu = T^*(\mu)$. That is, $d(T^*)$ is less concentrated toward the center than $a(m)$.

Proof: The transformation from the PDF $a(m)$ to the PDF $d(\cdot)$ is a simple coordinate transformation:

$$d(T) = a(T^{*-1}(T))\frac{dT^{*-1}}{dm}(T)$$

or

$$\frac{a(T^{*-1}(T))}{d(T)} = \frac{dT^*}{dm}(m) \tag{7.23}$$

A change in m is best envisioned using Figure 5.2 (page 88): a one-unit change in m translates the curve one unit to the left; $\frac{dT^*}{dm}$ is then how far to the left the intercept between the line and the curve moves. The slope of one minus the CDF is $-f(T^*(m)|m)$, and the slope of the function $k^*(m)$ is $-\frac{1}{2n}$. Plugging these slopes into Equation 7.7 from Lemma 47 (page 119), we get:

$$\frac{dT^*}{dm}(m) = \frac{\frac{1}{2n}}{\frac{1}{2n} - f(T^*(m)|m)}. \tag{7.24}$$

The premise of Theorem 34 guarantees that the denominator is always positive, meaning that a positive shift in m always leads to a finite, positive shift in T^*.

Equation 7.23 says that Equation 7.24 gives us the ratio $\frac{a(m)}{d(T^*(m))}$ which the Theorem describes. Since $f(\cdot)$ is assumed to be a single-peaked distribution, $f(T^*(m)|m)$

is largest where $T^*(m) = m$, and is monotonically decreasing as m diverges from that value in either direction; the left-hand side of Equation 7.24—and therefore the left-hand side of Equation 7.23—is also largest where $T^*(m) = m$ and is monotonically decreasing as m diverges from that point. $\diamondsuit$

Proposition 36 *Given the premises of Theorem 34, $T^*(\mu)$ is a strictly decreasing function.*

Proof: Let μ be the value of m which induces an equilibrium at τ, and let $\tau' > \tau$. The proof of Theorem 34 showed that Equation 7.22 should actually be an inequality:

$$1 - \int_{-\infty}^{\tau'} f(t|\mu)dt < \int_{-\infty}^{\tau'} \frac{\partial k^*(t)}{\partial t} dt, \qquad (7.25)$$

where μ is the median which leads to an equilibrium at T^*. Notice that the right-hand side is not a function of μ, while the left-hand side is an increasing function of μ. Therefore, if the two sides of Inequality 7.25 are equal, they will be for some value $\mu' > \mu$, meaning that the equilibrium value $\tau' > \tau$ can only occur given a median $\mu' > \mu$. In other words, the mapping from m to T^* is an increasing function; along with the fact from Lemma 46 (page 118) that $T^*(m)$ is one-to-one, this means that its inverse $T^*(m)$ is itself an increasing function. $\diamondsuit$

Proposition 37 *Assume uninformative prior information; that $f(t|0)$ is single-peaked and either symmetric or upward-leaning; and that $f(t|m)$ is a translation family. Then the only pure strategy Bayesian Nash equilibria are of a cutoff form.*

Proof: The proof is based on the fact that any interval of T^A that is bounded above and below must be bounded by two elements of T^*, t_1 and t_2. By the monotonicity of $V(\cdot, \cdot)$, if $t_2 > t_1$, but someone who drew either would be indifferent, it must be that one who draws t_1 expects more people to act than one whe drew t_2. This is not the case for the lowest interval in T^A. [Assumption 19 assures us that such a lowest interval exists.] The proof will mostly concentrate on the function $k(T^A, \cdot)$; the reader is referred to Figure 7.2 for depictions of the various dissections and translations.

