Four Puzzles in Information and Politics: Product Bans, Informed Voters, Social Insurance, & Persistent Disagreement

Thesis by

Robin D. Hanson

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

California Institute of Technology
Pasadena, California

1998
(Submitted June 16, 1997)

Acknowledgements

I thank the Sloan Foundation for a Dissertation Fellowship and the New Millennium Program Office of the Jet Propulsion Lab of NASA for financial support.

Thanks to my wife Peggy Jackson, my children Tommy and Andy, and my parents Don and Bonnie Hanson for suffering through an older student's return to low pay and prestige.

Thanks to my primary advisor, John Ledyard, for helping me see the economic theorist's concepts of generality and relevance. And thanks to the rest of my thesis committee, Richard McKelvey for unswerving intellectual honesty, Thomas Palfrey for setting high standards, and Simon Wilkie for his intuition regarding relevant theoretical applications.

I also thank other faculty I have relied on heavily. Thanks to Dave Porter for his experimental intuition, and for getting me to Caltech in the first place. Thanks to Matt Spitzer for our grounded policy discussions, and thanks to Michael Alvarez for help on the job market. And thanks to Kim Border for mathematical advice, especially regarding proving existence in chapter 3.

Thanks also to Jeff Banks, Pierpaolo Battigalli, Colin Camerer, Paolo Ghirardato, David Grether, Matt Jackson, Steven Knack, Susanne Lohmann, and Scott Page for discussions and comments on various parts of this research. And I thank seminar participants at Caltech, MIT, Northwestern U., and U. Rochester for their comments on the research in chapter 2.

Thanks to David Friedman for making economics seem worth the trouble

Thanks to my fellow students, Micah Altman, Oleg Bondarenko, John Duggan, Mark Fey, Gretchen Kaslow, Heidi King, Hye Yeong Lee, Mishe Myagkov, Charles Polk, Jason Savings, Fang Wang, Dean Williamson, and Jin Yu for sharing this journey with me.

Finally, I thank these older but still strong influences on my intellect, Eric Drexler, David Malament, Mark Miller, Riley Newmann, and many discussants on the Extropians email list.

Abstract

In four puzzling areas of information in politics, simple intuition and simple theory seem to conflict, muddling policy choices. This thesis elaborates theory to help resolve these conflicts.

The puzzle of product bans is why regulators don't instead offer the equivalent information, for example through a "would have banned" label. Regulators can want to lie with labels, however, either due to regulatory capture or to correct for market imperfections. Knowing this, consumers discount regulator warnings, and so regulators can prefer bans over the choices of skeptical consumers. But all sides can prefer regulators who are unable to ban products, since then regulator warnings will be taken more seriously.

The puzzle of voter information is why voters are not even more poorly informed; press coverage of politics seems out of proportion to its entertainment value. Voters can, however, want to commit to becoming informed, either by learning about issues or by subscribing to sources, to convince candidates to take favorable positions. Voters can also prefer to be in large groups, and to be ignorant in certain ways. This complicates the evaluation of institutions, like voting pools, which reduce ignorance.

The puzzle of group insurance as a cure for adverse selection is why this should be less a problem for groups than individuals. The usual argument about reduced variance of types for groups doesn't work in separating equilibria; what matters is the range, not variance, of types. Democratic group choice can, however, narrow the group type range by failing to represent part of the electorate. Furthermore, random juries can completely eliminate adverse selection losses.

The puzzle of persistent political disagreement is that for ideal Bayesians with common priors, the mere fact of a factual disagreement is enough of a clue to induce agreement. But what about agents like humans with severe computational limitations? If such agents agree that they are savvy in being aware of these limitations,

then any factual disagreement implies disagreement about their average biases. Yet average bias can in principle be computed without any private information. Thus disagreements seem to be fundamentally about priors or computation, rather than information.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

It has been barely two decades since adequate formal tools have been used in earnest to explore the role of information in social processes. And even then the community of explorers engaged in this quest has been dwarfed by the vast terrain to be covered. Thus a new explorer such as myself has the luxury of a wide range of rich, close-at-hand, at yet hardly examined prospects to investigate. There is a long list of basic, important, "in your face" social phenomena where information seems to be the key, and yet where we have at best only crude unsatisfying stories about what could be going on.

In the four chapters which follow in this thesis, I examine four important and puzzling areas of social information phenomena: paternalistic product and activity bans, voter incentives to become informed, adverse selection regarding collective choices, and the causes of persistent disagreement. In all of these puzzle areas simple intuition seems to be in conflict with simple theory, muddling important policy choices.

To help resolve each conflict between simple intuition and simple theory, I bring more sophisticated theory to bear, to identify plausible but overlooked social processes at work. When possible, I also use this more sophisticated theory to compare welfare across alternative institutions.

All of these puzzle areas are also either centered in or have strong relevance for important political phenomena. Yet they are also all familiar topics in economic and policy analysis. Thus this thesis can be thought of as centered either within formal political theory, within formal analysis of law and policy, or within the economics of information and public institutions.

1.0.1 The Puzzle of Product And Activity Bans

The puzzle examined in chapter 2 is that of paternalistic product and activity bans. At a political level, why do politicians and their regulators ban products, rather than labeling them as bad? Or at a personal level, why do parents forbid their children to engage in certain activities, such as drinking, driving, and sex, rather than simply advising children of possible dangers?

Pure preference divergence, such as regulatory capture, is an unsatisfying explanation of product bans; why don't captured regulators instead seek direct cash transfers? And how could bans be an attempt to hide the fact of transfers when bans are such a visible and easily monitored action? If, on the other hand, regulators simply have better information than consumers, why don't they just label certain products as "would have banned"?

Chapter 2 shows that either a small degree of regulator capture or a small deviation from fully competitive markets gives regulators a small incentive to lie about product quality. But small lies are corrosive, inducing a great deal of consumer skepticism regarding regulator statements. Thus even ideal regulators want to ban products which are bad enough, rather than live with the choices of ignorant consumers. Similarly parents can want to forbid their skeptical children from engaging in harmful activities.

When regulators and parents are forbidden from banning, however, consumers and children take regulator labels and parental warnings more seriously. And a welfare analysis reveals that for a wide class of cases, all parties on average prefer the outcomes when banning is forbidden. In these cases, bans can be seen as a commitment failure. This suggests that we consider the alternative institution of constitutional prohibitions on product bans, analogous to first amendment prohibitions on media bans.

1.0.2 The Puzzle of Informed Voters

The puzzle examined in chapter 3 is that of informed voters.

Many have wondered why independently acting voters in large electorates would have much instrumental reason to vote. After all, such a voter should discount the benefits of voting, but not the costs, by the probability of being pivotal (i.e., that the election is decided by one vote).

By analogy, why would voters acquire political information, such as via the ever-popular political news? Is politics that entertaining or valuable in day-to-day living? Yet actual political choices do not seem to reflect as uninformed an electorate as one might fear. (Though clearly the electorate is much less informed than many would wish.)

The analogy between voting and voter information has limits, however. While voters may find it hard to commit to vote, voters can commit to holding relevant political information, either by just acquiring it or by subscribing to an information source. And candidates who can observe such early efforts should adjust their positions to better favor informed voters. Since this influence is not diluted by the probability of being pivotal, it can give voters strong incentives to become informed.

Chapter 3 also considers voter preferences for being in large vs. small voter groups, where candidate positions cannot distinguish group members. Voters can prefer to be in large groups because scale economies in information production can override free-riding considerations.

Finally, it is shown that a certain type of voter ignorance, which prefers negative to positive news, can benefit voters both individually and collectively, by eliminating wasteful instabilities in candidate positions. This complicates consideration of the alternative institution of voluntarily-formed voting pools, where all the votes of pool members are given to one pool member at random. Such pools can induce better informed voters, but the value of ignorance can make voters want to commit not to join such pools.

1.0.3 The Puzzle of Group Insurance

Chapter 4 considers the puzzle of group insurance and other attempts to solve adverse selection problems via collective choice.

While there are obvious tax and overhead-reduction advantages of employer-based health insurance, it is often argued that a key function of group insurance is to avoid adverse selection problems in individual insurance. Many government-imposed restrictions, such as limits on hours of work, are explained similarly; by making a common choice we are said to avoid inefficient personal signaling.

Adverse selection happens in separating equilibria of signaling games. In such games, individuals vary in their innate "type," and good types attempt to distinguish themselves from "bad" types via their actions. A group making a collective choice also has a "type," however, which is the set of its member types. So why don't bad companies buy more health insurance than good companies?

The usual argument one hears is that a group, by averaging over its members, has a lower variance of possible risk types than each member. However, given the usual equilibrium refinements which select full separation, so that each type is distinguished from all the rest, equilibria and their inefficiencies depend only on the *support*, not the variance, of a distribution of types. Thus the usual argument is highly suspect.

Chapter 4 suggests that the key is not averaging to reduce variance, but limiting participation to narrow the support. For example, because a majority vote can fail to represent up to half of the electorate, this narrows the range of group types which can be inferred from democratic choices. And decisions by a random jury, who fail to represent most of a large group, can in the limit avoid all adverse selection losses from independent individual risks. This suggests an advantage of judge-made laws aimed at excessive signaling, such as liquidated-damages rules.

1.0.4 The Puzzle of Persistent Disagreement

Finally, chapter 5 considers the puzzle of persistent disagreement, in politics or anywhere else.

While honest differences of opinion seem ubiquitous in the world, simple theory suggests the remarkable conclusion that rational agents simply cannot agree to disagree. Consider two ideal Bayesian jurors of the O.J. (Simpson) trial. They start (at birth say) with identical beliefs, receive different private information before and during the trial, learn something about each other's beliefs during deliberation, and finally estimate the chance O.J. did it. If during deliberation these jurors reach a highly common belief about which of them estimates a higher chance that O.J. did it, this turns out to be enough information to allow these jurors to come to nearly the same estimates regarding O.J.'s chances.

Most researchers who are dissatisfied with explaining apparent disagreement as due to different initial (i.e., prior) beliefs, or as due to posturing by people who really agree, have looked to bounded rationality as the explanation. And it does seem that the calculations required of an ideal Bayesian are typically far beyond the ability of mere mortals.

Existing research on bounded rationality, however, has either assumed very specific computational strategies, or has stayed general at the cost of allowing nearly as much computation as an ideal Bayesian requires. For example, some models assume that agents know anything a Turing machine can compute in any finite time, and other models assume that agents can compute exact expected values over vast state-dependent sets of possible states, sets which satisfy various strong axioms.

In contrast, chapter 5 allows agents to make arbitrary state-dependent computational errors. These agents are constrained only to be *savvy* in the sense of being aware of certain easy-to-compute implications of the fact that agents make such errors. Even this minimal degree of rationality, however, implies that agents with common priors who agree to disagree about O.J.'s chances must agree to disagree about each of their average bias when making such estimates.

Since average bias could in principle be computed independently of private information, this situation is a computational disagreement, similar to a situation where one agent always estimates π to be 3.14, another always estimates π to be 22/7, and both are fully aware of the others' alternative method. It seems that disagreements are about computation or priors, not information.

Chapter 2 Why Regulators Ban Products

2.1 Introduction

2.1.1 The Puzzle of Product Bans

Governments currently prohibit the sale of a wide variety of products. Often such prohibitions take the form of mandating features in allowed products.

For example, cars, planes, and many household appliances may not be sold without a variety of safety and other standard features (such as seat belts). Building codes constrain which kinds of houses and other construction may be sold. Financial regulation, including laws against gambling and usury, prohibit the sale of a wide variety of financial instruments, and require a variety of information be disclosed.

Food must meet various standards before it can be sold, and drugs usually cannot be marketed without regulatory approval. Professional licensing laws prohibit unauthorized people from selling various medical, legal, educational, and other services. Finally, laws in many places limit the sale of sex, of pornography and erotica, and even of literature concerning disapproved medical, religious, and political ideas.

2.1.2 Previous Explanations

A wide variety of explanations have been offered to account for this regulatory behavior. One class of explanations involves non-competitive industries. Since a monopolist or other producer for a non-competitive market may prefer to offer an inefficient menu of product qualities [MR78], there can be situations in which regulators[1] can improve efficiency through restrictions on allowed product quality [Ron91]. Self-regulating

[1] In this paper, a "regulator" is any authority authorized by a government to ban or label products.

monopolists can gain similar benefits [GJ95].

Product bans may also mitigate excessive signaling, by preventing the product variation used to signal. Among the products that may be banned to prevent consumer risk signaling are various insurance products [RS76], product liability waivers [Ord79], and liquidated-damages contracts [AH90]. Wealth-signaling status-signals can be banned during national emergencies, and professional licensing may limit inefficient signaling via human capital investments [Sha86].

However, while non-competitive industries and inefficient signaling may be popular explanations with economic theorists, such concepts are rarely mentioned in political discussions where these policies are presumably decided. While this is not an overwhelming objection, it does deserve consideration.

A concept that political discussions more often refer to is that of use-externalities; advocates suggest many externalities which their favored bans would mitigate. For example, building codes may lower the risk of fires spreading to neighboring properties. Required health care product features may reduce the spread of infectious disease. Required product safety features may lower the risk that injured people will use public health care.

Usury laws and bans on liquidated damages contracts may also limit the risk that poor people will require public assistance [Pos95]. And allegations of crime and public-assistance mediated externalities have been pivotal in securing public support for drug prohibitions [JGP+85, Mil91], including alcohol prohibition [Bar05, Isa65].

Bans on sales of sex, pornography, body parts, and children are said to mitigate the externality of "commodification" which markets are said to create in the presense of "incommensurable" values [Rad96]. Similarly, bans on prohibited literatures are said to mitigate the externalities of contagious bad ideas. Finally, meddling preferences, where voters have specific preferences over other people's consumption patterns, are a type of externality which can (perhaps too easily) explain a wide variety of regulatory intervention.

Externalities are surely part of the story for many product bans. But there are difficulties in explaining most product bans this way. For example, it is not clear why legal liability rules or special product taxes would not usually better deal with such externalities [Bur93]. (Yes a product ban can be thought of as a a very high tax, but why are there are not more intermediate level taxes?)

Also, product bans do not seem to be very sensitive to known variations in the magnitude of externalities. For example, neither drug regulation or health professional licensing seem to distinguish contagious from non-contagious illnesses. And building codes do not seem to distinguish fire-promoting risks from risks, such as structural defects, with mostly local consequences. It is also hard to understand what substantial externality, signaling, or non-competitive behavior could be behind required airline safety features, especially on flights which are mainly over the open ocean.

Regulatory capture by special interests is another relatively general explanation of most product bans, which is mentioned often, though not especially frequently, in political debates. While this explanation has much intuitive appeal, it is hard to understand why rational voters would approve politicians who back such bans, if regulatory capture were the main story. Direct cash transfers seem to be a more efficient form of wealth redistribution, so the only obvious reason for using inefficient bans would be to disguise the transfer. But product bans could only disguise such transfers if such bans had other substantial accepted purpose in the minds of voters [CM95]. Thus while some degree of regulatory capture is surely present, it is hard to understand how it can be the main story.

Early empirical studies [Sti71, Mau74] did suggest that professional licensing was best explained as due to regulators captured by professionals in search of higher incomes. The weight of recent evidence, however, seems to support studies which find this to be a relatively small effect when compared to "consumer demand" and "public interest" type regulation [Lef78, LOR95, LM90, Jen92].

Public interest theories of product bans are also relatively general. Such theories focus on regulators who have special information about product quality, and who use this information to protect consumers from buying bad products. Such public interest theories seem to dominate the political discussion of product bans; critics may question regulator's judgement regarding a particular quality, but they rarely question the basic concept of banning very low quality products. For example, the Florida statute requiring medical licensing begins by explaining that "it is difficult for the public to make an informed choice when selecting a physician" [Fei85].

Such public interest justifications also have very high levels of public support. For example, a 1976 Harris poll found that 85 and 83 percent of the public favors federal regulation of product safety and quality standards, respectively [LS79].

When regulators have special information which they cannot otherwise communicate to consumers, and when consumers cannot or will not get equivalent information from other sources, product bans *can* improve consumer welfare. For example, when regulators can observe product quality, they can help consumers by imposing a minimum required quality level [Lel79]. And theories of special regulator information can help explain why old and familiar products, such as rock climbing, smoking, alcohol, and pornography, are banned less often than comparable and arguably less harmful new products.

Many authors have noted, however, that in such situations consumers could do no worse, and often better, if, rather than banning bad products, regulators instead communicated the same information by certifying good products [Hig95, Wei80]. For example, Leffler [Lef78] notes that

> Under a costly information argument for intervention, certification is the preferred response. Certification provides all the information of licensure while offering a wider choice set.

Confirming this intuition, many papers which suggest an advantage for product bans do not directly compare bans to certification of the ban information. Gale [Gal96] takes them to be equivalent, for example, and Shapiro [Sha86] compares bans with certification of more than the ban information. Shaked and Sutton [SS81], who do directly compare bans and labels in comparing professional licensing and certification, find that permitting entry of rival para-professionals is welfare improving.

A counter argument to certification is that it is costly to display labels, and costly for a customer to be constantly "checking for the certification every time he buys an unfamiliar product" [Kel81]. With perfect enforcement of bans, in contrast, it takes no effort to avoid banned products. In response, many authors have suggested that labeling costs could be borne by the unapproved products, for example by requiring a large red "Not FDA approved" warning label on unapproved drugs [Gie85, PS82, Wei80].

If watching for this label is still deemed too costly a burden for consumers, then "a case could probably be made for allowing establishment of certain stores, with warnings prominently posted, that sell only products that do not meet regulatory standards" [Kel81]. One might even require that consumers pass a test, something like a driver's test, before they are allowed to buy from such a store. The fact that no such exceptions are allowed to product bans is somewhat of a puzzle from the simple public interest regulation perspective.

A simple modified theory which can explain this behavior posits systematic biases in consumer beliefs. In this case regulators could complain that bans are required because consumers would simply not believe regulators who said that the product is bad. For example, various limits to contract have been explained as due to "limits of cognition" such as framing effects and availability and representativeness biases [Eis95].

Attention has focused in particular on consumer biases regarding low probability events [Spe77]. Ackerlof and Dickens [AD82] suggest that required safety features may help consumers if cognitive dissonance makes them prefer to believe products are safe rather than fear for their lives. And Viscusi [VVH95] instructs risk regulators to keep in mind that "individuals tend to overestimate the risks associated with lower-probability events ... [and to] underestimate the risks associated with higher-risk events."

Unless people are less biased as voters than they are as consumers, however, it is hard to understand why they would favor politicians who promise future product bans. The same biases that would lead voters to buy too much of a bad product should lead them to expect to benefit from the option to buy such products. Hence we would expect them to favor politicians who promise to allow them that choice, at least if voters base their electoral choices primarily on *prospective* evaluations of the future consequences of candidate policy positions.

In many electoral models, however, "the chosen candidate ... selects an action ..., where this action is unobserved by the voter, and (stochastically) determines the voter's reward for that period." Because of this "the voter employs a simple *retrospective voting rule*: retain the current incumbent as long as rewards remain above a certain level" [BS93]. Biased and retrospective-voting consumers could induce regulators to try to correct for their biases, and "take as societies objective the promotion of societal welfare based on the true risk levels, not the risk levels as they may be perceived by society more generally" [VVH95], as many regulation theorists suggest.

While there is no doubt that consumers are not always exactly rational, however, it is also far from clear that consumers are systematically biased in just the ways required to explain most product bans. That is, would consumers be biased even after taking a class informing them of their bias, after passing a special test, and then being limited to buying such products in special stores? And if strong ex-ante public support for product bans we observe [Kel81] means that consumers know they are often biased, why wouldn't they then believe regulator warning labels? Or why wouldn't they voluntarily agree ahead of time to a personal product ban, where they agree to be punished if they purchase some non-certified product?

2.1.3 Cheap Talk As An Explanation

This paper suggests that we might better understand the phenomena of product bans by realizing that, even with fully rational consumers, a small degree of either regulatory capture or non-competitive markets can trigger consumer behavior that looks to regulators a lot like an irrational unresponsiveness to regulator labels. This is because if retrospective voting explains voter behavior, then regulator product labeling is basically a cheap-talk signaling game, where a small divergence in preferences between regulator and consumers can induce large losses in information transfer [CS82]. And small regulatory capture or non-competitive market effects can induce the required small divergence in preferences.

When product quality is one-dimensional, a regulator-labeling cheap-talk equilibrium consists of a set of quality intervals. A regulator who knows the true product quality can only communicate which interval this quality lies within, and cannot make finer within-interval distinctions. In particular, there is a lowest interval, where consumers will buy an amount of the product corresponding to the average quality in this lowest interval. When the true quality happens to be near its lowest possible value, the regulator can prefer that no one buy the product, rather than deal with the consequences of future voters unhappy with their product experience.

If regulators can actually ban the product, in addition to announcing its quality, then the signaling game is changed; the lowest quality interval now consists of banned products, and the boundaries of the higher certification intervals move. In equilibrium not banning a product will be taken by consumers as an endorsement of product quality, which encourages regulators to ban even more products. And consumers who act on this perceived endorsement can as voters be understandably upset at a regulator who failed to ban a product of very low quality.