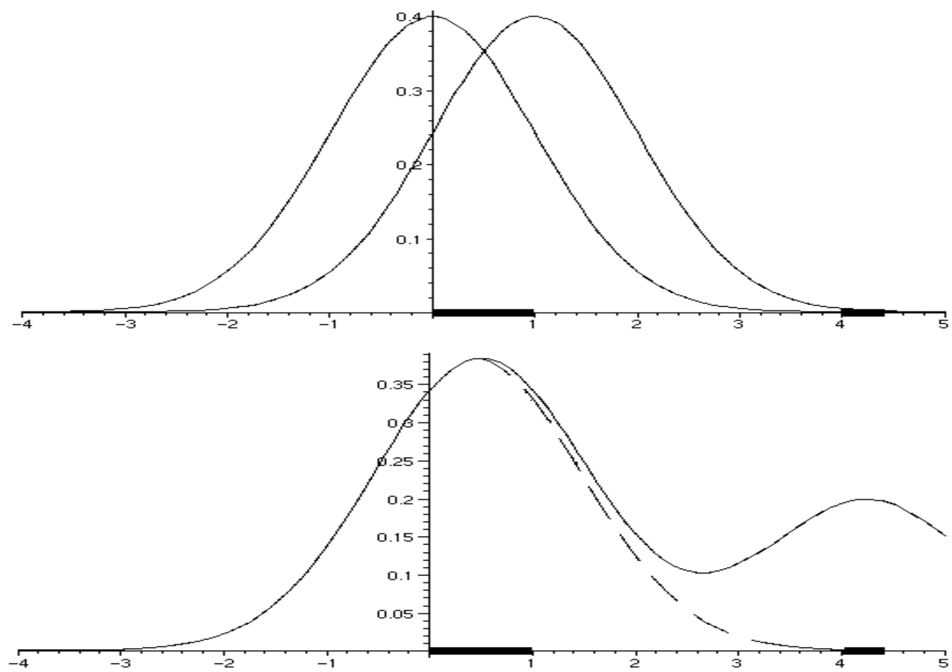

Figure 7.2: top: symmetric, single-peaked distributions centered at zero and one
bottom: $k(T^A, t)$; the mirror image at C of $k(T^A, t)$ for $t < C$ (middle dotted line)
both: T^A includes $A_1 = (0, 1)$ and $(4, 4.5)$

As proven in Lemma 27, T^A consists of a set of open intervals. Let A_1 be the lowest of these, and t_1 to be its infimum and t_2 its supremum. They are elements of T^*, but it suffices to know that they are not in T^A and all points between them are. It will be proven that t_2 actually should be in T^A, giving a contradiction which can only be resolved if A_1 is of the form (t_1, ∞).

Let the center of A_1, $(t_1 + t_2)/2$, be C. Let the crossing-point of $post(\cdot|t_1)$ and $post(\cdot|t_2)$ be X. If $f(\cdot|0)$ is symmetric, then $X = C$; if $f(\cdot|0)$ is upward-leaning, the posteriors are downward-leaning (as per the discussion on page 94), so $X < C$.

Since $post(\cdot|t_1)$ and $post(\cdot|t_2)$ are translations of the same single-peaked distribution, C is their unique crossing point.

$\triangleleft$ *Fact #1:* For $t \leq C$, $k(T^A, t)$ is an increasing function of t. *Why:* Shifting the single-peaked posterior distribution to the right means that the likelihood placed on everything to the right of the distribution's mode increases, which includes all of T^A for $x \leq t_1$, and all of $T^A \backslash A_1$ for $x > t_1$ and $x \leq C$. Symmetry/downward-leaning along with single-peakedness imply that $k(A_1, t)$ is increasing for $t \in [t_1, C]$. ▷

◁ *Fact #2:* $k(T^A, C-d) \leq k(T^A, C+d)$ for all $d > 0$. *Why:* By the symmetry/downward-leaning condition, for the integrals of the posteriors within A_1:

$$\int_{t_1}^{t_2} post(x|C+d)dx \geq \int_{t_1}^{t_2} post(x|C-d)dx,$$

with equality in the symmetric case. For those points above t_2,

$$\int_{T^A \backslash A_1} post(x|C+d)dx > \int_{T^A \backslash A_1} post(x|C-d)dx,$$

since $post(\tau|t_2) > post(\tau|t_1)$ for all $\tau > C$. ▷

By monotonicity, these facts about the function $k(T^A, \tau)$ also hold for t_1's value of action, $V\left(t_1, k(T^A, \tau)\right)$.

From here, break down the function $V\left(t_1, k(T^A, \cdot)\right)$ into a sum of two subfunctions:

$$m(t) = \begin{cases} V\left(t_1, k(T^A, t)\right) & t \leq C \\ V\left(t_1, k(T^A, 2C-t)\right) & t \geq C \end{cases}$$

$$n(t) = \begin{cases} 0 & t \leq C \\ V\left(t_1, k(T^A, t)\right) - V\left(t_1, k(T^A, 2C-t)\right) & t \geq C \end{cases}$$

The function $m(t)$ is symmetric about C, while $n(t)$ is the asymmetric part. In the lower graph of Figure 7.2, $V(T^A, k)$ is the upper solid line, $m(t)$ is the middle dotted line and $n(t)$ is the distance between the middle dotted line and the upper solid line. By fact #1, $m(t)$ is monotonically increasing up to C and therefore monotonically decreasing above C, and by fact #2, $n(t)$ is always positive.

◁ *Fact #3:* The expectation of $m(\cdot)$ given $post(\cdot|t_2)$ equals the expectation given $post(\cdot|t_1)$. *Why:* The situation is symmetric around C. ▷

◁ *Fact #4:* The expectation of $n(\cdot)$ given $post(\cdot|t_2)$ is greater than or equal to the expectation given $post(\cdot|t_1)$. *Why:* $n(t_2) > 0$, while $n(t_1) = 0$. ▷

Facts #3 and 4 show that the sum $m(t_2) + n(t_2)$ is greater than the sum $m(t_1) + n(t_1)$, meaning

$$V\left(t_1, k(T^A, t_2)\right) > V\left(t_1, k(T^A, t_1)\right); \tag{7.26}$$

by the monotonicity of $V(\cdot,\cdot)$,

$$V\left(t_2, k(T^A, t_2)\right) > V\left(t_1, k(T^A, t_2)\right). \tag{7.27}$$

Chaining together Inequalities 7.26 and 7.27,

$$V\left(t_2, k(T^A, t_2)\right) > V\left(t_1, k(T^A, t_1)\right),$$

and we may similarly prove that

$$V\left(0, 1 - k(T^A, t_1)\right) < V\left(0, 1 - k(T^A, t_1)\right),$$

so it can't be the case that people at both t_1 and t_2 are indifferent. If $t_1 \in T^*$, then t_2 must be in T^A; since this would be true for any lower upper bound of T^A, T^A can have no upper bound. $\diamondsuit$

Proposition 39 *Define $f(t|m)$ to be a translation family. If individuals have uninformative prior information about the true center of the distribution, then there is a unique cutoff equilibrium.*

Proof: Since we are considering the class of cutoff equilibria, $k(T^A, t)$ is assumed to be a monotonically increasing, invertible function of t, meaning that there is an inverse $t(T^A, k)$ which gives the unique value of k such that if k percent act, it must be that $m = t$. The distribution of possible values of k for person i, given that t_i is the cutoff, can therefore be described as:

$$\begin{aligned} p_i(k) &= post(t(T^A, k)|t_i) \\ &= post(t((t_i, \infty), k)|t_i) \end{aligned} \tag{7.28}$$

$$= \quad f(t_i|t((t_i,\infty),k)) \tag{7.29}$$

$$= \quad f(0|t((0,\infty),k)) \tag{7.30}$$

Equation 7.28 expands T^A to (t_i, ∞) under the assumption that t_i is the cutoff. Equation 7.29 restates Equation 5.5, that the posterior given an uninformative prior is the mirror of the likelihood function. Equation 7.30 follows from the translation property of the whole operation.

In conclusion, $p_i(k)$ does not depend on i or t_i, so it can be written without a subscript as one common function $p(k)$. Therefore, in the comparison between expected utility from acting and not acting,

$$\int_0^1 V\left(t_i, \kappa\right) p(\kappa) d\kappa \qquad \text{vs} \qquad \int_0^1 V\left(0, 1-\kappa\right) p(\kappa) d\kappa, \tag{7.31}$$

the left-hand side is monotonically increasing in t_i, since the expression $V\left(t_i, \kappa\right) p(\kappa)$ is monotonically increasing in t_i for a given κ, so the integral of these expressions for all κ must also be increasing, while the right-hand side now makes no mention of t_i and is thus a constant. A monotonically increasing function is equal to a constant at at most one point, so there is at most one equilibrium cutoff point. As discussed above, if the left-hand side of Comparison 7.31 is always less than the right-hand side, then no one would be willing to act if they were pivotal, and so the equilibrium cutoff is $T^* = \infty$. $\diamondsuit$

Theorem 40 *Assume $f(t|m)$ is a translation family of a PDF that is continuous at all but a finite number of points, and that the prior is uninformative.*

(i: no conviviality) In the symmetric case $[f(d) = f(-d)]$, the unique equilibrium satisfies $T^* = 0$, meaning that there is neither active nor inactive conviviality. That is, if the distribution is symmetric, people behave as if they only cared about their private signals.

(ii: active conviviality) If the distribution is upward-leaning, meaning that $f(d) = [f(-d)/A]$ for all $d > 0$ and some $A < 1$, then the unique cutoff equilibrium satisfies $T^* < 0$, meaning that everyone with a private signal to act will do so, but some people

with negative private signals will ignore them and act anyway.