Like the simple public interest theory which it modifies, this cheap-talk explanation of product bans is relatively general[2], and does not depend on special industry structures in the way that the non-competitive market and excessive signaling explanations do. This theory also works with fully rational consumers and regulators, and with costless labels.

This theory fits comfortably with recent empirical work which prefers a large public interest component and a small regulatory capture component to product ban behavior, since the theory can explain large rates of product banning from small levels of regulator bias. With the public interest theory, this theory explains the apparent regulatory focus on new harms.

Yet this theory also explains why consumers ignore fine label information, and

[2]The phenomena of product bans may of course be best explained by many dissimilar theories, each of which covers some narrow range of products. But it is surely worth looking for more general explanations.

why regulators do not allow exceptions for consumers who pass special tests in order to shop at special unapproved product stores. This theory also makes it easier to understand why voters re-elect regulators who ban, and it explains why regulators who are biased *in favor* of some product category may still ban some products in this category.

Finally, this explanation can take most political discussion of product bans at face value, instead of suggesting that such discussion is mainly a smoke screen obscuring the real motives. Given that parties already have asymmetric information and that the regulatory authority to ban is not in dispute, we should expect to see, and largely do see, debate on bans which centers on the actual quality of particular products.

2.1.4 Welfare Comparisons

Voters may be happy with the behavior of a particular regulator relative to an equilibrium of a game where regulators are authorized to ban products. This does not mean, however, that voters would not be even happier in an equilibrium of a game where regulators are not authorized to ban products. Since these games have different equilibria, it is sensible to make welfare comparisons between them.

After introducing a simple example of Cournot competition in two competing products with linear supply and demand, and after examining the divergent preferences which regulatory capture or non-competitive markets can induce in this example, this paper will focus on making welfare comparisons. Assuming a uniform distribution over product quality, assuming full competition, and assuming a regulator whose bias for one product over another is constant over the range of product quality, exact solutions and welfare values will be given.

Given these assumptions, it turns out that for any given level of regulator bias, the best (i.e., highest total welfare) equilibrium is always in the game where bans are not allowed. And if bans are not allowed, all groups ex-ante prefer a smaller magnitude bias.

The different groups have different ex-ante preferences, however, over bans and regulator bias given the game where bans are allowed. Consumers and producers of the possibly-banned product always ex-ante prefer no bans and a small bias, as do regulators biased in favor of these producers. But producers of the competing product ex-ante prefer that bans be allowed, and given bans they prefer that regulators be biased as far as possible in their favor. Regulators also prefer bans be allowed when they are biased in favor of these competing producers, though they still prefer to have a smaller magnitude bias.

Thus, in this example, eliminating the possibility of product bans aligns the interests of all parties ex-ante, so that they should all cooperate to minimize the magnitude of regulator bias.

The above conclusions were for a uniform distribution over product quality. For other distributions it will be shown that the best equilibrium can be in the game where bans are allowed. This does not contradict the intuition that consumers should prefer the information in a ban to the ban itself, since the lowest label in the no-ban game equilibrium communicates different information than a ban. This is because banning a product is a more severe action than assigning it the lowest quality label in the no banning game, so there are quality levels which a regulator is not willing to actually ban, but is willing to label as in the lowest quality interval without bans.

Finally, a general sufficient condition will be given regarding when the maximum welfare equilibrium is always in the game where bans are not allowed. We start with general preferences over one-dimensional types (e.g., quality levels) and actions (e.g., amount purchased) and consider Crawford and Sobel's [CS82] cheap-talk signaling game between a sender (e.g., a regulator) and a receiver (e.g., a market containing consumers), modified to allow the sender a single extreme "forced" or "ban" action. We assume, with Crawford and Sobel [CS82], concavity and sorting in preferences, convex signal support, that at no type do the two player's preferences coincide, and that interval boundaries all move in the same direction.

Given these assumptions, we find that allowing a ban is never better for either the sender, the receiver, or anyone with intermediate preferences, when the sender's preferences are biased *away* from the forced action. At least this is true if the ban outcome is between these two no-ban-possible outcomes: when the sender declares that the signal (e.g., quality) was the worst possible, and when the receiver knows for sure that the signal is the worst possible.

This result is counter to the plausible intuition that while it might be dangerous to allow bans by a regulator who is biased against a product, it couldn't hurt to allow bans if the regulator is biased in favor of the product. Actually, bans can only help ex-ante with regulators who are biased against the product.

When consumers ex-ante prefer the no-ban game to the ban game, banning can be thought of as a commitment failure. Without a commitment not to ban, even an ideal unbiased regulator will often want to ban in equilibrium, knowing that a failure to ban would be interpreted by consumers as a product endorsement. But consumers who know that regulators are constitutionally prevented from banning will interpret regulator labels differently, to their ex-ante benefit.

It was hard for this author to find combinations of functional forms and parameter values for which consumers prefer the ban game, and relatively easy to find functional forms where for all parameter values consumers prefer the no-ban game. This suggests, though only weakly, that preferring the no-ban game is the usual case. If so, perhaps consumers would benefit overall from a broad constitutional prohibition against product bans, such as currently applies to print media in the United States.

2.1.5 General Paternalism

The generalized "cheap-talk" model of Crawford and Sobel [CS82] mentioned above, with one-dimensional information held by a sender, and a matching one-dimensional choice made by a receiver, applies to a wide variety of contexts beyond that of product quality regulation. And the augmented game with a single extreme forced action available to the sender can be considered a general model of "extreme-act paternalism."

For example, a parent may face the choice between recommending to their child how far to go on a date, or simply prohibiting dating before a certain age. Similarly, a parent may either face the choice between recommending how much alcohol or other drugs to consume, or simply prohibiting any such consumption. Parents may also advise their children how carefully to drive, or simply prohibit driving. The above analysis suggests that such "paternalistic" extreme choices will often be made, even with fully rational children, if parents suspect that their children believe that those parents have slightly different preferences.

On a technical note, this paternalism game is somewhat unusual in mixing cheap talk and expensive signals. This paper has also made a modest technical contribution to the analysis of the original cheap-talk game of Crawford and Sobel [CS82]. They called the assumption that interval boundaries all move in the same direction *monotonicity*, and presented a sufficient condition for monotonicity involving two inequalities on preferences. A different and arguably more intuitive sufficient condition is presented here, involving only one preference inequality.

These general results on an applied model are somewhat unusual in the cheap-talk literature, most of which has been concerned either with models with specific functional forms, or with very general analysis, mostly regarding equilibrium refinements and existence.

2.2 Linear Supply and Demand

2.2.1 The Market Model

Let us first consider product bans in a case of not fully competitive markets with linear supply and demand. Specifically, let us consider two products, 1 and 2, with linear supply, i.e., industry marginal cost, given by

$$MC_1 = a_1 + b_1 Q_1$$

$$MC_2 = a_2 + b_2 Q_2$$

and linear demand, i.e., consumer marginal value, given by

$$MV_1 = c_1 - d_1 Q_1 - e_1 Q_2$$

$$MV_2 = c_2 - d_2 Q_2 - e_2 Q_1.$$

We assume non-negative slope coefficients $b_1, b_2, d_1, d_2, e_1, e_2$, and assume $d_1 d_2 > e_1 e_2$ to ensure the familiar sign of slopes in expressions like $Q_1 = A_1 - \alpha_1 P_1 + \beta_1 P_2$.

We also assume Cournot competition. Each of n_1 identical producers of product 1, facing an individual marginal cost of $MC_i = a_1 + n_1 b_1 Q_{1i}$, simultaneously chooses its quantity Q_{1i}. At the same time, n_2 identical producers of product 2 similarly choose Q_{2i}. Then consumers drive prices to their marginal value at the quantities produced, as in $P_1 = MV_1(Q_1, Q_2)$ with $Q_1 = \sum_{i=1}^{n_1} Q_{1i}$, and similarly for P_2 and Q_2.

Defining marginal welfare loss to be $L_1 = MV_1 - MC_1$, the Cournot equilibrium satisfies $L_1 = (d_1/n_1)Q_1$. Thus welfare loss is positive, but decreases as more firms compete (n_1 larger), as demand becomes more inelastic (d_1 smaller), and as total demand decreases (Q_1 smaller). (This last dependence on quantity may be an artifact of holding the number of firms fixed as quantity decreases, rather than modeling the entry and exist of firms from each industry.)

Defining

$$f_1 = \frac{1}{e_1}(d_1 \frac{n_1 + 1}{n_1} + b_1), \qquad g_1 = \frac{c_1 - a_1}{e_1}$$

(and L_2, f_2, and g_2 similarly, with the labels 1 and 2 switched), we can write the equilibrium quantities as

$$Q_1 = \frac{f_1 g_1 - g_2}{f_1 f_2 - 1}, \qquad Q_2 = \frac{f_2 g_2 - g_1}{f_1 f_2 - 1}.$$

Consider now the consequences of varying the quality q of product 1, which varies the parameter $c_1 = q + \tilde{c}_1$ while holding constant n_1, n_2 and all the other supply and demand parameters. The quantities $Q_1(q), Q_2(q)$ then vary linearly with g_1, and hence linearly with q, and they vary in opposite directions. At $Q_1 = 0$, $Q_2 = \bar{Q}_2 = g_2/f_1$ and at $Q_2 = 0$, $Q_1 = \bar{Q}_1 = g_2$.

For example, in the symmetric case where $a_1 = a_2 = 0$, $b_1 = b_2 = 0.5$, $c_2 = 1$, $d_1 = d_2 = 1.2$, $e_1 = e_2 = 1$, and $n_1 = n_2 = 100$, as Q varies over $[0, 1]$, c_1 varies over $[0.584, 1.712]$, Q_2 ranges over $[0.584, 0]$, P_1 ranges over $[0, 0.512]$, and P_2 ranges over $[0.299, 0]$. All these variables are linear in c_1, and hence are linear in each other.

2.2.2 Possible Biases

Risk-neutral consumers who are symmetrically uncertain about quality q will act according to their common expectation of quality $\hat{q} = \mathrm{E}[q]$, inducing a market quantity $\hat{Q} = Q_1(\hat{q})$, instead of the quantity $Q = Q_1(q)$ which would be induced by perfect information. Incentives for any agent to deceive consumers about product quality can then be described by the way in which that agent's payoff changes with consumer expectation $\hat{q}$, given a fixed true quality q.

Let π_1, π_2 be the producer profits for the two industries, let π_C be consumer surplus, let $W_0 = \pi_C + \pi_1 + \pi_2$ be total welfare given equal welfare weights, and for any X let $X' = dX/d\hat{q}$ evaluated at $\hat{q}$. Then we can write

$$W_0' = (q - \hat{q} + L_1)Q_1' + L_2 Q_2',$$

which combines the consequences of quality misperceptions with the welfare losses of non-competitive markets.

In the competitive limit $(n_1, n_2 \to \infty)$, W_0 is maximized at $q = \hat{q}$, so someone seeking to maximize total welfare would have no incentive to deceive consumers about quality. For a finite number of firms, however, this expression W_0' can be non-zero even when $q = \hat{q}$. This is because quantity increases due to misperceptions of product quality can be used to compensate for quantity reductions due to non-competitive markets.

Further incentives to deceive consumers about quality can arise from not giving equal welfare weights to the three groups, consumers and two producer industries, as in $W_\gamma = W_0 + \gamma_1\pi_1 + \gamma_2\pi_2$, with γ_1, γ_2 being the weight *deviations* of the two producer industries.

Before describing these more general incentives in detail, let us note that since W_0' and all π_j' are linear in both q and $\hat{q}$ (e.g., $\pi_1' = L_1 Q_1' + \hat{Q}_1 M V_1'$), all preferences are quadratic in $\hat{q}$ with a convexity W_γ'' independent of q. For agents with such quadratic preferences, we need only consider their ideal points. For such quadratic agents, we can also generalize our information structure; q and Q can now refer to the expected value of quality and quantity using the regulator's superior but not necessarily full information.

Thus we need only consider such an agent's *bias*, defined as the difference $\beta(Q) = \hat{Q}^* - Q$, where ideal point $\hat{Q}^*$ solves $W_\gamma' = 0$ for a given Q. Bias is the difference between the true quantity $Q = Q_1(q)$ and the ideal quantity $\hat{Q} = Q_1(\hat{q})$ one would like consumers to believe. A bit of algebra then reveals that

$$\beta(Q) = \frac{(H_1 + H_2)Q - H_2\bar{Q}_1}{H_0 + H_1 + H_2}$$

where $H_0 = e_1 f_1(f_1 f_2 - 1)$ and

$$H_1 = \gamma_1 H_0 + \gamma_1 f_1(e_2 - d_1 f_1) + (1 + \gamma_1) f_1^2 \frac{d_1}{n_1}$$

$$H_2 = \gamma_2 f_1(e_2 f_1 - d_2) + (1 + \gamma_2)\frac{d_2}{n_2}.$$

To ensure the concavity of W_γ, we assume $H_0 > H_1 + H_2$.

Bias vanishes, i.e., $\beta(Q^0) = 0$, at $Q^0 = \bar{Q}_1 H_2/(H_1 + H_2)$. The sign of the bias changes at this boundary, and for $Q > Q^0$, $\text{sign}(\beta) = \text{sign}(H_1 + H_2)$. Thus when Q^0 lies in $[0, \bar{Q}_1]$, $H_1 + H_2 > 0$ gives an *outward* bias, toward the extremes, while $H_1 + H_2 < 0$ gives an *inward* bias, toward the boundary Q^0. When $H_1 + H_2 = 0$, the bias is *constant*.

In the special case of zero weight deviations, $\gamma_1 = \gamma_2 = 0$, the bias is outward with $Q^0 = n_1 d_2 \bar{Q}_1/(n_1 d_2 + n_2 d_1 f_1^2)$. For example, as quality q rises so that Q_2 goes to zero, the welfare loss from a not fully competitive market in product 2 also falls to zero, while the welfare loss from a not fully competitive market in product 1 gets larger. In this case, one might prefer that consumers overestimate the quality of product 1 to compensate for the producer's strategic reductions in the quantity of that product.

In the special case of fully competitive markets ($n_1, n_2 \to \infty$), we can write $H_1 = \gamma_1 \tilde{H}_1$ and $H_2 = \gamma_2 \tilde{H}_2$, with $\tilde{H}_1, \tilde{H}_2$ independent of γ_1, γ_2. Thus for $\gamma_2 = 0$ and $\gamma_1 > 0$, $Q^0 = 0$ and for positive Q, bias β is positive for $\tilde{H}_1$ positive. That is, someone who gives extra weight only to producers of product 1 prefers consumers to overestimate the quality of that product. Similarly, for H_2 positive someone who gives extra weight only to producers of product 2 prefers consumers to underestimate the quality of product 1.

For example, in the specific symmetric example described earlier, zero weight deviations $(\gamma_1, \gamma_2) = (0, 0)$ imply an outward bias that ranges over $[-0.00368, 0.01079]$ as Q ranges over $[0, 1]$, with a zero bias at $Q^0 = 0.254$. That is, with one hundred competing firms for each of two symmetric products, bias is a bit less that one part in a hundred, and toward overestimating quality over most of the quality range. Exactly zero bias can come from symmetric negative deviations for both producers, specifically $(\gamma_1, \gamma_2) = (-0.0229, -0.0229)$. A constant negative bias of -0.01 comes from deviations $(\gamma_1, \gamma_2) = (-0.0444, 0.0402)$, while a constant positive bias of 0.01 comes from deviations $(\gamma_1, \gamma_2) = (-0.00138, -0.0860)$.

2.2.3 Modeling Regulators

Let us assume that consumers are symmetrically but not fully informed about the quality of some product, and that some regulator has obtained further information about this quality. Such a regulator will be empowered to make a pre-purchase announcement to consumers regarding product quality, such as via requiring a product label visible at the point of purchase, and may also be empowered to simply ban the product, ensuring the $Q = 0$ outcome.

If consumers can be divided into distinct groups with differing information or preferences regarding product quality, and if our regulator can selectively target these different groups with different labels and bans, we will consider these to be two different products, to be modeled separately.

We do not explicitly model a regulator who is also empowered to require that consumers purchase the product. We do this because few products are directly required in this manner; most product requirements are actually implemented via product bans. For example, those who buy cars must also buy seat belts, but people are not required to buy cars. Instead, they are forbidden from buying the product of cars without seat belts. This makes sense because there are very few products which regulators know that *all* consumers would buy, were it not for quality misperceptions.

Similarly, we do not allow our regulators to signal their information by punishing themselves, i.e, by "burning money" [ASB95], either directly or indirectly by imposing costs on consumers or producers.

Finally, we assume that regulators just care about the consequences of their actions for consumer and producer welfare, and not about the act itself. Thus in the above example we assume our regulator seeks to maximize W_γ for some values of γ_1, γ_2. Having a regulator maximize a linear welfare function such as this is a standard assumption in the economics of regulation literature [LT93]. Note that this assumption ignores the internal structure of the regulatory process, such as that examined by Hopenhayn and Lohmann [HL96]. We are implicitly assuming something like retrospective voting; come election day voters don't remember much about the specific actions taken by regulators, but they can estimate how happy they are with recent consequences in some regulatory area.

2.2.4 Alternative Theories of Bans

Before considering our cheap-talk model for this quadratic preference case, let us first consider some alternative theories of product bans.

First, consider the case of fully competitive markets ($n_1, n_2 \to \infty$) where consumers know that regulators have no private quality information. Here, product bans must be due purely to "captured" regulators, who are described by non-zero weight deviations γ_1, γ_2. Such a regulator will ban the product if $\beta(Q) < -Q/2$. For a constant bias, the fraction of products banned is linear in the bias, but the product is never banned if the regulator has a bias in its favor.

Second, consider a case of irrational consumers and completely uncaptured regulators. Here information about Q is symmetric between regulators and consumers, but regulators know that consumers are irrationally systematically biased, acting on $\hat{Q} = \alpha Q + \eta$. In this case regulators will ban the product when $Q < \eta/(2-\alpha)$. Here the fraction of products banned is roughly linear in consumer bias η, but product 1 is never banned if consumers are biased against this product (at least for $\alpha < 2$).

Third, consider a prohibitively-costly-labels explanation of bans. Assume that Q is distributed uniformly, that the regulator is unbiased, and that the regulator can only communicate with consumers by banning the product (labels will not be read). In this case the regulator bans the product if $Q < \frac{1}{3}$.

2.2.5 Cheap Talk With Constant Bias and Uniform Distribution

Consider now the cheap-talk explanation of product bans. There are effectively two actors in the cheap-talk model, a regulator and a market. The preferences of both these actors are quadratic over the regulator's private quality signal Q and the market's estimate of this quality $\hat{Q}$. The market effectively maximizes $-(\hat{Q} - Q)^2$, and the regulator maximizes W_γ, which is equivalent to maximizing $-(\hat{Q} - Q + \beta(Q))^2$.

The cheap-talk game proceeds as follows. The regulator sees a private quality signal Q, and then either bans the product (if given the authority) or announces a recommended quantity $\tilde{Q}$. Consumers then estimate product quality, which influences their product demand, which then determines the amount $\hat{Q}$ of the product actually purchased.

Relatively general results regarding such cheap talk games with bans are given in section 2.3, titled "A General Banning Game." Here we look in more detail at some specific closed form expressions for equilibria of this linear supply and demand game. To ease our task of finding closed form expressions, we will in this section assume both that the regulator quality signal Q is distributed uniformly on $[0, 1]$, and that the regulator has a constant bias $\beta(Q) = \beta$. (By assuming constant bias, we are for ease of analysis focusing on a particular one-dimensional subspace of the two-dimensional space of linear bias functions possible in this linear model.)

For this game, Bayes-Nash equilibria can be described by a simple partition of $[0, 1]$, that is, a set of n intervals $[Q_i, Q_{i+1}]$ such that $Q_0 = 0$, $Q_n = 1$, where a regulator who observes a $Q \in [Q_i, Q_{i+1}]$ can only communicate the fact that $Q \in [Q_i, Q_{i+1}]$. Consumers would not believe any more specific claims about Q.

If $Q \in [0, Q_1]$, the lowest quality interval, a regulator who is allowed to will ban the product, forcing $\hat{Q}_0 = 0$. Given a uniform distribution over Q, consumers who are told $Q \in [Q_i, Q_{i+1}]$ and have a choice will choose to buy an average amount $\hat{Q}_i = (Q_i + Q_{i+1})/2$. Finally, the Q_i are the points where the regulator is indifferent between the outcomes $\hat{Q}_{i-1}$ and $\hat{Q}_i$. Given the quadratic regulator preferences, Q_i solves

$$(Q_i + \beta) - \hat{Q}_{i-1} = \hat{Q}_i - (Q_i + \beta),$$

These equations can be solved to give closed form solutions[3], shown in Table 2.1. With no ban, there can be an equilibrium with n intervals for any integer $n \geq 1$ such that $1 \geq 2|\beta|n(n-1)$. For bans allowed with positive (pro product 1) regulator bias $(\beta > 0)$ the same expression holds, but for bans allowed with negative regulator bias $(\beta < 0)$ the condition is instead $1 \geq 2|\beta|(n-1)^2$. Thus for a negative bias, but not for a positive bias, allowing bans can increase the equilibrium "size," i.e., the number of intervals in the equilibrium.

The solutions with the maximal number of intervals n for any given bias level are graphed in Figure 2.1, for $n \in [1, 7]$. (To see the full solutions for lessor values of n, project the lines shown in Figure 2.1 toward the zero bias line.) Table 2.2 gives

[3]Closed form solutions for the case where bans are not allowed were given in Crawford and Sobel [CS82].

no ban	$\beta > 0$	Q_1	$=$	$\frac{1-2\beta n(n-1)}{n}$	Q_i	$=$	$iQ_1 + 2\beta i(i-1)$
ban	$\beta > 0$	Q_1	$=$	$\frac{1-2\beta n(n-1)}{2n-1}$	Q_i	$=$	$(2i-1)Q_1 + 2\beta i(i-1)$
no ban	$\beta < 0$	$1 - Q_{n-1}$	$=$	$\frac{1+2\beta n(n-1)}{n}$	$1 - Q_{n-i}$	$=$	$(n-i)(1-Q_{n-1})$
ban	$\beta < 0$	$1 - Q_{n-1}$	$=$	$\frac{1+2\beta(n-1)^2}{n-1/2}$			$-2\beta(n-i)(n-i+1)$

Table 2.1: Linear Cheap Talk Equilibria Formula

	Bias	Welfare	n max	Boundaries Q_i for n max
no ban	-.5	-83.3	1	0, 1
no ban	-.05	-15.9	3	0, .53, .87, 1
no ban	-.005	-1.66	10	0, .19, .36, .51, .64, .75, .84, .91, .96, .99, 1
no ban	.005	-1.66	10	0, .01, .04, .09, .16, .25, .36, .49, .64, .81, 1
no ban	.05	-15.9	3	0, .13, .47, 1
no ban	.5	-83.3	1	0, 1
ban	-.5	-333.3	1	0, 1 (all banned)
ban	-.05	-17.9	4	0, .31, .74, .97, 1
ban	-.005	-1.68	10	0, .1, .28, .44, .58, .7, .8, .88, .94, .98, 1
ban	.005	-1.68	10	0, .005, .04, .09, .16, .25, .356, .49, .64, .81, 1
ban	.05	-18.7	3	0, .08, .44, 1
ban	.5	- 83.3	1	0, 1 (nothing banned)

Table 2.2: Linear Cheap Talk Equilibria Examples

numerical values for some specific bias values.

Note that with bans allowed, product bans occur for regulators with both positive and negative bias, and given negative bias a large fraction of products are banned even with a very weak bias. For example, 10% of products are banned at a bias of $-.5\%$, and this banning fraction goes roughly as the square root of bias, at least for negative bias. In contrast, in the alternative models described above of pure regulator capture or consumer irrationality, the fraction of banned products is roughly linear in the bias, and hence smaller, and bans only happen for one sign of the relevant bias parameter.

If we assume a full competitive market, then for the class of games we have considered in this section, with quadratic preferences and a uniform distribution over best quantity Q, it turns out that allowing bans never increases expected total welfare W_0.

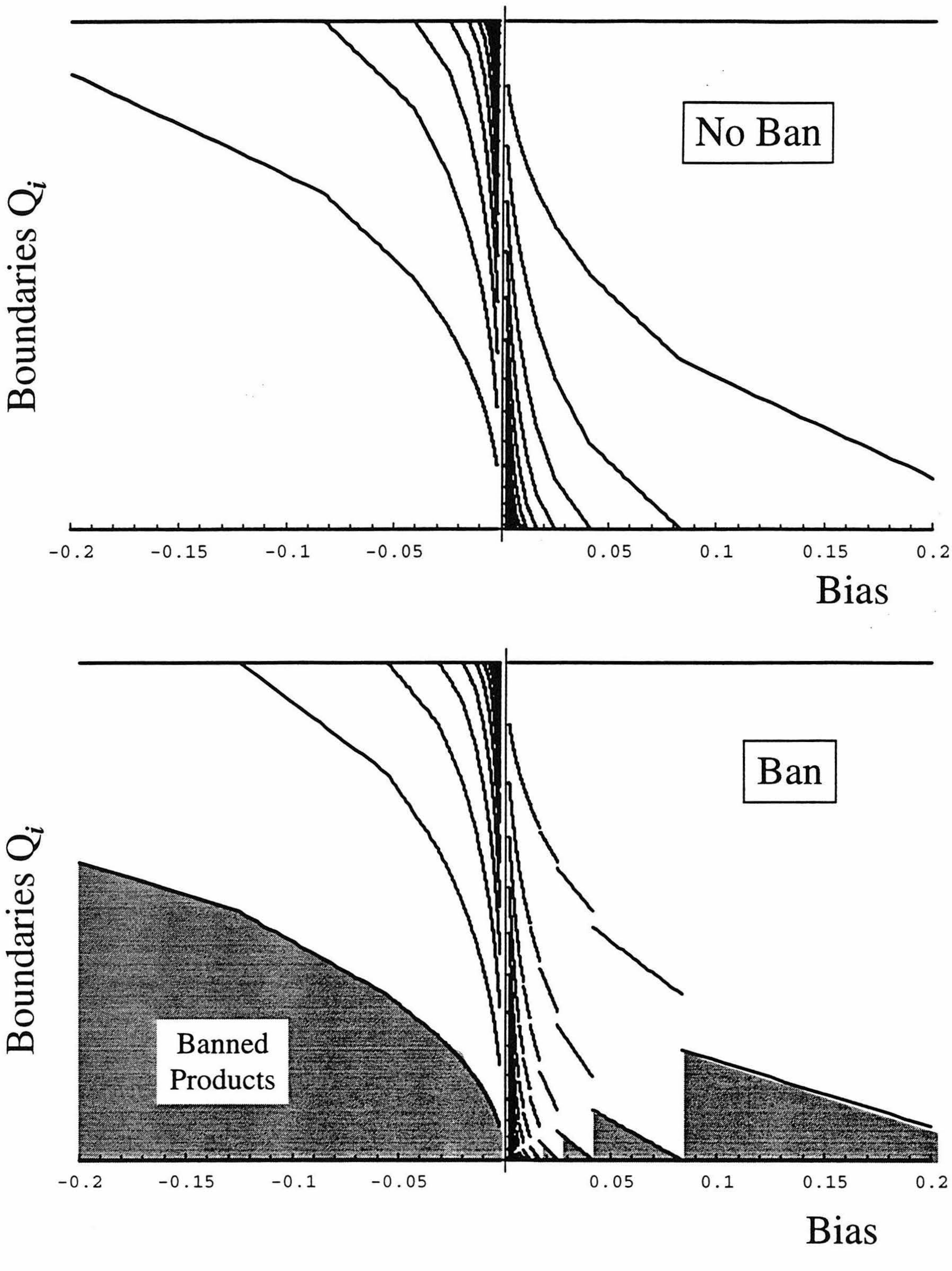

Figure 2.1: Solutions for Quadratic Preferences, Uniform Distribution

For any equilibria of such a game with bans, there is an equilibria of the corresponding no-ban game with higher ex-ante expected welfare. And if bans are not allowed, *all* groups ex-ante prefer that regulator bias be made as small as possible.

The different groups have different ex-ante preferences, however, over bans and regulator bias given bans. Specifically, producers of competing products ex-ante prefer that bans be allowed on this product, and given bans competing producers ex-ante prefer regulators to be biased as far as possible in their favor. Regulators also prefer bans be allowed when they are biased in favor of these competing producers, though they do prefer their bias to be as small in magnitude as possible. In contrast, potential consumers and producers of this product always ex ante prefer no bans and a small bias, as do regulators biased in favor of these producers.

2.2.6 A Counter Example

Given the fact that bans never improve ex-ante consumer welfare in the full competition case just described, one might conjecture that this result holds much more generally. After about a dozen trials of different functional forms, however, this author found a counter example. It can be constructed by making just one change to the above model.

Instead of a uniform distribution $F'(Q) = 1$ over Q, let us assume

$$F'(Q) = \tfrac{1}{2}(1 - Q)^{-1/2}.$$

This distribution is concentrated near $Q = 1$, but also has substantial weight over the whole $[0, 1]$ range. As shown in figure 2.2, for $\beta \in [-.5, -.18]$ the game with a ban has a two interval ($n = 2$) solution, while the no-ban game only has a one interval ($n = 1$) solution[4]. And for $\beta > -.24$, this two interval solution has a higher ex-ante expected consumer welfare. Thus for $\beta \in [-.24, -.18]$, the game with bans allowed is ex-ante better. For example, if $\beta = .2$, then the two interval solution has $Q_1 = .64$ and $\hat{Q}_1 = .88$. In this equilibrium 40% of products (those for which $Q \in [0, .64]$) are

[4]More precisely, Mathematica failed to find a solution.

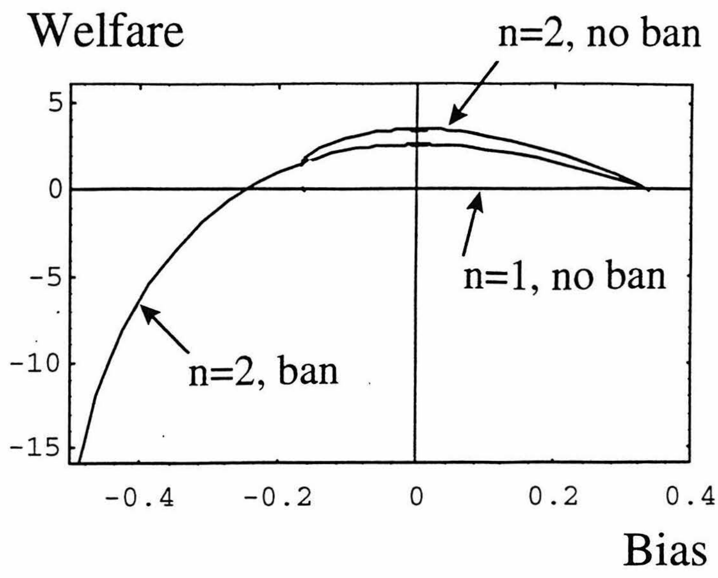

Figure 2.2: Counter-Example Welfare

banned.

It should be noted that for a wide range of other bias values in this model, the ban game is ex-ante worse, and often by much larger margins. See Figure 2.2.

Another counter example model arises from $F'(Q) \propto (1-Q)^{-1/4}$.

2.3 A General Banning Game

2.3.1 The Model

The following is a more general model of a cheap-talk labeling game, taken directly from the first published cheap-talk signaling game paper of Crawford and Sobel [CS82] (hereafter, C&S). (We will soon extend it to include a single forced "ban" action.)

There are two players, a Sender (S) who observes private information $a \in [\underline{a}, \bar{a}] \subset \mathcal{R}$ (the real line), and a Receiver (R) who takes an action $y \in [\underline{y}, \overline{y}] \subset \mathcal{R}$. This is a signaling game in that before R takes action y, but after S learns information a, S sends a signal s to R about a. It is a "cheap-talk" signaling game in that the player's twice continuously differentiable utility functions, $U^S(y, a)$ and $U^R(y, a)$, do not depend directly on the signal s. S's private information a is drawn from a differentiable c.d.f $F(a)$ with support $[\underline{a}, \bar{a}]$.

Sender S can be thought of as a regulator with special information $a = Q$ on product quality, and receiver R can be thought of as a market which chooses quantity purchased $y = \hat{Q}$ to maximize some effective objective. (For a competitive market, this objective would be total welfare.) Alternatively, S can be thought of as a "parent" who can either recommend a level of some "child" R's activity, such as driving, drugs, or sex, or can instead ban this activity entirely.

Instead of dealing directly with utilities $U(y, a)$, it will usually be more convenient to deal with marginal utilities $M(y, a) = U_1(y, a)$. Following C&S, we will assume *concavity*, $M_1 < 0$, and *sorting* (or single-crossing), $M_2 > 0$, everywhere. These imply strictly increasing and unique ideal points $\bar{y}(a) = \text{argmax}_y U(y, a)$. Futhermore, we w.l.o.g. assume[5] $M^R(y, y) = 0$, which implies $\bar{y}^R(a) = a$. Unless otherwise noted, we will also assume $[\underline{a}, \bar{a}] \subset [\underline{y}, \overline{y}]$. Finally, we will often want to assume that S's "bias" relative to R's preferences, $\bar{y}^S(a) - \bar{y}^R(a)$, has a constant sign, either positive from $M^S > M^R$ or negative from $M^S < M^R$.

In the game C&S considered, S first observes a and then sends signal s to R, who then chooses an action y. In this paper, we compare this basic labeling game with an extended labeling plus ban game, where S can choose to force a certain "ban" action x instead of sending a signal s to R. If S chooses x, the game ends immediately. Note that this extended game mixes cheap talk and costly signals within the same strategy space; the sender S chooses between "talking" and "doing."

2.3.2 Equilibria

Concavity ensures that R never uses mixed strategies, so R's strategy can be written as $y(s)$. S may use mixed strategies, so we write S's strategy as $q(s|a)$, a probability density of s given a, where $1 - \int q(s|a)da$ is the probability of x given a.

There will be some set of actions Y induced in any sequential equilibrium, and since talk is cheap and S has all private information, S essentially can choose any

[5]If this is not true for $U(y, b)$, use $\tilde{U}(y, a) = U(y, a(b))$ with $a'(b) = -M_1^R(b, a(b))/M_2^R(b, a(b))$.

$y \in Y$. C&S's lemma one shows that this set of equilibrium actions is finite given $M^S > M^R$ or $M^S < M^R$, and their proof is valid without modification in our extended game.

Sorting and concavity ensure that the set of S types a who choose y_i are of the form $[a_i, a_{i+1}]$. (We needn't be very precise regarding the behavior of the measure-zero set of boundary types a_i.) When Y has n elements, we have $a_0 = \underline{a}$, $a_n = \bar{a}$.

When several signals s induce the same action y_i, R's expected utility must be the same given each of them, which follows if $q(s|a)$ is the same for each such s, given $a \in [a_i, a_{i+1}]$. Since the signal space will be partitioned into sets whose members are not meaningfully distinguished by the players, all that really matters about the space of possible signals s is its cardinality, which we will assume is large enough to not be a constraint. Following C&S, we ignore signals from here on, describing equilibria by $\vec{a} = (a_i)_{i=0}^{n}$ and $\vec{y} = (y_i)_{i=0}^{n-1}$.

C&S's theorem one shows that, for the basic game, there exists a N such that for every integer n in $[1, N]$, there exists a sequential equilibrium where $\vec{a}, \vec{y}$ satisfy equations $(\mathcal{E}_i^S)_{i=1}^{n-1}, (\mathcal{E}_i^R)_{i=0}^{n-1}$, where equation $\mathcal{E}_i^S(y_{i-1}, a_i, y_i)$ is

$$U^S(y_{i-1}, a_i) = U^S(y_i, a_i),$$

and equation $\mathcal{E}_i^R(a_i, y_i, a_{i+1})$ is

$$y_i = \bar{y}(a_i, a_{i+1}) = \operatorname{argmax}_y \int_{a_i}^{a_{i+1}} U^R(y, a) dF(a).$$

Furthermore C&S show that every equilibrium is essentially equivalent to one of these.

These equations $\mathcal{E}_i^S, \mathcal{E}_i^R$ can be rewritten in terms of M as

$$\int_{y_{i-1}}^{y_i} M^S(y, a_i)\, dy = 0,$$

$$\int_{a_i}^{a_{i+1}} M^R(y_i, a)\, dF(a) = 0.$$

Alternatively, we can rewrite these equations $\mathcal{E}_i^S, \mathcal{E}_i^R$ as

$$U^S(y_{i-1}, a_i) = U^S(y_i, a_i),$$

$$I^R(y_i, a_i) = I^R(y_i, a_{i+1}),$$

where we have defined a pseudo-utility

$$I^R(y, a) = \int_y^a M^R(y, a')dF(a').$$

If x is too low or high, S may strictly prefer to never choose x. Such an equilibria of the extended game can be identified with an equilibria of the basic game with no forced act x.

For an equilibria of the extended game where the forced act x is chosen by S with positive probability, the only change to these equilibrium equations is that for some $i = \hat{\imath}$, the equation $\mathcal{E}_{\hat{\imath}}^R(a_i, y_i, a_{i+1})$ is replaced by $y_i = x$. In this case there is no longer any direct dependence between $a_{\hat{\imath}}$ and $a_{\hat{\imath}+1}$. Thus for *interior* forced acts $\hat{\imath} \notin \{0, n\}$, the equilibrium equations are *divided* into disjoint sets, sets where i is above and below the $\hat{\imath}$ where $y_{\hat{\imath}} = x$.

For every equilibria of the basic game there is for some x an equilibria of the extended game with the same $\vec{a}, \vec{y}$. This is because if we set $x = y_i$ for any y_i in the basic game, S won't want to change his strategy, and hence neither will R. Thus we may w.l.o.g. analyze only equilibria of the extended game where x is chosen with positive probability.

C&S's proof of their theorem one (which proves existence) applies to our extended game as well, if we simply replace the equation $\mathcal{E}_{\hat{\imath}}^R$ with $y_{\hat{\imath}} = x$. We again have a set of continuous non-linear difference equations with the same sort of properties. The only significant difference is that these equations may be divided. So $n = 1$ is possible only if either all or no S types a prefer the forced act x. Thus the following lemma applies.

Lemma 2.1 *Given a forced act x, there exists a $\hat{N}(x)$ such that for every $n \in$*

$[\underline{N}, \hat{N}(x)]$, for $\underline{N} \in \{1,2\}$, there is a Bayes-Nash equilibrium (which is also sequential) satisfying $(\mathcal{E}_i^R)_0^{n-1}, (\mathcal{E}_i^S)_1^{n-1}$, except that for some i, $\mathcal{E}_i^R$ may be replaced by $y_i = x$.

2.3.3 Comparing Welfare

To simplify their analysis C&S invoked a *monotonicity* assumption equivalent to $da_i/dx > 0$ for all $i \in [1, n-1]$, which implies $dy_i/dx > 0$ for all $i \in [0, n-1]$, and also implies that in the basic game there is a unique equilibrium for each size n. C&S's lemma three proves this, and their proof applies here. The only modification is that interior forced acts can induce some non-uniqueness; an $n = 7$ solution, for example, might have 3 parts on one side and 4 parts on the other, or these numbers might be reversed.

C&S showed that a sufficient condition for monotonicity is (something slightly weaker than) $M_1^R + M_2^R \leq 0$ and $M_1^S + M_2^S \geq 0$. An alternative and perhaps more intuitive sufficient condition for monotonicity is available, however.

Lemma 2.2 *Monotonicity is implied by I^R having steeper isoquants than U^S in (y,a) space, i.e., by*

$$\frac{I_1^R}{I_2^R} \leq \frac{U_1^S}{U_2^S}.$$

(Non-trivial proofs are in the Appendix.) Note that all isoquants of both I^R and U^S are in the $(+,+)$ direction.

Let us now collect together all of C&S's assumption we plan to use and call them *C&S's standard 1D cheap-talk* assumptions. These assumptions are: $[\underline{a}, \bar{a}] \subset [\underline{y}, \bar{y}]$, concavity $M_1 < 0$ and sorting $M_2 > 0$ in preferences, monotonicity $da_i/dx > 0$, and either $M^S > M^R$ or $M^S < M^R$. We will from here on make these standard assumptions unless we state otherwise.

Let us also define a *mixed* agent T to be one for which $M^T = \theta M^R + (1-\theta) M^S$ for some function $\theta(y,a) \in [0,1]$. A mixed agent has a marginal utility intermediate between the sender S and receiver R. (Both S and R are mixed agents.) Finally, let us define z_n to be the y_0 in the n step equilibrium of the basic game. z_n is the lowest act the receiver would voluntarily take.

Given monotonicity and a positive sender bias, we can show that both the sender and receiver, or any mixed agent in between, ex-ante prefers an n step equilibrium of the basic game to an n step equilibrium of the extended game where $z_n \geq x = y_0$. Such a forced act is the lowest act taken in equilibrium, and is no more than the lowest act taken when no forced act is possible.

Lemma 2.3 *Given $M^S \geq M^R$, any mixed agent prefers an n step equilibrium of the basic game to an n step equilibrium of the extended game where $z_n \geq x = y_0$. The preference is strict if the equilibria are distinct.*

We can also show that one can't get any more equilibrium steps by introducing a lowest-taken forced act x that is within the range of the actions the receiver might take if she were fully informed.

Lemma 2.4 *Given $M^S > M^R$, for any n step equilibrium of the extended game where $\underline{a} \leq x = y_0$, there exists an n step equilibrium of the basic game.*

Putting together lemmas 2.3 and 2.4, we can conclude that for $M^S > M^R$ and $y_0 = x \in [\underline{a}, z_n]$, a mixed agent ex-ante prefers the basic game.

Theorem 2.1 *Given C&S's standard 1D cheap-talk assumptions, and $M^S > M^R$, for any n step equilibria of the extended game where $y_0 = x \in [\underline{a}, z_n]$, there exists an n step equilibrium of the basic game which any mixed agent ex-ante prefers. This preference is strict if the equilibria are distinct.*

Since both S and R are mixed agents, an immediate corollary is that both S and R ex-ante prefer the basic game in this situation.