(iii: inactive conviviality) If the distribution is downward-leaning, so $f(d) = [f(-d)/A]$ *for all* $d > 0$ *and some* $A > 1$*, then the unique cutoff equilibrium satisfies* $T^* > 0$*, meaning that everyone with a private signal not to act will not act, but some people with positive private signals will ignore them and not act as well.*

Proof: (i) One who drew t_i is indifferent iff

$$\int_{-\infty}^{\infty} V(t_i, k((t_i,\infty),\tau))post(\tau|t_i)d\tau = \int_{-\infty}^{\infty} V(0, 1-k((t_i,\infty),\tau))post(\tau|t_i)d\tau \quad (7.32)$$

This would be equivalent if we ran the variable of integration backward, so let $\theta = -\tau$:

$$\int_{-\infty}^{\infty} V(0, 1-k((t_i,\infty),\tau))post(\tau|t_i)d\tau = -\int_{\infty}^{-\infty} V(0, 1-k((t_i,\infty),\theta))post(\theta|t_i)d\theta$$

But by symmetry, $1 - k((t_i,\infty),\theta) = k((t_i,\infty),\tau)$, and since the posterior is also symmetric, $post(\theta|t_i) = post(\tau|t_i)$. Making these substitutions and again substituting $\tau = -\theta$,

$$-\int_{\infty}^{-\infty} V(0, 1-k((t_i,\infty),\theta))post(\theta|t_i)d\theta = \int_{-\infty}^{\infty} V(0, k((t_i,\infty),\tau))post(\tau|t_i)d\tau. \quad (7.33)$$

We have thus solved for t_i in the left-hand side of Equation 7.32: the right-hand side of Equations 7.33 says that it is zero.

(ii) The proof of the skewed case is by comparison with an artificial symmetric distribution. We will mostly be interested in the case where $t_i = 0$, so wlog assume this is the case. Then define the symmetric function $\pi(\cdot)$ to be $post(x)$ when $x \geq 0$ and the mirror image $\pi(x) = post(-x)$ when $x < 0$. Then $post(\cdot)$ is now defined by:

$$post(x) = \begin{cases} \pi(x) & x \geq 0 \\ \frac{\pi(x)}{A} & x < 0. \end{cases}$$

Notice that $\pi(\cdot)$ is not really a PDF—it doesn't integrate to one:

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(x)dx \\ &= \int_{-\infty}^{0} \frac{\pi(x)}{A}dx + \int_{0}^{\infty} \pi(x)dx \\ &= (1+\frac{1}{A})\int_{-\infty}^{0} \pi(x)dx \\ &= \frac{(1+\frac{1}{A})}{2}\int_{-\infty}^{\infty} \pi(x)dx \end{aligned}$$

So we need a scaling factor

$$S = \frac{1+\frac{1}{A}}{2} = \frac{1+A}{2A}$$

so that $\int S\pi(x)dx$ will indeed integrate to one. Then the expected value of acting for someone with $t_i = 0$, if they were the cutoff and tastes had distribution $S\pi(\cdot)$, would be:

$$\int_{-\infty}^{\infty} V\left(0, \int_{0}^{\infty} S\pi(x|\tau)dx\right) S\pi(-\tau)d\tau$$

Notice that, as discussed above, $post(\tau) = \pi(-\tau)$. We know from (i) above that under this scenario, the person at $t = 0$ is indifferent between action and inaction.

Now returning to the asymmetric case of $post(\cdot)$, we can use the definition of $post(\cdot)$ in terms of $\pi(\cdot)$ to write the expected value of acting as:

$$\begin{aligned} &\int_{-\infty}^{0} V\left(0, \int_{0}^{\infty} \pi(x-\tau)dx\right) \pi(-\tau)d\tau + \\ &\int_{0}^{\infty} V\left(0, \int_{0}^{\tau} \frac{\pi(x-\tau)}{A}dx + \int_{\tau}^{\infty} \pi(x-\tau)dx\right) \frac{\pi(-\tau)}{A}d\tau. \end{aligned} \tag{7.34}$$

Why: as per the definition of a translation family in Section 2.2.1 , we can write $\pi(x|\tau)$ as $\pi(x-\tau|0)$, which, by the notational standard that $\pi(\cdot) \equiv \pi(\cdot|0)$, is what appears in the integrals above. In the first integral, $\tau < 0$, so the posterior likelihood $post(\tau) = f(-\tau) = \pi(-\tau)$, and since x is always greater than zero, $x - \tau > 0$, so $f(x-\tau) = \pi(x-\tau)$. In the second part, $\tau > 0$, so $f(-\tau) = \pi(-\tau)/A$, but $x - \tau$ is less than zero if $x < \tau$ and greater than zero when $x > \tau$; thus the integral over x

has to be broken down into the two parts shown above.