One of the standard 1D cheap-talk assumptions is that $\underline{y} \leq \underline{a}$, which implies says that there is a distinct best action $\bar{y}(a)$ for the receiver for any true signal a. In the product quality domain this is equivalent to saying that quality cannot be negative; there is only one fully-known quality level where none of the product would be purchased.

We can also extend the result of theorem 2.1 to the case where, while holding $\underline{a}$ fixed, we allow $\underline{y}$ to vary up into the range where $\underline{y} > \underline{a}$. This allows for negative product quality.

Let us define $\hat{z}_\infty = \max(\underline{y}, \underline{a})$, the lowest action the receiver would take given full information. Let us also define $\hat{z}_n = \max(\underline{y}, z_n)$, where z_n is as before, the y_0 of the n step equilibria of the basic game where $\underline{y} \leq \underline{a}$. Then $\hat{z}_n$ must be the y_0 of the n step equilibria of the basic game, assuming it exists, for any value of $\underline{y}$. (For $\underline{y} \leq z_n$ the equilibria is not changed, and for $\underline{y} > z_n$, monotonicity requires $y_0 = \underline{y}$.) Using these definitions, we can express a more general result.

Theorem 2.2 *Given C&S's standard 1D cheap-talk assumptions, except that we allow $\underline{y} \geq \underline{a}$, and given $M^S > M^R$, for any n step equilibria of the extended game where $y_0 = x \in [\hat{z}_\infty, \hat{z}_n]$, there exists an n step equilibrium of the basic game which any mixed agent ex-ante prefers. This preference is strict if the equilibria are distinct.*

2.3.4 Applications

These general results translate to the domain of product bans as follows. Assume that more of some product is purchased when consumers expect it to be of a higher quality. Assume that for any given fully-known quality level more of the product would be purchased in a fully competitive market than in the actual market. Assume that the regulator would prefer at least this much of the product be purchased. And assume the regulator has private information about product quality.

Finally, assume that a product "ban" results just in a situation equivalent to that where consumers expect some low quality level, without any further enforcement costs or losses. Assume that the amount of the product purchased under a ban is somewhere between the amount which would be purchased if consumers were certain that the product was of the worst possible quality, and the amount which would be purchased if the not-entirely trusted regulator, with no authority to ban, simply declared the product to be of the worst possible quality.

Given these assumptions, neither the regulator nor an external observer preferring the fully-competitive market quantity would ex ante prefer that the regulator have the authority to ban the product. With bans possible, however, the regulator will sometimes choose bans in equilibrium.

A similar translation can be made to the domain of a parent concerned about a child's level of some risky activity.

Assume that the child would do more of the activity if it were safer. Assume that for any given fully-known safety level, an external observer would prefer to see more activity than the child would, and that the parent would prefer even more activity than that. Assume the the parent knows more about the risk level than the child. Finally, assume that if the parent were to "ban" this activity the resulting activity level would be somewhere between what the child would choose if certain that the risk were the highest possible, and the level the child would choose if the not entirely-trusted parent, with no ability to ban, were to just tell the child that this risk is maximal.

Given these assumptions, neither the child, the parent, nor the external observer would ex-ante prefer that the parent be able to ban the activity. Given the power to ban, however, the parent will sometimes choose to ban.

2.4 Conclusion

This paper presents a game-theoretic model of product bans, intended to combine the best elements of the captured regulator, public interest, and irrational consumer models, and to explain a number of empirical regularities regarding product bans.

The basic intuition is that when a "nanny" keeps dangerous things out of someone's reach, that someone becomes more complacent about possible harms. This in turn encourages "paternalism" in the nanny, who is unwilling to live with the consequences when complacence encounters real danger. When this person knows that the nanny must treat them as an "adult," however, and can only warn them about dangers, they become more cautious. And overall, people may be better off being treated as adults. This intuition applies to the relationship between parents and children, and to the relation between a "nanny state" regulator and consumers.

This intuition is formalized in a simple model of a one-dimensional choice set and a single extreme action which can be forced. Focusing on the product ban application, we consider simple products with no use externalities, or relevant signaling function, and we imagine that regulators have quality information not otherwise available to fully rational consumers. We assume that regulators are rewarded only by retrospective voters attending to their ex post utility levels, we assume some small degree of either regulator capture or non-competitive markets, and we assume that product labeling or banning are the only policy options available to the regulator.

Given these assumptions, regulator labeling is a cheap-talk signaling game, and so there are cases where the regulator would rather ban a product than live with the consequences of consumers who don't believe their labeling advice, but who will nonetheless hold the regulator accountable for the consequences of consumer choices.

Giving the regulator the ability to ban products changes the equilibria, and in some situations this makes consumers ex-ante better off, while in other cases consumers become worse off. While it seems easier to generate examples where consumers become worse off, more work needs to be done to find a systematic way to distinguish these cases. A general sufficient condition for the superiority of prohibiting bans has been given, but there is much that it doesn't cover.

If bans are ex-ante worse, product bans can be viewed as commitment failures. Without a commitment not to ban, regulators will want to ban sometimes, and consumers will base their inferences on this possibility. With a commitment not to ban, consumers make different inferences, to their ex-ante benefit. A commitment not to ban also aligns the ex-ante preferences of all actors toward minimizing the magnitude of regulator bias. Perhaps we should consider a constitutional prohibition on regulator-information motivated product bans, similar to the U.S. prohibition on print media bans.

There are also many other directions for future work. It would be nice to better characterize when bans are ex-ante better, to consider bias functions which can be zero at points, and to consider the case of where a regulator can choose to either ban or require a product. And it is important to examine the degree to which the cheap-talk aspect of regulator recommendations is broken by repeated play with some later information revelation to consumers about true product quality.

Beyond this, we might consider imperfect enforcement, imperfect commitments to not ban, and regulator powers to place taxes or subsidies on products. We might allow consumers to be uncertain about regulator bias. And we might let regulators signal via "burned money", such as purposely expensive advertizing [ASB95]. We might also model the case where one or more private information sources can also certify the product, and we might explicitly model exogenous label-reading costs and ban enforcement costs. Finally, we might consider a model with endogenous quality, where producers decide what products to develop and market.

Empirical work to illuminate the range of application of this model is also appropriate. Lab experiments should verify that people really do play according to one of the analyzed equilibria of such games, and data on the rates and correlates of activity bans which parents impose on their nearly-mature children should be informative.

Finally, the correlates of regulator product bans may help us to confirm or reject this cheap-talk model of such bans. This model predicts political debates focused on quality levels, predicts that bans will focus on new unfamiliar harms about where better regulator information is plausible, predicts bans by regulators biased in favor of a product, and predicts large rates of bans from small levels of bias. The model also suggests larger rates of product bans from regulators whose advice consumers treat more skeptically. Finally, the model also suggests that voters and candidates will be unreceptive to proposals for ban-exceptions using special stores and tests, and suggests that consumer disregard for product labels is due more to skepticism than cognitive processing limits. One or more of these predictions may be testable empirically.

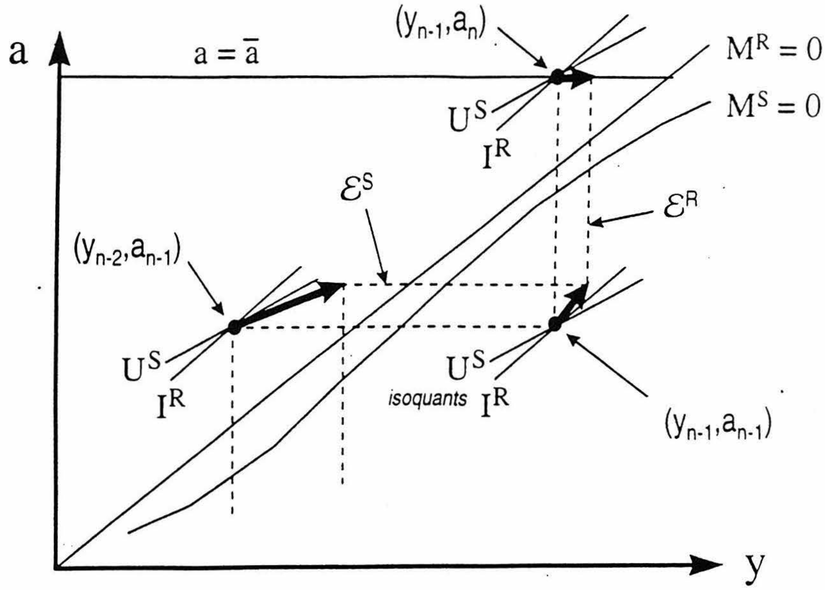

Figure 2.3: Aid To Monotonicity Proof

2.5 Appendix

2.5.1 Proof of Lemma 2.2

PROOF: (See Figure 2.3.) An equilibrium can be thought of as a path through the (y, a) plane.

The forced act point (x, a_{i+1}) for $i = \hat{\imath}$ is connected by a horizontal line representing equation $\mathcal{E}^S_{i+1}$ to a point (y_{i+1}, a_{i+1}). A vertical line representing equation $\mathcal{E}^R_{i+1}$ then connects this to a point (y_{i+1}, a_{i+2}). This zig-zag pattern continues until a vertical line reaches the last point $(y_{n-1}, \bar{a})$. If x is interior, then going in the other direction from point (x, a_{i+1}), the zig-zag pattern ends at a vertical line reaching $(y_0, \underline{a})$.

As x varies, each point (y, a) will vary with some vector $(dy/dx, da/dx)$, and the equations corresponding to each line connect the vectors on the two ends of the lines. The $\mathcal{E}^S_i$ equations say that if the vector on one side of a horizontal line cuts the U^S isoquants so as to move toward lower U^S, the other side must cut in the opposite direction, so as to also move toward lower U^S. Since points connected by lines are on opposite sides of the maximal Q^S (i.e., $M^S = 0$) curve, opposite movement means they both move "inside" (toward $M^S = 0$), or both move "outside" (away from $M^S = 0$). Similarly, the $\mathcal{E}^R_i$ equations say that the vector must cut the I^R isoquants in opposite directions from the $y = x$ (i.e., $M^R = 0$) line at the two ends of a vertical line.

The end point $(y_{n-1}, \bar{a})$ must vary along the line $a = \bar{a}$, which as y_{n-1} increases must cut inside the isoquants of I^R. Thus at (y_{n-1}, a_{n-1}) the vector must also be inside I^R, and is hence positive, and the assumption of a steeper I^R vector implies inside U^S as well. Continuing, at (y_{n-2}, a_{n-1}) the vector must cut inside U^S and be positive, which is now also inside I^R, since we are on the other side of $M^S = 0$.

Since all the constructions leave the vector in the $(+,+)$ direction, the vector at (x, a_{i+1}) is also in this direction, and so all da/dx and dy/dx are positive for $i > \hat{\imath}$. The same argument applies when starting from end point (y_0, r) which varies along the line $a = \underline{a}$, implying that all dy/dx and dy/dx are positive for $i < \hat{\imath}$. QED.

2.5.2 Proof of Lemma 2.3

In general ex-ante expected utility for any agent is

$$E[U] = \sum_{i=0}^{n-1} \int_{a_i}^{a_{i+1}} U(y_i, a) dF(a).$$

To compare utility between equilibria, we will continuously vary one equilibrium into another, and in a way such that the local derivative of $E[U]$ along the way maintains the same sign. (This proof closely follows C&S's proof of their theorem three.)

Let us define y_{-1} to be the solution to $\mathcal{E}_0^S(y_{-1}, \underline{a}, z)$. For $x < y_{-1}$, S never chooses x, so the extended and basic equilibria are identical.

For $x \geq y_{-1}$, as we vary $x = y_0$ in the extended game from y_{-1} to x to z, we move from an equilibria with the same y_i as an $n+1$ step solution of the basic game over $[y_{-1}, \bar{a}]$, to an n step solution of the extended game on $[\underline{a}, \bar{a}]$, to an n step solution of the basic game on $[\underline{a}, \bar{a}]$. (All the while we hold $[\underline{a}, \bar{a}]$ fixed.) Along this path, the local derivative of $E[U]$ with respect to x is

$$\frac{dE[U]}{dx} = \sum_{i=0}^{n-1} \left(\frac{dy_i}{dx} \int_{a_i}^{a_{i+1}} M(y_i, a) dF(a) - \frac{da_i}{dx} \int_{y_{i-1}}^{y_i} M(y, a_i) dy \right).$$

Given monotonicity all the da_i/dx and dy_i/dx are positive.

For S, the second integral is exactly zero for all i, and so given $M^S \geq M^T$, for T this integral term contributes non-negatively. For R the first integral is exactly zero for all $i \neq 0$, and for $i = 0$ is positive for any $x \in [y_{-1}, z)$, since x is less than R's ideal $\bar{y}(r, a_1)$ where this integral is zero. (x is clearly less at $x = \underline{a}$, and by continuity stays less across this range; otherwise it would be equal somewhere between, identifying another n step solution of the basic game, which contradicts monotonicity.) Thus for T the first integral contributes positively.

Thus the sum is positive, and so T strictly prefers a distinct n step equilibrium of the basic game to an extended n step equilibrium for $x \geq r$. QED.

2.5.3 Proof of Lemma 2.4

PROOF: We need to show that for $\underline{a} \leq x = y_0$ and $M^S > M^R$, introducing the forced act x does not increase the number of equilibrium steps possible, i.e., that $N \geq \hat{N}(x)$.

Let us focus first on $x = \underline{a}$ in the extended game. And let us consider allowing the range $[\underline{a}, \bar{a}]$ to vary within a larger range $[\underline{r}, \bar{r}]$. A varying $F_{\underline{a},\bar{a}}$ will be obtained by conditioning on a differentiable c.d.f $G(a)$ with support $[\underline{r}, \bar{r}]$, as in

$$F_{\underline{a},\bar{a}}(a) = \frac{G(a) - G(r)}{G(r') - G(r)}.$$

Observe that since the equations $\mathcal{E}_i^S, \mathcal{E}_i^R$ are in terms of continuous functions of the a_i, y_i, their solutions vary continuously as we vary $\underline{a}$, in both the case where $\underline{a} = x$ and where there is no forced act. The only thing preventing one from taking a solution with n steps over range $[\underline{a}, \bar{a}]$ and collapsing it all the way to $[\bar{a}, \bar{a}]$ is that at some point the solutions found this way will begin to violate one of the constraints $a_i \leq a_{i+1}$ or $y_{i-1} \leq y_i$.

If there were a sequence of solutions to $\mathcal{E}_i^S$ approaching $y_{i-1} = y_i$, then we would have $M^S(y_{i-1}, a_i) = 0 \geq M^R(y_{i-1}, a_i)$. $M_2 > 0$ and $\mathcal{E}_{i-1}^R$ imply that if this $M^R(y_{i-1}, a_i)$ is zero, then $a_i = a_{i-1}$, and otherwise $\mathcal{E}_{i-1}^R$ has no solution. Thus as r increases, $N_{rr'}$ and $\hat{N}_{rr'}$ can decrease only immediately after the r points which have *boundary solutions* where $a_i = a_{i+1}$ for some i.

$\mathcal{E}_i^R$, $M_2^R > 0$, and $M^R(y, y) = 0$ imply that if any two of a_i, y_i, a_{i+1} become equal, all three must be equal. Similarly, $M^S \geq M^R$, $M^R(y, y) = 0$, and $M_1 < 0$ imply that $\mathcal{E}_i^S$ has no solution for $y_{i-1} < a_i = y_i$. Taken together these imply that boundary solutions must satisfy $\underline{a} = a_0 = a_1$, in both the basic and extended game, with $a_i < a_{i+1}$ for all $i > 0$. In the basic game $\mathcal{E}_0^R$ implies $y_0 = a_0$, while in the extended game $x = \underline{a}$ implies the same thing. Thus both games will have a boundary solution for a given $\underline{a}$ if either does. Since for $\underline{a} = \bar{a}$ we have $N = \hat{N} = 1$, and since $N, \hat{N}$ change by one unit at a time, we must have $N_{\underline{a}\bar{a}} = \hat{N}_{\underline{a}\bar{a}}$ for all $\underline{a}, \bar{a}$ for $x = \underline{a}$.

Now consider forced acts $y_0 = x \geq \underline{a}$. For any equilibrium of the extended game with $x = y_0$, a_0 only appears in the constraint $\underline{a} = a_0 \leq a_1$. Thus given such an equilibrium, we can construct other equilibria on other ranges by simply varying $\underline{a} = a_0$ within $[\underline{r}, a_1]$. This implies that for $y_0 = x \geq \underline{a}$, $N_{\underline{a}\bar{a}} \geq \hat{N}_{\underline{a}\bar{a}}(x)$.

Thus for any equilibria of the extended game meeting the conditions, there is an equilibria of the basic game with at least as many steps. QED.

2.5.4 Proof of Theorem 2.2

The only change that $\underline{y} > \underline{a}$ induces in the basic game that in some equilibria the equation $\mathcal{E}_0^R$, which is $\bar{y}(\underline{a}, a_1) = \underline{y}$, is replaced by $y_0 = \underline{y}$ when $\bar{y}(\underline{a}, a_1) < \underline{y}$. Since $\hat{z}_\infty = \max(\underline{y}, \underline{a})$ and $\hat{z}_n = \max(\underline{y}, z_n)$, then $x \in [\hat{z}_\infty, \hat{z}_n]$ is the same as $x = \underline{y}$ or $x \in [\underline{a}, z_n]$. For $x = \underline{y}$ then, if there is an n step equilibrium of the extended game, then the only way there can fail to be an identical n step equilibrium of the basic game is if $\bar{y}(\underline{a}, a_1) > \underline{y}$ for the a_1 of this extended equilibrium. But by monotonicity $\underline{y} < z_n$, and so in this case $x \in [\underline{a}, z_n]$. For this condition, the n step basic game equilibria exists without modification from the $\underline{y} \leq \underline{a}$ case. Here the result of theorem 2.1

applies directly. QED.

Chapter 3 Voter Incentives to Become Informed

3.1 Introduction

3.1.1 The Question

In his classic 1957 book, "An Economic Theory of Democracy," Anthony Downs considered whether, in a democracy, voters have too little incentive to become informed:

> [a voter's] value of voting correctly ... is compounded from his estimates of his party differential and of the probability that his vote will be decisive. ... [and] is nearly infinitesimal under most circumstances. ... The result is an enormously diminished incentive for voters to acquire political information before voting. ... democratic election systems always operate at less-than-perfect efficiency. [Dow57]

Downs has recently reflected [Dow93] that "the way information costs are treated in that book is perhaps its most important contribution." However, while attention has recently turned to ways in which voters can make great use of inexpensive information shortcuts [Pop91, MO86, GW93], and to various other incomplete information models of electoral processes, little formal work has followed up on the question of voter incentives to become informed. (There are exceptions [Ger95].)

This neglect is unfortunate because we can imagine alternative political institutions in which voters would have much stronger incentives to become informed. For example, if we randomly selected a small jury to decide an election, each juror would have a high probability of being decisive. We would like to know whether to consider such alternative institutions.

This paper therefore attempts to develop a relatively-general fully-game-theoretic model of voter choices of efforts to become informed, and of candidate responses to those choices. Prominent prior analyses of related issues [Uhl89, FS89, FS90], have, in contrast, considered only simple partial-equilibrium models, with for example only two possible candidate positions and agents who even then do not fully optimize their behavior in response to the strategies of others.

With this general model in hand, this paper then attempts to clarify what does and does not dilute voter incentives to become politically informed. We will find that voters can have surprisingly strong incentives to become informed, that information scale economies can overwhelm large group free-riding problems, and that even when information is free, voters can prefer ignorance to full information.

3.1.2 Model Overview

This paper analyses a fundamentally prospective two-candidate election. That is, before an election each of two candidates must simultaneously announce their policy positions, policies which they are committed to implementing if elected. In contrast to most prospective election models, however, we here allow "policy-day" to occur well before election-day, to allow for the time and repetition required to communicate policies to wider audiences.

Also, voters do not directly base their election-day voting decisions on announced policies, for two reasons. First, information may be revealed between policy-day and election-day regarding the consequences of candidate policies. For example, a candidate may announce a policy in March favoring free trade, but the consequences of this policy may become clearer by November as the national economy weakens, as trading partners clarify their intentions, and as pundits analyze the situation. Thus candidate policy-day positions are really lotteries over election-day positions.

Second, voters need not be perfectly informed about candidate election-day positions when they vote. Instead, voters may receive only imperfect signals regarding such positions, signals whose quality (i.e., relation to candidate positions) may come at some personal cost. Voter efforts to become politically informed are conceived of very broadly here, and may include any effort which improves one's political signals or ability to interpret them, such as attending college, subscribing to a newspaper, or watching a debate.

Strategically, it will be useful to divide voter information efforts into two categories: *early* efforts, like attending college, made before candidates announce their policies and visible to candidates, and *late* efforts, like watching a debate on TV, either made after the candidates announce their policies, or not visible to those candidates.

The main space of possible candidate policies modeled here will be roughly "distributive." That is, a candidate's election-day position says how large a "pie-slice" each of the finite set of voters would receive if this candidate were elected. The size of the total "pie," however, is not fixed. Instead, a candidate's policy-day lottery can contain election-day positions with different total pie sizes. An announced policy of an aggressive military buildup may, for example, increase the national pie if potential adversaries back down, but decrease the national pie if adversaries respond in kind.

This space of candidate positions is also generalized in two ways. First, candidate distributive positions can be constrained by the existence of voter "groups," which are such that all members of a group must be given the same size pie slice. There are no group leaders here, however, nor any other non-trivial form of group organization. Second, conditions are identified which allow one to add any arbitrary policy space to the analysis without changing the conclusions.

Putting this all together, then, this paper's main model contains the following stages

1. Voters choose early (visible) information efforts.

2. Candidates choose policy-day positions (= lotteries on election-day positions).

3. Voters choose late (or invisible) information efforts.

4. Nature chooses election-day positions from policy-day lotteries, and chooses each voter's noisy signals regarding election-day positions.

5. Voters chooses their votes, after seeing their signals regarding the "pie slice" their group would receive from each candidate.

3.1.3 Key Insights

This model can be used to present a number of insights regarding voter incentives to become informed, some of them counter-intuitive. While some of these insights can be, and have been, presented in the context of simple partial-equilibrium models, their elaboration in this more general context should increase our confidence that they are relatively general features of electoral games, and should help us understand their range of applicability.

Downs, in addition to remarking on infinitesimal voter "incentives to acquire political information" (as in the previous quote) also remarked that

> the more information a citizen has, the more influence over government policy he is likely to exercise, provided he informs the government what his preferences are. ... an influencer's intervention value may suffer hardly any discount because only a small number of others are interested in the policy he wants to influence. ... Such ignorance ... stems from the great cost of obtaining enough information to exert effective influence. ... The complexity of these areas often forces influencers to become experts before they can discover what policies best suit their own interests. [Dow57]

That is, Downs noticed that efforts by citizens to become informed in order to communicate their preferences to politicians do not suffer a dilution due to low probabilities of being pivotal in an election, if these communications are made *before* these politicians choose their policy positions.

Downs claimed that this vast leverage in influence is available only to a small minority, apparently because a large minimum effort is required before one can become informed enough to use this channel of influence. Similarly, most discussions of interest group politics [Uhl89, Ros93, Wal91] allow group leaders, but not group members, to expect their actions to have a non-negligible influence on candidate policies.

Downs' large-minimum-effort claim is difficult to understand, however. In theory, any credible visible signal about how informed one will be regarding candidate positions on election day should have a direct influence on those candidates, undiluted by the probability of being pivotal. And, contrary to Downs, this should be true *regardless* of the level of information signaled, though of course a signal of a high information level should induce a stronger reaction than a signal of a low information level. For example, simply subscribing to a newspaper is a relatively cheap visible signal of a minimal information level.

The basic game-theoretic concept here is that the first-mover in a game often gains an advantage by her ability to commit to a course of action up front. Voters can commit to becoming informed, just by becoming informed, and then credibly signaling that fact.

The other major reason cited (most famously by Olson [Ols65]) for weak voter information incentives is free-riding within large voter interest groups. If candidate positions must treat all interest group members the same, then without some special group organization each group member might rather that the other members make whatever efforts are required to convince candidates to fear their group's wrath.

What group member efforts need to produce, however, is information, and it is worth pointing out that there can be large scale economies in information production. Imagine, for example, a situation where if one group member bothered to learn a candidate's position toward his group, he could costlessly communicate this information to all other members of his group. In such a case, information-production scale economies could outweigh even a very extreme group free-riding problem, so that voters prefer to be members of larger groups.

Note, however, that voters having strong incentives regarding their political information need not imply that such voters want more information, even when information is free. There are many games, including electoral games [Ree89], where players prefer less information to more, at least if other players become aware of and can respond to this fact. You can prefer, for example, that potential extortionists believe that you don't know whether they would carry out a threat to hurt you.

As suggested by Fiorina and Shepsle [FS90, FS89], voters can prefer a "negative news" form of ignorance, where one's probability of supporting a candidate is concave in the position they take. Voters who are more likely to learn of candidates who hurt them than of candidates who help them will be effectively "negative voting." Not only can this concavity induce stability (or zero variance) in candidate positions, but the optimal form of ignorance will induce the same expected benefits from candidates as full information.

This possible preference for strategic ignorance offers a caveat regarding mechanisms like the voting juries mentioned above and other stochastic voting mechanisms which have been suggested [Lev89, Sto90, Dah89]. While such mechanisms can in fact induce voters to become better informed, strategically ignorant voters can prefer to commit to not using such mechanisms.

3.1.4 Technical Features

The game model used here has some noteworthy technical features.

First, this model assumes that candidates maximize something like plurality. This is a common assumption, and is equivalent to maximizing the probability of winning under a variety of conditions, reviewed by Coughlin [Cou90]. For example, Hinich [Hin77] justifies it by assuming everyone votes, their vote distributions are independent, and the electorate is large.

Second, noisy voter signals imply that candidates are not sure about each voter's response to their positions. Hence this is a probabilistic voting model [Cou92, Led84]. However, in contrast to probabilistic voting models which are driven either by voter mistakes or by exogenous candidate uncertainty over voter preferences, here all actors are full rational and candidate uncertainty is endogenous – all the model specifies is a mapping between voter effort and the resulting conditional signal distribution.

Third, this model differs from previous probabilistic voting models in considering mixed strategy equilibria for the candidate-policy-position game. While previous models, such as Coughlin's [Cou92], have been able to demonstrate results such as that candidates give better offers to voters who they expect will be better informed, these results have been limited to the case of probability distributions which imply concave candidate expected payoffs, and hence pure strategy candidate choices. This excludes, for example, small normally-distributed errors over distributive policies.

Mixed strategies are examined here even though good reasons have been offered [Ord86] to be wary of using them to model what may be considered an intrinsically dynamic process of position adjustment. Others, such as Myerson [Mye93], have gained useful insights through mixed-strategy electoral models. And while it seems clear that real candidates can often react to opponent positions, it also seems clear that there remains some inertia in the process, and that candidates often remain uncertain on election day about effective opponent positions (how an opponent's position will be perceived by voters). A fully multi-stage adjustment process is more complex to analyze, and even so can reduce to a single-stage game in the last period [Led89].

Finally, the use of mixed strategies over policy-day positions, which are themselves lotteries over election-day positions, implies that candidates are in effect choosing from a convex set of distributions over election-day positions. Instead of explicitly representing this convex set in terms of a finite number of "pure strategy" boundary-vertex lotteries, however, we will instead describe this convex set in terms of a finite number of boundary-surfaces specified by linear inequalities.

That is, we will allow a candidate to choose any lottery over pie-slices which satisfies a given set of inequalities regarding the expected value of various functions. For example, candidate lotteries over pie-slice allocations might have to keep the expected value of the total pie size below some bound. Also, a candidate's ability to "gamble" with the total pie size might be limited by a constraint on the variance of total pie size within each lottery.

This may seem an odd way to describe candidate strategy sets. But given this approach, conditions are found under which it is a sequential equilibrium for candidates to choose their distributions to be *independent*, rather than correlated, across groups. And given such independent equilibria, candidates can be thought of as playing a number of independent electoral games, one for each voting group.

Thus independent equilibria allow us to dramatically simplify the description and analysis of such games. In contrast, previously known mixed-strategies over divide-the-pie games, such as the "Colonel Blotto" game [GW50], contain rather complex correlations. This simplification enables a simplified presentation of most of the applications in the paper, and should simplify future analysis of models in this family.

If one finds it difficult to accept candidate strategy spaces bounded by expectation inequalities, this approach to candidate strategy spaces can be alternatively understood as a technical trick to relax of the usual constraint of an exact fixed-pie, made primarily to simplify the resulting analysis. If only a small pie-size variance is allowed, for example, this might be considered only a slight relaxation of the usual constraint.

3.2 The Model

3.2.1 Players and Actions

A two-candidate election where voters have incomplete information is modeled as an extensive form game, with the sequence of events sketched in Figure 3.2.1.

The players are a set of two candidates $C = \{1, -1\}$, and a finite set of voters I, who are divided by a partition G into disjoint voter groups g, so that for each voter i, $i \in g(i) \in G \subset 2^I$. There are thus $|I|$ voters divided into $|G|$ groups, each of size $n_g = |g|$. Feminine and masculine pronouns will be used to denote candidates and voters respectively.

Each voter i will eventually chose a vote $v_i \in C$. The net vote for candidate 1 is

$$V = \sum_{i \in I} w_i v_i + \epsilon,$$

where w_i is the number of votes controlled by voter i, ϵ is some independent noise, a combination of coin flips used to break ties, vote miscounts, etc., with a continuously

Stage I Each voter i chooses early efforts $\tilde{a}_i$, then all players are told $\tilde{a} = (\tilde{a}_i)_{i \in I}$.

Stage II Each voter i chooses late efforts $\bar{a}_i(\tilde{a})$. Simultaneously, each candidate $c \in C$ chooses a distribution $\mathcal{F}_c(\tilde{a})$ over election-day positions x_c.

Nature Moves Nature picks $(x_{cg})_{g \in G} = x_c \sim \mathcal{F}_c(\tilde{a})$, and then $(s_i)_i = s \sim H(s|x_1, x_{-1}, \bar{a}, \tilde{a})$.

Stage III Each voter i sees signal s_i but not $\bar{a}_{-i}$, and then chooses a vote $v_i(s_i, \bar{a}_i, \tilde{a}) \in C$.

Nature Again Nature picks vote noise $\epsilon \sim N(\epsilon)$, and the max plurality candidate wins.

Figure 3.1: Order of Events in Game

differentiable c.d.f. distribution $N(\epsilon)$. The total ballot is $v = (\epsilon, (v_i)_{i \in I})$, and the winning candidate will be $\hat{c}(v) = 1$ if $V \geq 0$, and -1 otherwise.

Before the election, each candidate $c \in C$ must choose and commit to a probability distribution $\mathcal{F}_c$ over "divide a pie" election-day group positions $(x_{cg})_{g \in G} = x_c \in \mathcal{X}_c$ (compact and convex) $\subset R_+^{|G|}$, where $x_{cg} \in R_+$ is the "pie" amount given to each member of group g. (Distributions are also assumed to be independent across candidates.) This distribution $\mathcal{F}_c$ must satisfy M *distribution constraints* of the form

$$E_{\mathcal{F}_c}\left[f_{mc}(x_c)\right] \leq K_{mc} \tag{3.1}$$

for some continuous functions $f_{mc}(x_c)$ and constants K_{mc}.

For example, one such constraint might be an *average pie-size* constraint

$$E_{\mathcal{F}_c}\left[X_c(x_c)\right] \leq \bar{X}_c > 0 \tag{3.2}$$

where $X_c(x_c) = \sum_{g \in G} n_g x_{cg}$. Here candidates need not divide the same total distributive pie X_c, but can instead in effect make limited "gambles" with their total expected pie $\bar{X}_c$. Another example constraint is

$$E_{\mathcal{F}_c}\left[(X_c(x_c) - \bar{X}_c)^2\right] \leq K,$$

where K here constrains the variance of X_c. Setting $K = \bar{X}^2$ would constrain the variance to be zero. We assume, however, that there exists some distribution such that all these constraints are not binding (i.e., Slater's condition).

Before candidates choose their positions, each voter chooses an early effort level $\tilde{a}_i \in \tilde{A}_i$ (compact and convex) regarding how informed he will be about candidate positions. All these efforts $\tilde{a} = (\tilde{a}_i)_{i \in I}$ are then revealed to all players.

After candidates choose their positions, but before voters vote, each voter chooses a late effort level $\bar{a}_i \in \bar{A}_i$ (compact and convex), also regarding how informed he will be about candidate positions. Nature then gives each voter a noisy signal $s_i \in S_i$ about candidate positions, and each voter then chooses a vote $v_i(s_i)$. This signal process will now be explained in more detail.

3.2.2 Information

Most real politicians are so much better informed about politics than most voters that it seems reasonable to consider a first approximation where each candidate knows everything voters know, and everything other the candidates know, but where voters may know much less.

Thus this game models voter ignorance, but not asymmetric candidate ignorance. Specifically, each voter i only observes a noisy signal $s_i \in S_i$ (compact and convex) before he votes, instead of directly observing candidate positions x_c. The total signal $s = (s_i)_{i \in I}$ is distributed according to a c.d.f.

$$s \sim H(s|x_1, x_{-1}, \tilde{a}, \bar{a})$$

continuous in $x_1, x_{-1}, \tilde{a}, \bar{a}$, where $\bar{a} = (\bar{a}_i)_i$. Again, $\tilde{a}_i$ and $\bar{a}_i$ denote voter i's early and late efforts, respectively, to become informed about candidate positions.

The information structure of the game is encoded in the following notation of the strategies of each player at each choice point. First, all voters choose early efforts $\tilde{a}_i$. Then the two candidates simultaneously choose distributions $\mathcal{F}_c(\tilde{a})$ over x_c, while all voters simultaneously choose their late efforts level $\bar{a}_i(\tilde{a})$. Next, nature samples

$x_c \sim \mathcal{F}_c$ and then $s \sim H(s|x_1, x_{-1}, \tilde{a}, \bar{a})$. Finally, voters choose $v_i(s_i, \tilde{a}, \bar{a}_i)$ and nature samples $\epsilon \sim N(\epsilon)$. After the election, the winner $\hat{c}(v)$ implements their policy $x_{\hat{c}}$.

3.2.3 Preferences

Each candidate c's utility $W_c(v)$ depends in general on the total ballot v. Each candidate c prefers more votes, so $cW_c(v)$ is non-decreasing in each v_i.

Each voter i cares about the winning position $x_{\hat{c}}$, and their efforts, $\tilde{a}_i$ and $\bar{a}_i$, to become informed. Specifically, each voter i gets utility

$$u_i(x_{\hat{c}g(i)}, \tilde{a}_i, \bar{a}_i)$$

which is continuous, non-decreasing, and strictly concave in its first argument, x, and continuous in the remaining arguments. Note that this utility form $u_i(x, \tilde{a}, \bar{a})$ can express non-electoral advantages of political information, such as in performing one's job or in making better consumer decisions.

3.2.4 Expected Payoffs

Putting it all together, we can think of this as the three "stage" game in Figure 1.

In the first stage, each voter knows only the game form, and chooses early effort $\tilde{a}_i$ to maximize

$$\bar{U} = E_{J(\mathcal{F}_1(\tilde{a}), \mathcal{F}_{-1}(\tilde{a}), \tilde{a}, \bar{a})}\left[u(x_{\hat{c}(v(s))g(i)}, \tilde{a}_i, \bar{a}_i)\right]$$

where $v(s) = (\epsilon, (v_i(s_i, \bar{a}_i, \tilde{a}))_{i \in I})$, and where we define the joint distribution J as

$$dJ(\epsilon, s, x_1, x_{-1}|\mathcal{F}_1, \mathcal{F}_{-1}, \tilde{a}, \bar{a}) = dN(\epsilon)dH(s|x_1, x_{-1}, \tilde{a}, \bar{a})d\mathcal{F}_1(x_1)d\mathcal{F}_{-1}(x_{-1}).$$

In the second stage, each voter i knows all early efforts $\tilde{a}$, and chooses late effort $\bar{a}_i$ to maximize this same $\bar{U}$ form. The difference is that a voter knows that later-stage strategies $v_i, \mathcal{F}_c$ of other players can depend on his first-stage choice of $\tilde{a}_i$, but they cannot depend on his second-stage choice of $\bar{a}_i$.

In the second stage, each candidate c also knows only $\tilde{a}$, and chooses distribution $\mathcal{F}_c$ to maximize

$$\overline{W}_c = E_{J(\mathcal{F}_1, \mathcal{F}_{-1}, \tilde{a}, \bar{a})}[W_c(v(s))]$$

subject to M constraints of the form $E_{\mathcal{F}_c}[f_{mc}(x_c)] \leq K_{mc}$.

In the third and last stage, each voter i, knowing his signal s_i, his late effort $\bar{a}_i$, and all early efforts $\tilde{a}$, chooses strategy $v_i \in C$ to maximize

$$\bar{u}_i = E_J\left[u_i(x_{\hat{c}(v_i, v_{-i}(s))g(i)}, \tilde{a}_i, \bar{a}_i)\middle| s_i\right]$$

where $v_{-i} = v \backslash v_i = (\epsilon, (v_{i'})_{i' \in N_i})$.

Since $\tilde{a}$ is told to all after the first stage, the game starting with the second stage is a proper subgame of the whole game. We will call this subgame the *late game.*

3.3 Existence of Equilibria

Before considering this game further, let us consider whether equilibria exist.

If S_i is not finite, then let a finite approximation to S be a finite partition $\hat{S}_i$ of S_i, which implies a discrete distribution $\hat{h}(\hat{s}) = \int_{s \in \hat{s}} dH(s)$ for each $\hat{s} \in \hat{S}_i$ (suppressing arguments $x_1, x_{-1}, \tilde{a}, \bar{a}$). To use such an approximation, one replaces integrals $\int_{s \in S_i} dH(s)$ with sums $\sum_{\hat{s} \in \hat{S}_i} \hat{h}(\hat{s})$ in the expressions for $\bar{u}, \bar{U}, \overline{W}$.

We can prove the following result (see section 3.9.1).

Theorem 3.1 *For any finite approximation to S, there exists a mixed-strategy sequential equilibrium to the late game.*

Our analysis will be easier, however, if we can find equilibria where candidates *treat each voter group independently,* so that $\mathcal{F}_c(x_c) = \prod_g F_{cg}(x_{cg})$. Defining *group* *signals* $s_g = (s_i)_{i \in g}$ and *group votes* $v_g = (v_i)_{i \in g}$, such independent equilibria can be found if we assume

1. *group-independent signals*, so that $H(s|x_1, x_{-1}, \tilde{a}, \bar{a}) = \prod_g H_g(s_g|x_{1g}, x_{-1g}, \tilde{a}, \bar{a})$,

2. *group-linear candidate preferences*, as in $W_c(v) = W_{c0}(\epsilon) + \sum_g W_{cg}(v_g)$, and

3. *group-linear distribution constraints*, so $f_{mc}(x_c) = \sum_g f_{mcg}(x_{cg})$ and $\mathcal{X}_c = \times_g \mathcal{X}_{cg}$, with f_{mcg} continuous and $\mathcal{X}_{cg}$ compact and convex.

For example, group variances might, given constants $K_c, k_{cg}, \bar{x}_{cg}$, be constrained via

$$\sum k_{cg} \mathrm{Var}_{\mathcal{F}_c}[x_{cg}] \leq K_c - \sum_g k_{cg}(E_{\mathcal{F}_c}[x_{cg}] - \bar{x}_{cg})^2.$$

We can prove the following (see section 3.9.2).

Theorem 3.2 *If signals are group-independent and candidate preferences and distribution constraints are group-linear, then there exists a mixed-strategy equilibrium to the late game where candidates treat groups independently.*

The assumptions of this theorem are not that unreasonable for independent equilibria.

For independent equilibria, group should strive to avoid having their signal of candidate positions toward them mixed up with candidate positions regarding other groups, since those other positions are completely irrelevant to their choice. That is, groups should strive for group-independent signals.

Regarding the assumption of group-linear candidate preferences, all we really need is group-linearity of the expected candidate payoff, as in

$$0 = \frac{\partial^2}{\partial x_{cg'} \partial x_{cg''}} \left(\int_s W_c(v(s)) dN(\epsilon) \prod_g \int_{x_{-cg}} dH_g(s_g|x_{1g}, x_{-1g}, \tilde{a}, \bar{a}) dF_{-cg}(x_{-cg}) \right).$$

This should be approximately true, for example, for in an independent equilibrium with $|G|$ large and W_c a smooth function of total votes V. With fractionally-small groups, helping out one group should keep V with a range where W_c is approximately linear, and so not much effect the benefits of helping out other groups.

Independent equilibria with fractionally-small groups also provide a reason to be less wary of assuming group-independent distribution constraints, and more generally of using distribution constraints to constrain candidate policy positions. That is because with enough small groups, the only binding constraint should be an average pie-size constraint, which is group-independent.

In an independent equilibria, the total variance the total pie $\sum_g n_g x_{cg}$ is the sum of the variances of the group distributions F_{cg}, each of which depends mainly on within-group properties. As we get more fractionally smaller groups, the total fractional variance declines to zero, and so reasonable distribution constraints on the variance should be slack. Similarly, with enough groups finite higher-moment group distributions F_{cg} should also make reasonable higher-moment distribution constraints slack. Thus in the limit of many groups, only a simple average-pie-size distribution constraint (such as equation 3.2) should bind.

We can thus think of this model as an exact description of a finite-group approximation to the asymptotic case where variance goes to zero with enough small voter groups, and budgets can be balanced exactly even with independent equilibria [Mye93].

The big advantage of dealing with group independent equilibria is their simplicity. For example, modulo a few coordinating parameters, candidates can be thought of as playing many independent games, one for each group. Let us define each candidate's *group vote payoff* to be the expectation over any vote or effort strategy mixtures of

$$Q_{cg}(x_{cg}, x_{-cg}|\tilde{a}, \bar{a}) = \int_{s_g} W_{cg}(v_g(s_g))dH_g(s_g|x_{1g}, x_{-1g}, \tilde{a}, \bar{a}),$$

which is the convolution of voting strategies, vote payoffs and the voter signal process.

We can prove the following (see section 3.9.3).