We now wish to compare the expected value in the symmetric and the asymmetric case. For example, the first half of Expression 7.34 is bounded below by:

$$\int_{-\infty}^{0} V\left(0, \int_{0}^{\infty} \pi(x-\tau)dx\right) \pi(-\tau)d\tau \;>\; \int_{-\infty}^{0} V\left(0, S\int_{0}^{\infty} \pi(x-\tau)dx\right) \pi(-\tau)d\tau$$

because $S < 1$ when $A < 1$ as in the upward-leaning case we are considering, and $V(0,\cdot)$ is monotonically increasing. In the downward-leaning case, $A > 1$, the inequality is reversed.

As for the second half, break down the argument to $V(0,\cdot)$ as follows:

$$\begin{aligned}
\int_{0}^{\infty} f(x-\tau)dx &= \int_{0}^{\tau} \frac{\pi(x-\tau)}{A}dx + \int_{\tau}^{2\tau} \pi(x-\tau)dx + \int_{2\tau}^{\infty} \pi(x-\tau)dx \\
&= \left(1+\frac{1}{A}\right)\left[\int_{0}^{\tau} \pi(x-\tau)dx\right] + \int_{2\tau}^{\infty} \pi(x-\tau)dx \qquad (7.35)\\
&> 2S\left[\int_{0}^{\tau} \pi(x-\tau)dx\right] + S\int_{2\tau}^{\infty} \pi(x-\tau)dx \qquad (7.36)\\
&= S\left[\int_{0}^{\infty} \pi(x-\tau)dx\right]
\end{aligned}$$

Equation 7.35 is due to the symmetry of $\pi(\cdot)$. Inequality 7.36 uses the fact that the coefficient $\left(1+\frac{1}{A}\right)$ is equal to $2S$, and the fact that $S < 1$. The inequality reverses when $S > 1$.

Every step in the following chain of inequalities has now been shown:

$$\begin{aligned}
EV(\text{act}|\text{skew}) &= \int_{-\infty}^{\infty} V\left(0, \int_{0}^{\infty} f(x-\tau)dx\right) post(\tau)d\tau \\
&= \int_{-\infty}^{0} V\left(0, \int_{0}^{\infty} \pi(x-\tau)dx\right) \pi(-\tau)d\tau \\
&\quad + \int_{0}^{\infty} V\left(0, \int_{0}^{\tau} A\pi(x-\tau)dx + \int_{\tau}^{\infty} \pi(x-\tau)dx\right) \frac{\pi(-\tau)}{A}d\tau \\
&> \int_{-\infty}^{0} V\left(0, S\int_{0}^{\infty} \pi(x-\tau)dx\right) \pi(-\tau)d\tau \\
&\quad + \int_{0}^{\infty} V\left(0, S\int_{0}^{\infty} \pi(x-\tau)dx\right) \frac{\pi(-\tau)}{A}d\tau
\end{aligned}$$

$$
\begin{aligned}
&= \left(1+\frac{1}{A}\right)\int_{-\infty}^{0} V\left(0, S\int_{0}^{\infty}\pi(x-\tau)dx\right)\pi(-\tau)d\tau \\
&= S\int_{-\infty}^{\infty} V\left(0, S\int_{0}^{\infty}\pi(x-\tau)dx\right)\pi(-\tau)d\tau \\
&= EV(\text{act}|\text{symmetry})
\end{aligned}
$$

Similarly, $EV(0, 1-k)$ is smaller in the asymmetric case, and so where a person who drew zero was indifferent in the symmetric case, he strictly prefers to act in the asymmetric case. So it must be that $T^* < 0$.

(iii) Reverse the inequalities noted in the proof of (ii). ◇

Theorem 41 *Given that $m = \mu$ is known.*

(i: no conviviality) In the symmetric case, there is an equilibrium at $T^* = 0$, meaning that there is neither active nor inactive conviviality.

(ii: active conviviality) If the distribution is upward-leaning, then there is an equilibrium at some point $T^* < 0$.

(iii: inactive conviviality) If the distribution is downward-leaning, then there is an equilibrium at some point $T^* > 0$.