Theorem 3.3 *An independent equilibrium has the same candidate strategies as a set of equilibria of one-group candidate games where, given fixed constants $(\lambda_{cm})_m$, each* *candidate c's net payoff is*

$$Q_{cg}(x_{cg}, x_{-cg}) - \sum_m \lambda_{cm} f_{mcg}(x_{cg}).$$

3.4 Rationalizing the Policy Spaces

We have so far described candidates strategy spaces very abstractly; candidates choose distributions $\mathcal{F}_c$ over divide-a-pie positions $x_c = (x_{cg})_g$, positions which are constrained to lie within a convex set defined by a finite set of bounding planes given by inequalities of the form $E_{\mathcal{F}_c}[f_{mc}(x_c)] \leq K_{mc}$. But is this very abstract characterization of strategy spaces consistent with any more concrete models of candidate policy spaces?

Imagine that each candidate could, on "policy day", choose from among a finite set of policy positions $\alpha \in A_c$, and that between policy day and election day, information would come out about the consequences of these positions for post-election group pie-slices x_g which would result from policy α. In this case, each policy-day position α would be associated with a distribution $\mathcal{F}_\alpha$ over election-day positions $(x_g)_g$.

Candidates able to choose mixed strategies over policy-day positions α could thus choose any mixture $\mathcal{F}_c = \sum_{\alpha \in A_c} \pi_{c\alpha} \mathcal{F}_\alpha$ where $\sum_{\alpha \in A_c} \pi_{c\alpha} = 1$. Given a finite set A_c of possible policy-day positions α, the set of possible mixtures $\mathcal{F}_c$ would be a convex hull with a finite set of vertices $\mathcal{F}_\alpha$, and also a finite set of bounding surfaces. Thus in this model candidate strategy spaces could as well be described by a finite set of inequalities, one for each bounding plane. So the abstract formulation of this paper would apply to this more specific model.

An alternative concrete model is as follows. Let each candidate choose both a distributive policy $d_c = (d_{cg})_g$, where $\sum_g d_{cg} = 1$, and a macroeconomic policy m_c. Let each pair of such policies d, m be associated with a distribution $\Psi_m(X|d)$ over the total pie size X. Given a realized total pie X, each group would get a pie-slice given by $x_g = X\delta_g(X, d)$ for continuous pie fraction functions $\delta_g(X, d)$ where $(\delta_g(X, \cdot))_g$ is a one-to-one and onto function on the $|G|$ dimensional simplex. Note that for fixed distributive policies d we allow the pie fractions δ_g to vary with the total pie X, but we only allow macro policy m to influence outcomes x via the total pie X.

Given these assumptions, every outcome $x = (x_g)_g$ (where all $x_g \geq 0$) is associated with a unique pair d, X such that $x = (X\delta_g(X, d))_g$. Thus any distribution $\mathcal{F}(x)$ over election-day positions x can be associated with a distribution $D(d)$ over distributive policies d together with a set of total-pie distributions $\Psi(X|d)$, one for each distributive policy d. Thus if there exists a wide enough range of macro policies m, so that for each d there exists a mixture $\pi(m|d)$ such that $\Psi(X|d) = \int \Psi_m(X|d)\, d\pi(m|d)$, then a distribution $\mathcal{F}$ can be implemented by a mixed-strategy randomization $D(d)$ over distributive policies d, followed by a conditional mixed-strategy randomization $\pi(m|d)$ over macro policies m.

Imagine, for example, that a presidential candidate could, on policy-day, commit to the following deal with a foreign power. In exchange for other concessions, the candidate would if elected arrange for his nation to buy from that foreign power some of a certain product at a certain price. This product might, for example, be a foreign currency. If the foreign power was risk-neutral, if the product was resellable, and if the agreed on purchase amount could vary arbitrarily with the later exogenous market price of the product, then the candidate could in effect agree to any constant-expected-value conditional bet of his nation's total assets. If there were no other sources of total-pie uncertainty than the future market price of this product, then the candidate could obtain any distribution over X which satisfied $E[X|d] \leq \bar{X}(d)$ and $\text{Prob}[X < 0] = 0$.

If we further assumed that the average total pie satisfied $\bar{X}(d) = \bar{X}$, being independent of distributive policies d, then the candidate could in effect commit to any distribution $\mathcal{F}$ subject to an average-pie-size distribution constraint $E_{\mathcal{F}}[X] \leq \bar{X}$. This space of distributions $\mathcal{F}$ includes many group-independent distributions where $\mathcal{F} = \prod_g F_g$.

3.5 Extending the Model

It turns out that this model can be straight-forwardly extended to endogenize the spaces of possible pie-divisions $\mathcal{X}_c$, and to allow candidates to take positions on more general policy issues, in addition to positions on distributive questions. In this extended model, all our previous results will still hold.

Let candidates now take general policy positions $z_c \in Z_c$ (compact and convex) in addition to choosing distributions over divide-the-pie positions $y_c = (y_{cg})_g \in [\underline{y}, \overline{y}]^{|G|}$, subject to only an average pie-size distribution constraint, $E_{\mathcal{F}_c}\left[\sum_g n_g y_{cg}\right] \leq \bar{Y}_c$. (Recall that this is the only constraint expected to bind in the limit of a large number of groups.)

Voter utility will now be of the form

$$u_i(\mu_g(z_{\hat{c}}) + y_{\hat{c}g}, \tilde{a}_i, \bar{a}_i),$$

with μ_g continuous, except that after the winning candidate is chosen, each voter will be given the option to "revolt." If all of the members of any group revolt, they will all be guaranteed $u_i(0, \tilde{a}_i, \bar{a}_i)$. Note that while we have made the strong assumption that voter preferences are linearly separable in general policy and group pie-slices, the choice of zero for the voter revolt payoff is without loss of generality, since utility u_i is specified only up to an affine transformation.

If the winning candidate ever has $\mu_g(z_c) + y_{cg} < 0$ for some group g, then that group will clearly revolt. If candidates sufficiently abhor the prospect of revolution then they will not choose any mixed strategy which gives a finite probability that their positions z, y will violate this constraint. Thus if we define $x_{cg} = \mu_g(z_c) + y_{cg}$ we are assured that all x_{cg} chosen will be non-negative, if such a choice exists. Thus voter utility will again be of the previous standard form $u_i(x_g, \tilde{a}_i, \bar{a}_i)$.

The single distribution constraint can be rewritten in terms of the x_{cg} as

$$E\left[\sum_g n_g x_{cg}\right] \leq \bar{X}(z_c) = \bar{Y}_c + E\left[\sum_g n_g \mu_g(z_c)\right].$$

If we further assume *utility signals*, so that $H(s|y_1, y_{-1}, z_1, z_{-1}, \tilde{a}, \bar{a}) = H(s|x_1, x_{-1}, \tilde{a}, \bar{a})$, with voters only receiving signals about their utility-relevant package $x_{cg(i)}$, instead of about the individual components $y_{cg(i)}$ and z_c, then $\bar{X}(z_c)$ in the distribution constraint is the only place where the z_c still appear in our reformulation of this game.

Thus if this distribution constraint is binding, then in any equilibrium each candidate must choose a z_c^* which maximizes $\bar{X}(z_c)$ over Z_c. Thus, taking $\bar{X}_c = \bar{X}_c(z_c^*)$, this extended game has reduced to the unextended game, inducing the same distributions over the x_{cg}.

Result 3.1 *Allowing general policy positions in addition to divide-a-pie positions does not change the results of this paper if voter preferences are linearly separable in these two kinds of policies, if voters receive only utility signals, and if only an average pie-size distribution constraint is binding.*

3.6 Model Applications

3.6.1 Simplified First-Order Conditions

We can gain a clearer understanding of the conflicting influences on voter information effort by examining equilibrium first-order conditions regarding those efforts. But first we should simplify the model to make these conditions more transparent.

Let us make the assumptions required for independent equilibria, let us further assume identical candidate abilities $\bar{X}_c$, $f_{mcg}(\cdot)$, K_{mc}, $\mathcal{X}_c$, and let us focus on the symmetric strategies $\mathcal{F}_c = \mathcal{F}$ $(= \prod_g F_g)$ which should exist in this now candidate-symmetric game. Let us assume that for all $i, j \in g, u_i = u_j$, so all voters in a group have the same risk preferences and cost of information. Furthermore, let us assume *group-independent information production*, so that

$$H_g(s_g|x_{1g}, x_{-1g}, \tilde{a}, \bar{a}) = H_g(s_g|x_{1g}, x_{-1g}, \tilde{a}_g, \bar{a}_g)$$

where *group efforts* are $\tilde{a}_g = (\tilde{a}_i)_{i \in g}$ and $\bar{a}_g = (\bar{a}_i)_{i \in g}$. Also assume *group information* *pooling*, where all group members $i \in g$ become identically informed about candidate positions, with $s_i = s_j$ when $g(i) = g(j)$, even though their information efforts vary. Finally, assume pure strategy votes and effort choices.

With identical group members, group information pooling, and pure-strategy votes and late efforts, the entire group will either choose one candidate or the other, and so candidate group vote payoff must be of the form $Q_{cg}(x_{1g}, x_{-1g}) = \nu_{cg} q_g(x_{1g}, x_{-1g})$, where $q_g + \frac{1}{2}$ is the probability that voter $i \in g$ will vote for candidate 1, given by

$$q_g(x_{1g}, x_{-1g}, \tilde{a}_g, \bar{a}_g) = \frac{1}{2} \int v_i(s_i) dH_g((s_i)_{i \in g} | x_{1g}, x_{-1g}, \tilde{a}_g, \bar{a}_g).$$

Since we have independent F_g and H_g here, the x_{cg} are independent as well, and so knowing the offers, signals, or efforts of other groups tells a voter nothing about how he should vote. (So there is no "swing voter's curse" here [FP96].) Thus, defining $\mathcal{P}_g(V_{-g})$ to be the c.d.f. distribution over other-group votes $V_{-g} = V - \sum_{i \in g} w_i v_i$, and defining $w_g = \sum_{i \in g} w_i$, we can refer to the probability that some voter in g is pivotal in this election as $p_g = \mathcal{P}_g[w_g] - \mathcal{P}_g[-w_g]$, independent of the late efforts $\bar{a}_g$, signal s_g, and offers x_{1g}, x_{-1g} of this group g.

Given all this, expected voter utility is, for $i \in g$,

$$\bar{U}_i = \int_x u_i(x, \tilde{a}_i, \bar{a}_i) dF_g(x|\tilde{a}) + 2p_g(\tilde{a}) \int_x u_i(x, \tilde{a}_i, \bar{a}_i) \int_y q_g(x, y, \tilde{a}_g, \bar{a}_g) dF_g(y|\tilde{a}) dF_g(x|\tilde{a}). \tag{3.3}$$

Assuming an interior solution and differentiability, the first-order condition for maximizing late effort $\bar{a}_i$ is (suppressing obvious g, a notation)

$$0 = \frac{dU}{d\bar{a}_i} = \int \frac{\partial u_x}{\partial \bar{a}_i} dF_x + 2p \int u_x \int \frac{\partial q_{xy}}{\partial \bar{a}_i} dF_y dF_x + 2p \int \frac{\partial u_x}{\partial \bar{a}_i} \int q_{xy} dF_y dF_x$$

and for maximizing early effort $\tilde{a}_i$ is

$$0 = \frac{dU}{d\tilde{a}_i} = \int \frac{\partial u_x}{\partial \tilde{a}_i} dF_x + \int u_x \frac{\partial (dF)}{\partial \tilde{a}_i} + 2 \frac{d}{d\tilde{a}_i} \left(p \int u_x \int q_{xy} dF_y dF_x \right),$$

where we've compacted the notation, writing $u_x = u(x), F_x = F(x)$ and $q_{xy} = q(x, y)$. In both expressions, the left-most term is the expected marginal cost of effort, and for p small enough the right-most term can be neglected compared to the other terms.

The choice of late effort $\bar{a}_i$ trades off marginal cost of effort against the marginal value of $\bar{a}_i$ in improving the choice of candidate when the voter's group is pivotal, discounted by the probability of being pivotal. The choice of early effort $\tilde{a}_i$, however, trades off marginal cost of effort against the marginal value of influence over candidate strategies, an influence *not* diluted by the probability of being pivotal.

There is, however, a common pool problem here within each group regarding both efforts $\tilde{a}_i, \bar{a}_i$; each group member trades off the personal cost of these efforts with the group benefits these efforts produce in the group terms F_g and $q_g(x, y)$.

Thus, as claimed:

Result 3.2 *In a symmetric independent equilibrium with group-independent information production, group information pooling, and identical group members, only late, not early, voter information efforts are diluted by the group probability of being pivotal. Both efforts, however, can be diluted by a group common pool problem.*

Note that if we modeled the incentives to vote, instead of to become informed, we should find that because the effort required to actually vote cannot be pooled the way information can, a voter's incentive to actually vote is diluted by the probability that his vote is pivotal, not by the probability that someone in his group will be pivotal. On the other hand, there is no group common-pool problem with voting.

3.6.2 A Parameterized Example

Let us now examine an instance of this last model, using specific functional forms to allow us to examine closed-form expressions for equilibrium strategies.

Assume H_g is such that each voter group g either learns both candidate positions x_{cg} exactly, with probability α_g, or it learns nothing. Let the production of information α_g from efforts $\tilde{a}_i, \bar{a}_i \in [0, 1]$ be given by

$$\alpha_g = \bar{\alpha}_g^{\bar{\delta}_g} \tilde{\alpha}_g^{\tilde{\delta}_g}.$$

This describes voters who need a conjunction of both early information $\tilde{\alpha}_g$ and late information $\bar{\alpha}_g$ to learn about candidate positions. For $\delta_g < 1$ there are diminishing returns to estimating candidate positions from the information available.

Groups jointly produce early and late information according to

$$1 - \tilde{\alpha}_g = \prod_{i \in g} (1 - \tilde{a}_i)^{\kappa_g} \qquad 1 - \bar{\alpha}_g = \prod_{i \in g} (1 - \bar{a}_i)^{\kappa_g}.$$

where group ignorance $1 - \alpha_g$ is a product of individual ignorance $1 - a_i$. This form of the information (and communication) production function would allow a small subset of the group to produce all of the group information if they wanted, by eliminating their own ignorance.

The coordination factor κ describes possible economies of scale in group information production. There are *no scale economies* when $\kappa = 1/n$, since here identical individuals are in the same situation as they would be in groups of size one. For $\kappa = 1$, in contrast, there is *zero marginal cost* to including and informing a larger group, even if the added members contribute nothing to information production

Regarding preferences, let voter utility be a "Cobb-Douglas" power law

$$u_i = x_g^{\beta_g} (1 - \bar{a}_i)^{\bar{\gamma}_g} (1 - \tilde{a}_i)^{\tilde{\gamma}_g}$$

where all $\delta, \beta, \gamma, \kappa > 0$, so that voters prefer *information leisure* (equals ignorance) $1 - a_i$. Let candidates seek plurality $W_c = cV$, where $w_g > 0$ for all g.

Finally, assume only an average pie-size constraint (so $M = 1$); candidates choose independent F_g on $x_{cg} \geq 0$ constrained only by $\sum_g n_g \hat{x}_{cg} \leq \bar{X}$, where $\hat{x}_{cg} = E_{F_g}[x_{cg}]$.

A little algebra reveals that a sufficient condition for this model to have a unique interior independent equilibrium is $\delta, \gamma, \kappa \leq 1$, $\tilde{\gamma} \geq 1.5\beta\tilde{\delta}(1 - \tilde{\kappa})$ and $\bar{\gamma} \geq .5\beta\bar{\delta}(1 - \bar{\kappa})$

for all groups. In this equilibrium, candidate distributions are uniform on $[0, 2\hat{x}_{cg}]$, and divide the expected pie according to vote-weighted probabilities of being informed[1], as in

$$\frac{n_g \hat{x}_{cg}}{\bar{X}} = \frac{w_g \alpha_g}{\sum_{g'} w_{g'} \alpha_{g'}}.$$

Voters choose their efforts to satisfy (suppressing g subscripts):

$$\frac{\tilde{\alpha}}{1-\tilde{\alpha}} = \frac{\tilde{\kappa}\tilde{\delta}\beta}{\tilde{\gamma}}\left(\frac{p\alpha}{2+\beta(1+p\alpha)} + 1 - \frac{w\alpha}{\sum_g w_g \alpha_g}\right), \quad (3.4)$$

$$\frac{\bar{\alpha}}{1-\bar{\alpha}} = \frac{\bar{\kappa}\bar{\delta}\beta}{\bar{\gamma}}\left(\frac{p\alpha}{2+\beta(1+p\alpha)}\right). \quad (3.5)$$

In the limit of *low information levels* $\tilde{\alpha}, \bar{\alpha} \ll 1$ and *small groups* $n_g \hat{x}_{cg} \ll \bar{X}$, we get the approximation

$$\alpha^{1-\bar{\delta}} \approx \left(\frac{p}{(2+\beta)}\frac{\beta\bar{\kappa}\bar{\delta}}{\bar{\gamma}}\right)^{\bar{\delta}} \left(\frac{\beta\tilde{\kappa}\tilde{\delta}}{\tilde{\gamma}}\right)^{\tilde{\delta}}. \quad (3.6)$$

Examining this expression, we can conclude the following.

Result 3.3 *With low information levels, small groups, and decreasing returns to late information ($\bar{\delta} < 1$) in this parameterized example, group information levels, and hence mean candidate offers to the group, increase with economies of scale in group information production κ, and with the group probability of being pivotal p. Information levels decrease with the personal costs of information γ, and with voter risk-aversion $1/\beta$.*

Note that group information levels α_g depends on group size n_g through *both* decisiveness p_g, which one expects should be approximately linear in w_g for a small group, and through the coordination factors κ_g, which should typically decline with group size n_g. If, for example, we assume that for n small, $p = kn$, $\bar{\delta} = \tilde{\delta}$, and

[1] This equilibrium has the same marginals as the related Colonel Blotto equilibria [GW50].

$\bar{\kappa} = \tilde{\kappa} = k'n^{-\tau}$, then for $\tau < 1/2$ the equilibrium group information level α *increases* with group size.

A perhaps clearer, though more extreme, example is where $\bar{\kappa} = \tilde{\kappa} = 1$ in equations 3.4 and 3.5, in which case a larger group is clearly better informed, since p_g must be increasing in w_g. Note also that even if group information levels stay constant as group size increases, larger groups still have higher member utilities for $n\kappa > 1$, since in equilibrium $(1 - \bar{\alpha}_{g(i)}) = (1 - \bar{a}_i)^{n\kappa}$ and similarly for early efforts. With a larger group, each member need contribute less to information production to get the same total pooled information. Thus:

Result 3.4 *If group scale economies in information production are large enough, and if late information efforts are important, members of larger groups will be better informed and better off.*

3.6.3 Ignorance Can Be Bliss

The following model shows how voters can strongly prefer to be ignorant in certain ways about candidate positions.

Assume again group-independent information production, make the assumptions required for independent equilibria, assume the only binding distribution constraint is the average pie-size one (equation 3.2), and assume symmetric candidates who take symmetric strategies $F_{cg} = F_g$.

Assume further that for some group g we effectively have $|\bar{A}_i| = 1$ for all $i \in g$, so there are no late effort alternatives. This captures the idea that there is some small minimum effort required to obtain any late information, making effort $\bar{a}_i$ so costly relative to a voter's tiny incentive to obtain it that in equilibrium there is no temptation to make late efforts. Let us also assume that early effort is completely costless, so that $u_i(x, \tilde{a}_i, \bar{a}_i) = u_i(x)$ for all $\tilde{a}_i, \bar{a}_i$ and $i \in g$. Finally, assume group information pooling.

Now define a *negative news* signal distribution as $H_g(s_g|x_{1g}, x_{-1g}) = \prod_c H_{cg}(s_{cg}|x_{cg})$ where each c.d.f. $H_{cg}(s|x) = s/2x$. That is, each voter get two independent signals

$s_c \in R_+$, one from each candidate, such that given a candidate offer x the signal s is distributed uniformly on the interval $[0, 2x]$. For small x, or "negative news," this signal distribution is more concentrated, and hence this distribution is especially informative about such situations. While good candidates can appear bad, bad candidates cannot appear good.

Since this negative news distribution has a monotone likelihood ratio, it induces a voting strategy of $v_i = 1$ if $s_1 > s_{-1}$, and 0 otherwise, which implies a voting probability

$$\frac{1}{2} + q(x,y) = \begin{cases} x/2y & x \leq y \\ 1 - y/2x & x \geq y \end{cases} . \tag{3.7}$$

This function is concave in x and induces a unique pure strategy candidate equilibrium. We can also prove the following (see section 3.9.4).

Theorem 3.4 *If, for some group, early information efforts are costless, and late efforts effectively cannot vary, then if there are some early group efforts which induce a negative news group signal distribution, it is a group-Pareto equilibrium for group members to choose those efforts. In particular, these strictly dominate any group efforts which induce fully informative signals.*

This negative-news signal distribution induces the same expected offer x_{cg} as the fully informed case. The difference is that it induces a pure-strategy response, instead of the variance associated with the candidate response to a full-informed group. Thus a switch from full-information to negative-news signals can help this group without hurting any other groups; optimal ignorance can in principle produce large welfare gains by stabilizing the political system.

In this model, a candidate, considering proposing some policy change, knows that the voters who would be hurt by that change are more likely to find out about it than voters who would be helped, and is thus discouraged from proposing the change. The tendency for people to more easily join an interest group to stop a bad change than to effect a good one [Han85] is perhaps weak support for this model.

Note that the model in this section can apply to any situation in which someone with an indivisible prize to award wants to induce two suitors to both curry as much favor as possible, and where this prize-holder will want to award it to the suitor who did the most currying. Imagine an older person with a valuable inheritance, seeking more attention from his children. Such a person can prefer ahead of time to structure his system of information so that he will not learn too much about how much each of his suitors has favored him. Knowing about this ignorance, his suitors may then be induced to reduce the variance of their choices.

3.7 Voting Lotteries

Let us make the assumptions necessary for group-independent equilibria, and consider again the simplified model of section 3.6.1.

Imagine that after candidates take their positions, but before voters take late efforts, the voting weight w_g for some group g is subject to a fair lottery, so that with probability $1/b_g$ group g will have $b_g w_g$ votes, and otherwise it will have zero votes.

This might happen, for example, if random juries were used to decide elections, and if no more than one juror could come from any one group. Alternatively, voters might be allowed to join voting "pools" where all pool votes are randomly given to one pool member, and all members of g might join the same pool. Or perhaps voters could just directly gamble their vote at fair odds with an electoral agency.

If we assume that $Q_{cg} = w_g q_g$, so that candidate utility is linear in the group's voting weight w_g, then candidate behavior will not be directly sensitive to the lottery size b_g, since the expected voting weight is independent of this size.

Voter information efforts will, however, be influenced by the vote lottery. Voters will still maximize an expected utility $\bar{U}$ of the form of equation 3.3, but the decisiveness term $p_g(b_g)$ will change. When choosing $\bar{a}_i$, group members will, instead of considering $p_g(1)$ as they would without the lottery, now consider $p_g(0) = 0$ when their group loses the voting lottery, and $\bar{p} = p_g(b_g)$ when they win. And when choosing $\tilde{a}_i$, group members will consider the expected value $\tilde{p} = p_g(b_g)/b_g$. In all cases the relevant function $q_g(x, y)$ will be the one when the voter has won the lottery.

The calculation of $\bar{p} = p_g(b_g)$ should of course take into account the fact that winning the lottery might tell one something about the distribution of votes of other groups. But since in equilibrium the candidates still treat groups independently, this $\bar{p}$ remains independent of this group's efforts, signal, or offers.

For simplicity, let us now consider the case of *group-independent lotteries*, where knowing whether a group won or lost their lottery says absolutely nothing about other group weights. In this case $p_g(b_g) = \mathcal{P}_g[b_g w_g] - \mathcal{P}_g[-b_g w_g]$, and $\bar{p} = p_g(b_g)$ is clearly increasing in b_g. If we further assume a *uni-modal* vote distribution, with the p.d.f $\mathcal{P}'$ being maximal at zero (since candidates are symmetric) and decreasing away from zero, then $\tilde{p}$ should be decreasing in b_g, but decreasing only slightly for $b_g w_g$ small.

Substituting $\tilde{p}$ into equation 3.4, and $\bar{p}$ into equations 3.5 and 3.6, notice that a small $\bar{\alpha}$ increases almost linearly in $\bar{p}$ for $b_g w_g$ small, while $\tilde{a}$ depends only sightly on a small $\tilde{p}$. Thus:

Result 3.5 *In the parameterized example with a uni-modal vote distribution, late efforts are strongly increasing in the lottery size, while early efforts are only very slightly decreasing.*

In section 3.6.3, we assumed that some voting group had zero early effort costs, and had late efforts effectively prohibited, perhaps by weak incentives and some small fixed cost. We showed such a group would prefer early efforts that produced a form of ignorance which induced the best possible candidate offers.

Once candidates have taken their positions, however, such voters would have no such strategic incentive to decline low cost information about candidate positions. But such a temptation could lead candidates to expect voter's signals would deviate from the negative-news distribution on election day, inducing worse candidate offers. Thus a voter might regret the option to have stronger incentives to make late efforts, and hence may regret the option to gamble their vote.

For example, imagine that there were only two possible late efforts, a higher effort which made one's group fully informed, and a lower effort which gave no information.

If there was a b_g large enough to make $p_g(b_g)$ large enough that a voter would want to choose the higher effort if they won the lottery, then once candidate choices are made, this voter could not resist the temptation to play this lottery, even if he would, before candidates take their positions, regret having this lottery option. Note also that both effort choices result in the same expected candidate offers, and so other groups remain unaffected by this choice. Thus:

Result 3.6 *A voter group could prefer the option to visibly commit, before candidates take their positions, to not use a voting lottery. Allowing this commitment can aid this group without harming any other group.*

3.8 Discussion

A fully-game-theoretic model of a two-candidate election has been presented where candidate policy positions are lotteries over election-day positions, and where the space of mixture of such lotteries is delimited by a set of expectation inequalities. In effect, candidates can make small gambles with a distributive pie, and voters can choose how informed to become about candidate election-day positions.

This model illustrates the important points that 1) early visible voters efforts to become informed are not diluted by the probability of being pivotal, that 2) scale economies in information production can outweigh group free-riding problems, and that 3) voters can prefer ignorance and negative news signals in order to induce political stability.

While Downs seems to have understood this first point about early visible effort, he seems to have mistakenly narrowed its scope of application. I have not seen any other prior discussion of the first and second points, and the third point has only been suggested in the context of less than fully-game-theoretic models.

These insights offer many ways to understand existing political phenomena, and suggest directions for future empirical work.

Writing or calling a representative is easily understood as a direct early signal of voter information, though it does require a non-trivial minimum effort. Education and age, however, should also signal information levels, and are very cheap to signal. And both of these indicators seem to be strongly correlated with political participation [WR80]. Subscribing to a newspaper, magazine, or interest group newsletter also signal future information levels, and require relatively little effort. This approach might even help explain why many people seem so much more interested in politics in general, rather than in learning about specific candidates just before an election.

The models in this paper also suggests that successful and unsuccessful interest groups should differ by more than just factors moderating the free-rider problem, such as group size, fat-cat members, and private benefits of joining formal group organizations. In the models of this paper, the technology of information production and distribution is central to the democratic process. Technological progress in information production, however, may not be the political windfall many hope them to be if strategic ignorance is an important factor. Strategically ignorant voters may not want to learn any more about politics.

Testable implications regarding negative voting and negative news have begun to be explored [FS89], and there is a rich literature on negative voting to draw on.

There are also many directions for further exploration of related formal models. One could try a two-period model with some candidate uncertainty over voter costs and see if voters can use the fact that they voted in the first-period, rather than abstaining, to signal their information level, and so influence second-period candidate offers. Such a model might explain the observed stronger-than-expected voter incentives to vote. This signaling model would be somewhat analogous to a model of Glazer [Gla87] on voting to signal compatibility with colleagues, and less analogous, though still related, to the models of Lohmann [Loh94] on protest signaling.

Rather than requiring candidates to observe credible signals regarding the information levels of each and every voter, one could also try a sampling model where each candidate examines the information levels of some random sample of voters, and extrapolates the sample results to the rest of each voting group. Such a model would plausibly still give voters a strong incentive to become informed, if they did not know whether they would be sampled.

Finally, we most need to model some more direct welfare effects of voter information on the quality of policies adopted by candidates. One presumes that better informed voters can induce candidates to adopt better policies, but we would like to see if this intuition can be realized in a precise model.

3.9 Appendices

3.9.1 Proof of Theorem 3.1

The relevant strategies for the claimed mixed equilibrium are the $\mathcal{F}_c$ distributions over x_c, distributions B_i over the $\bar{a}_i$, and $p_i(\hat{s}) = \text{Prob}[v_i = 1 \mid \hat{s}]$ for all $\hat{s} \in \hat{S}$. The relevant payoffs are the expected approximate payoffs

$$E_{B,p}[\bar{u}], \quad E_{B,p}\left[\bar{U}\right], \quad E_{B,p}\left[\overline{W}\right]$$

where $B = (B_i)_i$ and $p = (p_i)_i$. To show that the claimed equilibrium exists, it is enough to show that these payoff functions are continuous and quasi-concave in these strategies, and that the strategy spaces are convex and compact. (By Berge's maximization theorem, best-reply correspondences which maximize a continuous function over a compact set exist and are upper-hemi-continuous with compact values. Quasi-concavity implies convex best-reply values. By Kakutani's fixed-point-theorem then, a best-reply fixed-point exists [BA94].)

Since all the strategies are distributions, all strategy spaces are convex, and all payoffs are linear and hence quasi-concave in these strategies. Under the weak topology, an unconstrained space of probability distributions is a compact metric space. For the constrained $\mathcal{F}_c$, the M constraints $E_{\mathcal{F}_c}[f_{mc}(x_c)] \leq K_{mc}$ are all inequalities linear in the $\mathcal{F}_c$, and so intersect the compact and convex set of possible $\mathcal{F}$ with half-spaces, leaving a compact convex set of constrained $\mathcal{F}$.

Finally, given the continuity assumptions made on u_i and $\hat{h}$, the expected approximate payoffs are continuous in x_c and $\bar{a}$, and therefore in the strategies $\mathcal{F}_c, B_i, p_i$.

After all, the p_i specify a mixture over a finite set of possible ballots v, expressions of the form $\int f(z)dF(x)$ are continuous in the distribution F when the function f is continuous, and the ratio of two non-zero continuous functions is continuous, as in the form

$$\bar{u}_i = \frac{\int u_i dN dH d\mathcal{F}_c d\mathcal{F}_{-c}}{\int dN dH d\mathcal{F}_c d\mathcal{F}_{-c}}.$$

QED.

3.9.2 Proof of Theorem 3.2

The proof strategy will be to show first that there exists an equilibrium when candidates are constrained to choose independent distributions, and then to show that when candidates are allowed to choose any distribution, they don't have a reason to deviate from this independent equilibrium.

Call $\mathcal{P}(\mathcal{X}_c)$ the space of distributions over x_c, where $\mathcal{X}_c$ is the space of possible x_c. If we constrain each $\mathcal{F}_c$ to satisfy $\mathcal{F}_c(x_c) = \prod_g F_{cg}(x_{cg})$, then the resulting subset of $\mathcal{P}(\mathcal{X}_c)$ is *not* convex under the stochastic combination

$$t\mathcal{F} + (1-t)\mathcal{F}' = t\prod_g F_g + (1-t)\prod_g F'_g.$$

However, if we consider the space of candidate strategies to be $\times_g \mathcal{P}(\mathcal{X}_{cg})$, where $\mathcal{X}_{cg}$ is the space of possible x_{cg}, then this space *is* convex under the combination

$$t\mathcal{F} + (1-t)\mathcal{F}' = \prod_g (tF_g + (1-t)F'_g).$$

Thus we can show existence of equilibria in this constrained strategy space, using the same proof structure as for Theorem 3.1, if we can show that, under this alternative convex combination, the M distribution constraints preserve convexity, and that candidate payoffs are still quasi-concave.

For any function f of the form $f(x) = \sum_g f_g(x_g)$, we have

$$\int f(x)d\mathcal{F}(x) = \int \left(\sum_g f_g(x_g)\right)\left(\prod_g dF_g(x_g)\right) = \sum_g \int f_g(x_g)dF_g(x_g)$$

and so

$$\int f(td\mathcal{F}+(1-t)d\mathcal{F}') = \sum_g \int f_g\left(tdF_g + (1-t)dF_g'\right) = t\left(\sum_g \int f_g dF_g\right)+(1-t)\left(\sum_g \int f_g dF_g'\right)$$

so that f is exactly linear in the mixing parameter t. Thus since the distribution constraints are of this form $\int \sum_g f_g d\mathcal{F} \leq K$, those constraints preserve convexity of the distribution space.

Since candidate payoff is of the form $W_c = W_{c0} + \sum_g W_{cg}$, expected payoff can be rewritten as $\overline{W}_c = \mathcal{W}_{c0} + \sum_g \mathcal{W}_{cg} dF_{cg}$, which is linear, and hence quasi-concave, in such mixtures $\mathcal{F}$. Thus there exists a constrained equilibrium.

To complete our proof, we now need only to show that when the space of strategies allowed for each candidate is expanded to the full $\mathcal{P}(\mathcal{X}_c)$, the same independent distribution still maximizes each candidate's expected payoff, holding the other candidate's strategy fixed at their independent equilibrium strategy.

The expression $\overline{W} = \bar{W}_0 + \int \sum_g \bar{W}_g d\mathcal{F}_c$ is linear in $\mathcal{F}_c$, and hence concave as well. And since we've assumed there is some distribution where no constraint binds, Slater's condition is satisfied. Thus a maximum implies a Lagrangian saddle-point and vice-versa [Fra80].

For the "restricted" case where $\mathcal{F}_c$ must be independent, then in equilibrium $\mathcal{F}_c$ must, for some $(\lambda_m)_m, (\eta_g)_g$, be a saddle-point of the Lagrangian $\mathcal{L} =$

$$\int \sum_g \bar{W}_g d\mathcal{F} + \sum_m \lambda_m \left(K_m - \int \sum_g f_{mg} d\mathcal{F}\right) + \sum_g \eta_g (1 - \int dF_g) \qquad (3.8)$$

(suppressing obvious c notation). For the "unrestricted" case where any correlated $\mathcal{F}_c$ is allowed, the last term is replaced by $\eta_0(1 - \int d\mathcal{F})$.

Note that a solution $\mathcal{F}$ to either of these problems, restricted or unrestricted, will be a solution to the matching "linearized" problem, where the $\sum_m \lambda_m$ part is taken to be part of the payoff function, instead of a constraint multiplier term. Conversely, a solution of either linearized problem with the right λ_m is a solution of the problem it was derived from.

In the linearized expanded problem, the only constraint directly on the distribution is that it be normalized. Thus the added constraint in the linearized non-expanded problem that the distribution $\mathcal{F}_c$ must be independent does not reduce the payoff, since for any solution distribution of either problem, a distribution concentrated at a single point within the support of that solution distribution will do just as well, and is clearly independent. Thus a solution to the linearized non-expanded problem is a solution to the linearized expanded problem, and hence is a solution to the non-linearized expanded problem. Thus there is no incentive to deviate. QED.

3.9.3 Proof of Theorem 3.3

The Lagrangian of equation 3.8 can be rewritten as $\mathcal{L} = \sum_m \lambda_m K_m + \sum_g \mathcal{L}_g$, where $\mathcal{L}_g =$

$$\int \left(\bar{W}_g - \sum_m \lambda_m f_{mg} \right) dF_g.$$

Since the only dependence of $\mathcal{L}$ on F_g is via $\mathcal{L}_g$, a maximum of $\mathcal{L}$ with respect to F_g is a maximum of each component $\mathcal{L}_g$ with respect to F_g, given the λ_m. And since each $\mathcal{L}_g$ is linear in F_g, the support of any distribution F_g which maximizes $\mathcal{L}_g$ must maximize the integrand $\bar{W}_g - \sum_m \lambda_m f_{mg}$. All that remains is to note that, given our definitions, $\bar{W}_g = \int Q_g dF_{-cg}$, and that $\sum_m \lambda_{mc} f_{mcg}(x_{cg})$ is independent of x_{-cg}.

3.9.4 Proof of Theorem 3.4

Since voter utilities $u_i(x_{g(i)})$ are concave in x_g, it is enough to show that, given the vote probability $q(x, y)$ of equation 3.7, the optimal candidate choice is a pure-strategy $x^* = x_{1g} = x_{-1g}$ equal to the maximum possible expected value $\hat{x} = \int x dF(x)$ in any equilibrium. Such an $x^* = \hat{x}$ would then first and second-order stochastically dominate any other equilibria, and hence be a group-Pareto equilibria.

From theorem 3.3 we know that there is some λ such that candidates choose x_{cg} to maximize $cq_g(x_{cg}, x_{-cg}) - \lambda x_{cg}$. Plugging a pure-strategy x^* into first-order conditions, using $q_g(x, y)$ of equation 3.7, we get $x^* = \frac{1}{2\lambda}$. And by the concavity of $q(x, y)$ this is a maximum.

For all x in the support of any F, we have

$$\int q(x, y)\, dF(y) - \lambda x \;\; \geq \;\; \int q(0, y)\, dF(y) \;\; \geq \;\; 0$$

since a maximal x should be at least as good as $x = 0$, and since $q \geq 0$. Integrating this over F, we have

$$\frac{1}{2} - \lambda \int x\, dF(x) \geq 0,$$

since the average vote per candidate is $\frac{1}{2}$ in a symmetric equilibrium. This implies that the maximum expected offer is $\frac{1}{2\lambda}$.

If the group were to be perfectly informed (say with $s_{cg} = x_{cg}$), then candidates would face $\frac{1}{2} + q(x, y) = \text{step}(x - y)$, inducing an equilibrium candidate c.d.f. of $F(x) = \lambda x$, uniform on $[0, \frac{1}{\lambda}]$, with mean $\int x dF = \frac{1}{2\lambda}$. Since this signal distribution has the same mean as the negative news distribution, but has substantial variance, it is dominated by that distribution.

Chapter 4 Adverse Selection and Collective Choice

4.1 Introduction

4.1.1 When Is Intervention Efficient?

When can (coercive) government intervention make an economy more efficient? A standard answer is: when there is "market failure," so that an equilibrium of a feasible coercive mechanism scores higher on some efficiency measure than an equilibrium of the no-intervention situation.

This answer is, however, subject to several familiar caveats. In particular, there is the possibility of "democratic failure" [Wit95]. Even when efficient interventions exist, a political process may actually select an inefficient intervention. Furthermore, a political process which chooses an intervention may constitute an informed principal [MT90], thereby effecting the intervention's equilibrium.

Since this last caveat has, it seems, received no formal attention, this paper examines some consequences of asymmetrically-informed centralized collective choice for economists' second-favorite prototypical situation calling for intervention: adverse selection and other excessive-signaling.[1] Our goal: to clarify the extent to which and the mechanisms by which democratic choice can mitigate adverse selection.

4.1.2 Adverse Selection Is Widely Cited

Soon after the first signaling game models [Spe74, RS76] were developed, researchers noted that pooling equilibria can Pareto-dominate separating equilibria, and that

[1] Economists' favorite prototype of justified intervention is arguably that of public goods and other externalities.

externally-imposed limits on allowed signals can force more efficient pooling.

For example, low-risk insurance customers may offer to buy less than full insurance in order to signal their low-type, and thereby convince insurance companies to offer them lower rates. This can work if high-risk types would rather admit their high type than do without full insurance. If sufficient limits are imposed on allowed signals, however, such as requiring that everyone buy some minimum amount of insurance, then everyone will pay average-type rates for this insurance. This can in some cases make all types better off [AH90].

Economists have long mentioned these facts in a wide range of policy contexts, as support for state-imposed signal limits, taxes, or forced pooling, or at least as support for favoring group insurance and collective-bargaining. For example, various forms of excessive-signaling arguments have been cited in support of taxes on status-signaling luxuries [Ire94], progressive taxation in general [And96], limits on work hours [LRT96], and taxes on job market signals such as education.

More prominently, adverse selection has been central to economist's recent arguments for greater government intervention in the health care market [Dia92]. Employer-based group insurance is said to mitigate adverse selection, and standard public finance texts (such as Hyman's [Hym93]) cite adverse selection as a major explanation for Medicare and government unemployment-insurance. (Curiously, though, the leading political science analysis of the politics of insurance regulation, by Meier [Mei91], doesn't even mention the mitigation of adverse selection among the six main goals of insurance regulation. And the empirical evidence that excessive-signaling is a real insurance problem is weak [Hem92, BD93].)

Finally, adverse selection is the primary formal justification offered in law and economics for limits to freedom of contract regarding "private" affairs which basically affect no one else. (Non-formal justifications offered by legal scholars focus largely on paternalism and irrationality [Tre93].) For example, excessive signaling arguments have been offered in support of liquidated-damages rules [AH90] and inalienable producer product liability [Ord79].

4.1.3 Problems With Collective Choice

Even if efficiency could be improved by forced pooling, signal limits, or signal taxes imposed from the outside, however, there is room for doubt about the efficiency of such signal restrictions chosen by real collective-choice processes involving asymmetrically-informed agents. Similarly, there remain questions about the degree to which adverse selection problems are really reduced when large employers choose group-insurance, or when labor unions choose employee benefits.

In any of these cases of collective choice[2], if a different actual distribution of types in the relevant population would induce a different collective choice for that population, then this choice should serve as a signal to receivers, such as insurance companies, in the remainder of the signaling game. Insurers should, for example, be wary that employers or labor unions with riskier employees will ask for more group insurance.

Similar concerns should in principle exist even with national political choices, such as national health care reform. Instead of signaling directly via their choice of insurance policy, voters might instead signal indirectly via their political choices.

While there have been some recent complete-information models of voting over health care and other public provision of private goods, [Gou93, ER96], I know of no such incomplete information models. And while there is a literature on informed principals in mechanism design, I know of no such models with multiple principals. Thus this issue seems to have not yet been addressed.