Proof: Recall that $t^*(0)$ is the point at which one who drew $t_i = 0$ would be indifferent if $t^*(t_0)$ were the cutoff. It is the point where

$$
V(0, \int_{t^*(0)}^{\infty} f(x|m)dx) = V(0, 1 - \int_{t^*(0)}^{\infty} f(x|m)dx)
$$

or

$$
\int_{t^*(0)}^{\infty} f(x|m)dx = \int_{-\infty}^{t^*(0)} f(x|m)dx
$$

In the symmetric case, where $f(x|m) = f(-x|m)$, this is clearly true at $t^*(0) = 0$, and therefore $T^* = 0$ does indeed describe a Bayesian Nash equilibrium.

In the case where $f(x|m) > f(-x|m)$, then the point which solves the above equation must satisfy $t^*(0) > 0$. By Assumption 19, there is some value of t, t_∞, such that the associated cutoff is $t^*(t_\infty) = -\infty$. Since $t^*(t_i)$ is a continuous, monotonic function of t_i, we know that $t_\infty < 0$, and that there is some point $t_e \in (t_\infty, 0)$ such that $t_e = t^*(t_e)$.

In the case where $f(x|m) < f(-x|m)$, then $t^*(0) < 0$. By continuity, there is some range such that $t^*(t_i) < t_i$ for all $t_i \in (0, t_e)$. If this range has an upper bound, then at that bound, $t_e = t^*(t_e)$; if it has no upper bound, then $T^* = \infty$ is an equilibrium. $\diamondsuit$

Theorem 42 *Assume prior distribution $a_1(m)$ satisfies the monotone likelihood ratio property with respect to $a_2(m)$, and assume cutoff equilibria. Assume nothing about the distribution $f(t)$. Then someone who is indifferent between action and inaction given prior $a_2(m)$ will strictly prefer acting given $a_1(m)$.*

Proof: By Theorem 5, the assumptions about the priors imply that for any fixed t_i, the posterior $post(\cdot|t_i, a_1)$ FOSDs the posterior $post(\cdot|t_i, a_2)$. The expression

$$V\left(t_i, \int_{t_i}^{\infty} a_1(x|\tau)dx\right)$$

is an increasing function of τ, by FOSD. Therefore,

$$\int_{-\infty}^{\infty} V\left(t_i, \int_{t_i}^{\infty} a_1(x|\tau)dx\right) post(\tau|t_i, a_1)d\tau > \int_{-\infty}^{\infty} V\left(t_i, \int_{t_i}^{\infty} a_1(x|\tau)dx\right) post(\tau|t_i, a_2)d\tau, \tag{7.37}$$

by FOSD. By FOSD,

$$\int_{-\infty}^{\infty} V\left(t_i, \int_{t_i}^{\infty} a_1(x|\tau)dx\right) post(\tau|t_i, a_2)d\tau > \int_{-\infty}^{\infty} V\left(t_i, \int_{t_i}^{\infty} a_2(x|\tau)dx\right) post(\tau|t_i, a_2)d\tau \tag{7.38}$$

If one who drew t_i and has prior $a_2(\cdot)$ is indifferent, then

$$\int_{-\infty}^{\infty} V\left(t_i, \int_{t_i}^{\infty} a_2(x|\tau)dx\right) post(\tau|t_i, a_2)d\tau = \int_{-\infty}^{\infty} V\left(0, 1 - \int_{t_i}^{\infty} a_1(x|\tau)dx\right) post(\tau|t_i, a_2)d\tau. \tag{7.39}$$

Similarly,

$$V\left(0, 1 - \int_{t_i}^{\infty} f(x|\tau)dx\right)$$

is monotonically decreasing in τ, so by FOSD:

$$\int_{-\infty}^{\infty} V\left(0, 1 - \int_{t_i}^{\infty} a_2(x|\tau)dx\right) post(\tau|t_i, a_2)d\tau > \int_{-\infty}^{\infty} V\left(0, 1 - \int_{t_i}^{\infty} a_2(x|\tau)dx\right) post(\tau|t_i, a_1)d\tau \tag{7.40}$$

and, by FOSD:

$$\int_{-\infty}^{\infty} V\left(0, 1 - \int_{t_i}^{\infty} a_2(x|\tau)dx\right) post(\tau|t_i, a_1)d\tau > \int_{-\infty}^{\infty} V\left(0, 1 - \int_{t_i}^{\infty} a_1(x|\tau)dx\right) post(\tau|t_i, a_1)d\tau \tag{7.41}$$

Chaining together Expressions 7.37, 7.38, 7.39, 7.40, and 7.41 shows that

$$\int_{-\infty}^{\infty} V\left(t_i, \int_{t_i}^{\infty} a_1(x|\tau)dx\right) post(\tau|t_i, a_1)d\tau > \int_{-\infty}^{\infty} V\left(0, 1 - \int_{t_i}^{\infty} a_1(x|\tau)dx\right) post(\tau|t_i, a_1)d\tau.$$

So if one who drew t_i is indifferent given $a_2(\cdot)$, he strictly prefers to act given $a_1(\cdot)$. $\diamondsuit$

Theorem 43 *The no-advertising cutoff T_0^* is decreasing in A_0 and ν_0. The cutoff with advertising, T_A^*, is decreasing in A, ν, A_0, and ν_0.*

Proof: If i believes that the original distribution of tastes was $f_i(\cdot)$, and believes that A_0 percent of consumers have expanded utility by ν_0, then i believes that the post-advertising distribution of tastes (herein $f_i'(t)$) is

$$f_i'(t) = A_0 f_i(t - \nu_0) + (1 - A_0) f_i(t).$$

In other words, A_0 percent of the distribution was shifted over by ν_0 while $1 - A_0$ percent didn't change. Say that the cutoff without advertising was $t^*(t_i)$, and so one who drew t_i is indifferent between action and inaction given no advertising. But (since $F_i(t) - F_i(t-\nu_0) > 0$ if $\nu_0 > 0$)

$$\begin{aligned} 1 - F_i'(t^*(t_i)) &= 1 - [A_0 F_i(t-\nu_0) + (1-A_0)F_i(t)] \\ &= 1 - [A_0 F_i(t) + (1-A_0)F_i(t) - A_0[F_i(t) - F_i(t-\nu_0)]] \\ &> 1 - F_i(t). \end{aligned} \tag{7.42}$$

In Expression 7.42, the first two terms in the brackets sum to $F_i(t)$, and the remaining part within brackets represents the shortfall by which $1 - F_i'(t^*(t_i))$ does not reack $1 - F_i(t)$. So i now believes that there are more people acting given the same cutoff $t^*(t_i)$, and therefore where i was indifferent, i now strictly prefers acting, and the cutoff must fall. Since $A_0[F_i(t) - F_i(t-\nu_0)]$ is increasing in both A_0 and in ν_0, the cutoff $t^*(0)$ must fall further for larger values of A_0 and ν_0.

Similarly, the distribution of tastes given A and ν,

$$f_i''(t) = Af_i(t-\nu) + (1-A)f_i(t),$$

will also have a higher percentage of people over any given cutoff as A and ν rise. Also, if i was indifferent between action and inaction given the cutoffs t_A^* and t_0^*, and A_0 or ν_0 fall, then t_0^* falls, meaning more people will consume, meaning that i strictly prefers acting given the cutoff t_A^*, meaning that cutoff must fall. $\diamondsuit$

Bibliography

[1] George A Akerlof and Rachel E Kranton. Economics and identity. Technical report, 1999.

[2] Lisa R Anderson and Charles A Holt. Information cascades in the laboratory. *The American Economic Review*, 87(5):847–862, December 1997.

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

[4] Abhijit V Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817, Aug 1992.

[5] Berkeley Symposium on Mathematical Statistics and Probability. *Comparison of Experiments*, 1951.

[6] B Douglas Bernheim. A theory of conformity. *Journal of Political Economy*, 102(5):841–877, 1994.

[7] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(51):992–1026, 1992.

[8] Sushil Bikhchandani, Uzi Segal, and Sunil Sharma. Stochastic dominance under Bayesian learning. *Journal of Economic Theory*, 56:352–377, 1992.

[9] William A Brock and Steven N Durlauf. Discrete choice with social interactions. *Review of Economic Studies*, 68:235–260, 2001.

[10] 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.

[11] Colin Camerer and Teck-Hua Ho. Experience-weighted attraction learning in normal form games. *Econometrica*, 67(4):827–874, July 1999.

[12] Colin Camerer, George Loewenstein, and Matt Rabin, editors. *Advances in Behavioral Economics*. Princeton University Press, Forthcoming.

[13] Bruce A Campbell. A theoretical approach to peer influence in adolescent socialization. *American Journal of Political Science*, 24(2):324–344, May 1980.

[14] Yin-Wong Cheung and Daniel Friedman. Individual learning in normal form games: Some laboratory results. *Games and Economic Behavior*, 19:46–76, 1997.

[15] Jay Pil Choi. Herd bahavior, the "penguin effect," and the suppression of informational diffusion: an analysis of information externalities and payoff interdependency. *RAND Journal of Economics*, 28(3):407–425, Autumn 1997.

[16] Michael Suk-Young Chwe. *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton University Press, 2001.

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

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

[19] Louis Eeckhoudt and Christian Gollier. Demand for risky assets and the monotone probability ratio order. *Journal of Risk and Uncertainty*, 11:113–122, 1995.

[20] William N Evans, Wallace E Oates, and Robert M Schwab. Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy*, 100(5):966–991, October 1992.

[21] Paul Frijters. A model of fashions and status. *Economic Modelling*, 15:501–517, 1998.

[22] Drew Fudenberg and David Levine. Learning in games. *European Economic Review*, 42:631–639, May 1998.

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

[24] Kenneth M Goldstein. *Interest Groups, Lobbying, and Participation in America*. Cambridge University Press, 1999.

[25] John R Graham. Herding among investment newsletters: Theory and evidence. *Journal of Finance*, 54(1):237–268, February 1999.

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

[27] Sanford J Grossman. On the efficiency of competitive stock markets where trades [sic] have diverse information. *The Journal of Finance*, 31(2):573–585, May 1976.

[28] Sanford J Grossman. An introduction to the theory of rational expectations under asymmetric information. *Review of Economic Studies*, 48:541–559, 1981.

[29] Harrison Hong, Jeffrey D Kubik, and Amit Solomon. Security analysts' career concerns and herding of earnings forecasts. *RAND Journal of Economics*, 31(1):121–144, Spring 2000.

[30] Tony Honoré. *Responsibility and Fault*. Hart Publishing, 1999.

[31] C J Jepma, H Jager, and E Kamphuis. *Introduction to International Economics*. Addison–Wesley, 1996.

[32] Robert E Kennedy. Strategy fads and competitive convergence: An empirical test for herd behavior in prime-time television programming. *The Journal of Industrial Economics*, 50(1):57–84, March 2002.

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

[34] Dipjyoti Majumdar. *Essays in Social Choice Theory.* PhD thesis, Indian Statistical Institute, October 2002.

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

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

[37] Paul R Milgrom. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 12(2):380–391, Autumn 1981.

[38] Deborah Minehart and Suzanne Scotchmer. Ex post regret and the decentralized sharing of information. *Games and Economic Behavior*, 27:114–131, 1999.

[39] Peter Murray. *Basic Trial Advocacy.* Aspen Publishers, 1995.

[40] André Orléan. Baysean interactions and collective dynamics of opinion: Herd behavior and mimetic contagion. *Journal of Economic Behavior and Organization*, 28:257–274, 1995.

[41] Nancy Pennington and Reid Hastie. Evidence evaluation in complex decision-making. *Journal of Personality and Social Psychology*, 51(2):242–258, 1986.

[42] 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.

[43] 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.

[44] Judea Perl. *Causality.* Cambridge University Press, March 2000.

[45] Wolfgang Pesendorfer. Design innovation and fashion cycles. *The American Economic Review*, 85(4):771–792, September 1995.

[46] Roy Radner. Rational expectations equilibrium: Generic existence and the information revealed by prices. *Econometrica*, 47(3):655–678, May 1979.

[47] William H Riker and Peter C Ordeshook. A theory of the calculus of voting. *American Political Science Review*, 62(1):25–42, March 1968.

[48] Alan Rosenthal. *The Third House: Lobbyists and Lobbying in the States.* CQ Press, second edition, 2001.

[49] Alexander A Schuessler. *A Logic of Expressive Choice.* Princeton University Press, 2000.

[50] Glenn Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, 1976.

[51] Bernard W Silverman. Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society, Series B (Methodological)*, 43:97–99, 1981.

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

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

[54] Amos Tversky and Daniel Kahneman. Judgement under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, September 1974.

[55] Arthur De Vaney and W David Walls. Bose-Einstein dynamics and adaptive contracting in the motion picture industry. *The Economics Journal*, 47:1493–1514, November 1996.

[56] Bruce A Weinberg, Patricia B Reagan, and Jeffrey J Yankow. Do neighborhoods affect hours worked: Evidence from longitudinal data. Working paper, Ohio State University and Furman University, November 2001.

[57] Ivo Welch. Herding among security analysts. *Journal of Financial Economics*, 58:369–396, 2000.

[58] Ward Whitt. A note on the influence of the sample on the posterior distribution. *Journal of the American Statistical Association*, 74(366):424–426, June 1979.

[59] Arnold Zellner. *An Introduction to Bayesian Inference in Econometrics.* Robert E Krieger Publishing Company, 1971.