4.1.4 Model Overview

To explore this collective signaling effect, I examine the ex-ante Pareto-efficiency of "extreme" models. That is, models in which the adverse selection problem is the most severe, in which voters are otherwise identical, and in which the democratic processes considered are the most benign, minimizing agency costs and revealing the least information to signal receivers. This is an attempt to find the best case for

[2] I take "non-collective choice" to be the case where each agent independently plays their signaling game.

democratic and other collective choice processes to mitigate adverse selection problems. A significant adverse-selection effect in these extreme models should suggest that this problem be taken seriously in less-extreme models as well.

After all, when adverse selection problems are mild, attention should shift to other tradeoffs regarding intervention. And when collective choice has high agency costs, or when voters are very diverse, there are other reasons to avoid collective choice. Finally, the more information that is revealed about the collective choice process, the more opportunities there are to signal via that process.

To model a severe adverse-selection problem, I use a basic insurance model [CL87] with no hidden actions or other allocative benefits of sorting agents, and where agents are distinguished by a continuous one-dimensional risk-type distributed over a wide support. In this insurance model, potential insurees signal their risk level by offering an insurance contract, which insurance companies must then either accept or reject.

To keep the adverse selection problem severe, I also focus on separating equilibria, which are preferred by refinements such as intuitive equilibria [CK87], as opposed to pooling equilibria, which tend to be preferred by refinements such as undefeated equilibria [MOFP93].

The focus is on collective choice of group insurance, where every group member must get the same insurance policy. I also, however, consider a variation where the group chooses a certain minimum insurance level required of each member.

I compare two models of collective choice of group insurance to individual choice of insurance. One model of collective choice has a single full-informed profit-maximizing group owner choosing group insurance. This owner must ensure that each member has at least an individual reservation utility level.

The other model of collective choice uses a direct-democracy process for voting on group insurance. In particular, I consider a median voting rule, where everyone submits a number and the median number is implemented. This median voting model has the same equilibrium outcomes as a two candidate prospective voting model, where the two candidates both know the median type of the electorate, and where insurance companies only observe the position of the winning candidate.

To minimize information revealed by collective choice, the insurance company observes the collective choice itself, but not any of the individual votes that produced this result. Having the insurer observe more process information would allow more opportunities for group members to signal their types, which would presumably allow even more adverse selection.

Finally, to further give the best chance for a collective choice process to successfully mitigate adverse selection problems, voters are assumed identical aside from their insurance risk level, and risks are assumed to be distributed independently across voters.

4.1.5 Results Overview

When group insurance is chosen by a profit-maximizing group owner, and when each member's reservation utility is what he would get buying individual insurance outside the group, then it turns out that group insurance offers no advantages over individual insurance for separating equilibria. Assuming no wealth effects, the utility of insurees is no better, and usually worse.

This is because separating equilibria depend on the support of the distribution of types, and the distribution of group averages has the same support as the distribution of individual types. While the variance of group types is substantially reduced, this is irrelevant in separating equilibria.

Regarding voting, a simple differential equation characterizes the median vote separating equilibrium when voters choose either a single pooled insurance contract, or a common signaling limit, i.e., a minimum insurance requirement. This median vote equation generalizes easily to apply to any "pivotal" collective choice process, in which each choice among the equilibrium alternatives is the favorite choice of some pivotal member of the collective.

In a particular example, a computed median vote separating equilibrium demonstrates explicitly that democratic choice is not always ex-ante efficient, though it can improve on the no-intervention equilibrium. This improvement can be thought of as arising from the fact that democracy can fail to represent the opinions of up to almost half of the electorate. Insurers who know that the median type is the worst possible type, for example, know that half the electorate is also of this worse type, but know nothing about the other half.

This suggests that we might get better outcomes by using collective choice mechanisms which fail to represent even larger fractions of the electorate. For example, if a random "jury" is selected to hold a median vote on signal restrictions, selfishly following its own preferences, this jury should not signal any types but its own. And it turns out that in the limit of large but fractionally small juries, the population gets the favorite insurance amount of the median type, given the mean insurance rate. When the mean equals the median, this implies asymptotic ex-ante efficiency. Thus random juries can eliminate adverse selection problems.

4.2 Individual Insurance

The following is the standard adverse-selection insurance game, widely treated as a prototype of excessive-signaling correctable by external restrictions. We now review standard results for separating equilibria of such games.

4.2.1 The Basic Insurance Signaling Game

This signaling game has two players, a risk-averse insuree desiring insurance against a possible loss L, and a risk-neutral insurance company. (We will denote the insuree and insurer by male and female pronouns respectively.) If x is the insuree's loss when an accident occurs, y is his loss otherwise, and $p(t) = 1/(1+t)$ is his probability of avoiding a loss (with $t = 1/p - 1$ being the insuree's risk type), then the player's expected payoffs are

$$\begin{aligned} \text{insuree:} \quad U(x,y,t) \equiv & \quad (1-p(t))u(-x) + p(t)u(-y)) = \quad p(t)(tu(-x) + u(-y)) \\ \text{insurer:} \quad V(x,y,t) \equiv & \quad (1-p(t))(x-L) + p(t)y = \quad p(t)(t(x-L) + y) \end{aligned}$$

where u is assumed continuously differentiable, strictly increasing, and strictly concave.

The insuree, knowing his risk type t, signals by offering a contract $(x, y) \in R^2$. The insurer, knowing only a prior c.d.f. $F(t)$ on t, and whatever she can infer from the insuree's offer, accepts or rejects that offer. Given rejection, $(x, y) = (L, 0)$.

When $F(t)$ has convex support $[\underline{t}, \bar{t}]$, then separating sequential equilibria exist with just barely accepted offers $(x(t), y(t))$ for $t \in [\underline{t}, \bar{t}]$ satisfying $V(x(t), y(t), t) = 0$, so that $y(t) = y^*(t, x(t))$, where $y^*(t, x) \equiv t(L - x)$. The local incentive compatibility condition for separation is

$$0 = tu'(-x)x'(t) + u'(-y)y'(t). \tag{4.1}$$

An interim-Pareto-dominant separating equilibrium must also satisfy $x'(t) < 0 < y'(t)$ and $x(\bar{t}) = x^*(\bar{t}, \bar{t})$, where we define favorite points as

$$x^*(s, t) \equiv \operatorname{argmax}_x U(x, y^*(s, x), t).$$

(This argmax is unique by the strict concavity of u.) While the worst type $\bar{t}$ gets his favorite level of insurance, in this case being fully insured with $x = y$, better types are less than fully insured with $x > y$.

We will denote this basic equilibrium as $x_1(t)$ from here on, with $y_1(t) = y^*(t, x_1(t))$. Note that the equilibrium menu of x, y choices depends on the distribution F only via the support $[\underline{t}, \bar{t}]$ of F. No other features of F are relevant.

There are many possible insurer belief profiles which can support this dependence $x_1(t)$. One example is where insurer beliefs result in these insurer expectations

$$\mathrm{E}[t \mid \text{offer } (X, Y)] = \begin{cases} \bar{t} & \text{if} \quad X < x_1(\bar{t}) \\ x_1^{-1}(X) & \text{if} \quad X \in [x_1(\bar{t}), x_1(\underline{t})] \\ \underline{t} & \text{if} \quad X > x_1(\underline{t}) \end{cases}$$

If we temporarily re-express utility as $U(x, y, t) = \tilde{U}(u_x, u_y, t) = p(t)(tu_x + u_y)$, where $u_x = u(-x), u_y = u(-y)$, then for all $t \in R$, $\tilde{U}$ everywhere satisfies the standard single-crossing condition in terms of u_x and u_y, since

$$\frac{d}{dt}\left(\frac{\partial \tilde{U}}{\partial u_x} \Big/ \frac{\partial \tilde{U}}{\partial u_y}\right) = 1. \tag{4.2}$$

Thus the local incentive compatibility equation (equation 4.1) implies global incentive compatibility, and a solution to this differential equation exists. (See Fudenburg and Tirole [FT91], chapter 7.)

Note that the single-crossing condition of equation 4.2 also guarantees that insurer preferences will be single-peaked along *any* separating equilibrium x-y curve satisfying a local incentive compatibility equation such as equation 4.1. After all, if preferences were not single-peaked, at some point the local incentive compatibility equation would have to be satisfied for two distinct points for the same type.

4.2.2 Doing Better

The basic equilibrium $x_1(t)$ is generally ex-ante worse than the ex-ante optimal pooling insurance, which would result from an ex-ante contract between the insuree and insurer which was non-renegotiable. (More precisely, the basic separating equilibrium is worse for the insuree, and the same for the insurer.) Under this optimal ex-ante contract, all types would get the same result, being fully insured as if they were of the average type $\hat{t}$ where $p(\hat{t}) \equiv E_F[p(t)] = \int p(t)dF(t)$. That is, for all t, we would have $x = y = \hat{x} \equiv x^*(\hat{t}, \hat{t})$.

Depending on the game-theory equilibrium refinement used, this efficient pooling result can actually be an equilibrium of our signaling game under asymmetric information. (And, as mentioned above, the empirical evidence for adverse selection in insurance is weak.) But to ensure that there is a real adverse selection problem, we will assume a refinement such as the intuitive criterion [CK87] which selects the fully separating equilibrium even when it is less efficient than pooling.

Full insurance with no adverse selection losses can also result when the insurer is the one to make a take-it-or-leave-it offer, instead of the insuree. For some distributions $F(t)$, the insurer will offer (x, x) where x solves $U(x, x, \underline{t}) = U(L, 0, \underline{t})$ and $V(x, x, \hat{t}) \geq 0$. But with a relatively competitive insurance market the privately informed insuree plausibly has most of the negotiating power, making the above insuree-signaling model a better approximation.

If available, an ideal external agent who shared the interests of the insuree, but who was somehow prevented from becoming informed by that insuree, could directly impose the optimal pooling contract, at least if it were empowered to do so.

Similar to the gains from forcing all types to pool at the same contract, more efficient pooling can also result from an externally-imposed maximum allowed loss $\bar{x}$, which limits the signals an insuree can send. With such a signal limit, there are three possible equilibrium forms, depending on the value of $\bar{x}$. There is a $\tilde{x}$ such that for $\bar{x} \leq \tilde{x}$ everyone pools at $(x, y) = (\bar{x},\ y(\hat{t}, \bar{x}))$. (If $\bar{x} = \hat{x}$, the ex-ante optimal full pooling is achieved.) For $\tilde{x} < \bar{x} < x_1(\underline{t})$ there is partial pooling, and for $x_1(\underline{t}) \leq \bar{x}$ the limit has no effect, leaving full separation.

With partial pooling, bad types $t > \tilde{t}$ separate with $(x_1(t), y_1(t))$, and good types $t \leq \tilde{t}$ pool at $(\bar{x},\ y(p^{-1}(E_F[p(t)|t < \tilde{t}], \bar{x}))$. The cutoff type $\tilde{t}(\bar{x})$ solves

$$U(x_1(\tilde{t}), y_1(\tilde{t}), \tilde{t}) = U(\bar{x}, y(p^{-1}(E_F[p(t)|t > \tilde{t}]), \bar{x}), \tilde{t}),$$

so this type is just indifferent between separating and pooling. The cutoff limit $\tilde{x}$ solves $\tilde{t}(\tilde{x}) = \bar{t}$, so that the worst type is indifferent.

Since the best limit mimics forced pooling, we will focus in the following focus on forced pooling, and just mention in passing its relation to signal limits.

4.3 Group Insurance

Presuming that adverse selection is a real problem, that renegotiation cannot be prevented, that negotiating power cannot be shifted, and that ideal powerful agents are not available, many observers have considered group insurance to be a cure for adverse selection. More generally, many have considered collective choice, especially democratic choice, as a cure for excessive signaling of many sorts.

The intuition seems to be that groups can limit excessive signaling internally by forcing internal pooling or signal restrictions. Also, group level adverse selection is said to be limited because for large groups the variance of group risks are vastly reducted relative to independent individual risks.

Assume there are n copies of the above players, with n insurees and n insurers. There is some joint c.d.f. $J((t_i)_i)$ with support $[\underline{t}, \overline{t}]^n$ over the risk types t_i of each insuree i, and assume that this joint J is symmetric in the t_i, with individual marginals $F(t_i)$. Now instead of having each insuree i directly propose a contract (x_i, y_i) to his insurer, the whole group will jointly propose a single contract (x, y) to all of their insurers. Each insurer will again accept or reject this offer, but since each risk-neutral insurer has the same preferences and gets the same information, the insurers are in identical situations. Thus since there are no coordination effects and there are strict preferences over actions, all insurers will take identical actions.

When signal limits are considered, the group will jointly choose a signal limit $\bar{x}$, and then each insuree and insurer pair will play the individual insurance signaling game.

4.3.1 Profit-Maximizing Group Insurance

Let us now consider a single well-informed group *owner* empowered to choose the group offer (x, y). (Think of an employer choosing a group insurance policy.) This owner is able to offer state-independent compensation c_i to each group member, desires to minimize the total group compensation $\sum_i c_i$, knows the risk type t_i of each group member, and is constrained to give each agent at least a reservation utility level $\underline{U}_i$, so that $U(x - c_i, y - c_i, t_i) \geq \underline{U}_i$.

The risk type of a group is the n dimensional $\vec{t} = (t_i)_i$. Insurers should be interested only in the actual average risk type s, defined by $p(s) = E_J[\bar{p}(\vec{t})|\vec{t} \text{ picked } x, y]$, where $\bar{p}(\vec{t}) \equiv \sum_i p(t_i)/n$. Given this fact and the small dimensional space of possible signals, the most separation we can expect is a semi-pooling equilibrium where the set of all possible $\vec{t}$ is partitioned into a one-dimensional continuum of sets S, such that every type $\vec{t} \in S$ picks the same (x, y) in equilibrium. We can index these sets S by their average risk type s, writing $S(s)$, and the equilibrium choices as well, writing $(x(s), y(s))$.

Each insurer now makes her accept/reject decision based on $V(x(s), y(s), s)$, implying that now $y(s) = y^*(s, x(s))$. When $n = 1$, we must of course have $s = t$, $S(s) = \{t\}$, and the equilibrium $(x_1(s), y_1(s))$ over $[\underline{t}, \bar{t}]$ as before.

The group owner's optimization has a lagrangian $\sum_i c_i + \lambda_i(U(x - c_i, y - c_i, t_i) - \underline{U}_i)$. Substituting the first-order conditions for optimizing the various c_i into the local incentive compatibility condition for separation yields

$$\frac{u'(-y)y'(s)}{u'(-x)x'(s)} = -p^{-1}(\bar{p}(\vec{\gamma})) \tag{4.3}$$

where $\vec{\gamma} = (\gamma_i)_i$ and

$$\gamma_i = t_i \frac{u'(c_i - y)/u'(-y)}{u'(c_i - x)/u'(-x)}.$$

The compensation c_i can be said to have induced a *wealth effect* when $\gamma_i \neq t_i$; in this case the relative marginal value of money in the two states changes as the agent's wealth varies. This wealth effect disappears (so that $\gamma_i = t_i$) with *exponential* utility, where $u'(a + b) = u'(a)u'(b)$, and with *compensating reservation utilities* $\underline{U}_i = U(-x_2(s), -y_2(s), t_i)$ where t_i is in a group indexed by s, and where $x_2(s), y_2(s)$ is the profit-maximizing group insurance equilibrium. We have compensating reservation utilities if and only if we have vanishing equilibrium compensations $c_i = 0$ for all i and $\vec{t}$.

The wealth effect also disappears at $x = y$, which implies that for an interim Pareto dominant equilibrium, the worst possible group type where $t_i = \bar{t} \forall_i$ reveals its type $s = \bar{t}$ and chooses full insurance $x_2(\bar{t}) = y_2(\bar{t}) = x_1(\bar{t})$.

Ignoring wealth effects, profit-maximizing group insurance offers no advantage over individual insurance.

Theorem 4.1 *Given no wealth effects we have $x_2(s) = x_1(s)$, so that the interim Pareto Dominant separating equilibrium curve of profit-maximizing group insurance curve is identical to the interim-Pareto dominant separating equilibrium curve of individual insurance. Given compensating reservation utilities, so all $c_i = 0$, all insurees are no better off than with individual insurance, and are strictly worse off when $t_i \neq s$.*

Proof: When $\gamma_i = t_i$ for all $\vec{t} \in S(s)$, the right-hand side of equation 4.3 is $-p^{-1}(\bar{p}(\vec{t}))$. Thus all the types $\vec{t}$ who choose this $(x(s), y(s))$ have the same value of $\bar{p}(\vec{t})$, and so by $p(s) = E_J[\bar{p}(\vec{t})|\vec{t}$ picked $x, y]$ we have $p(s) = \bar{p}(\vec{t})$. Thus the right side of equation 4.3 is $-s$, making this equation the same as equation 4.1 with s substituted for t. Since we also know that $x_2(\bar{t}) = y_2(\bar{t}) = x_1(\bar{t})$, the interim-Pareto-dominant separating equilibrium for profit-maximizing group insurance is $x_2(s) = x_1(s)$, with the same x, y curve as individual insurance over the same range $s \in [\underline{t}, \bar{t}]$.

With individual insurance, each insuree gets her favorite point along the curve $x_1(t), y_1(t)$ over $t \in [\underline{t}, \bar{t}]$. With compensating reservation utility, each insuree instead gets the point $x_1(s), y_1(s)$ where $p(s) = \bar{p}(\vec{t})$ and $s \in [\underline{t}, \bar{t}]$. Thus the insuree is no better off, and is strictly worse off when $t_i \neq s$. QED.

With no wealth effects, the group chooses contracts off the same x-y curve as with individual insurance. Since separating equilibria depend only on the support of the type distribution, and not its variance, the lower variance of types under group insurance does not help. Thus each individual gets no better an insurance package than he would individually. And when his risk type differs from the average of his group, he gets a worse insurance package.

Note however that even though $(x_2, y_2) = (x_1, y_1)$, the equilibrium is not the same for $n > 1$. Not only might some agents getting compensated for having different insurance amounts, but the distributions over s are not the same, i.e., $F(s) \neq \Pr_J[\bar{p}(\vec{t}) \leq s]$. Thus an equilibrium refinement which does not always select a separating equilibrium may possibly select the separating equilibrium in one case but not the other.[3]

[3]I have not analyzed what further equilibrium refinements are satisfied by these separating equilibria. It should be noted that for reasonable type distributions $F(t)$ and larger electorates, the

4.3.2 Voting on Group Insurance

Democratic collective choice has also been considered a cure for many sorts of adverse selection problems. So let us now consider some simple models of voting over group insurance.

To make the best case for voting to solve signaling problems, let us minimize the amount of information revealed by the voting process, informing the insurers only of the final joint offer (x, y) made, and not of any other statistics regarding the vote. Let us also try to avoid any additional agency costs by using a direct vote on policy, rather than have intermediating politicians. Finally, let us simplify our analysis by focusing on democratic mechanisms which have *pivotal voters*, where the result chosen implies a certain type t was pivotal.

For example, consider a direct two-dimensional (2D) median vote, wherein each voter submits a pair (x_i, y_i), and the chosen announced values are the medians $x = \text{median}_i\ x_i$ and $y = \text{median}_i\ y_i$. In a separating equilibrium where all other voters submit their favorite point along an x-y curve which separates accepted from rejected offers, for n odd each voter of type t_i effectively chooses a point in a box with opposing corners $x(t_{i-1}), y(t_{i-1})$ and $x(t_{i+1}), y(t_{i+1})$, for $t_{i-1} \leq t_i \leq t_{i+1}$, a choice which only matters if t_i happens to be the median type t.

The optimal strategy in this case is for voter i to submit his favorite point $x(t_i), y(t_i)$ along this curve, regardless of what he knows about the preferences of others. The winning (x, y) is thus the favorite point along this x-y curve of the voter with the median type t among the n actual voters. This sample median type is then pivotal.

We can generalize this 2D median vote mechanism to a 2D median jury, where voting is done only by a random sample of j (odd) jurors chosen from the n insurers. Will discuss this option more below.

Another variation is to use a one-dimensional (1D) median vote over the limit

---

distribution of s should be highly concentrated, with very thin tails out to the extreme values. Thin tails should also appear in the distribution of s in the next section on voting. Many have questioned the realism of separating equilibria in this sort of circumstance [MOFP93]. Signaling game models have only been experimentally tested with two near-equal-weight types [BCP94].

$\bar{x}$, where after the vote each pair of insuree and insurer play their signaling game separately. If preferences over $\bar{x}$ are single-peaked, 1D voters should also submit their favorite $\bar{x}$. And for all $\bar{x} \leq \tilde{x}$, voter preferences are the same as in the 2D case, since in this case the effect of a limit $\bar{x}$ is that everyone will pool at $x = \bar{x}$.

Finally, note that we get the same equilibria in a simple two candidate prospective voting model. When candidates are constrained to pick positions on the above equilibrium x-y curve, then if both candidates know the median type of the electorate, and if insurers only observe the position of the winning candidate, then the standard one-dimensional results apply. Since by single-crossing voter preferences are single-peaked along this separating curve, both candidates should pick the ideal point of the median type in the population.

The same equilibrium outcomes also result when candidates are free to pick positions in the entire x-y plane, because no majority prefers any other point to the median type's favorite point $x(t), y(t)$ along the x-y curve which separates accepted from rejected offers.

First, no majority prefers offers which would be rejected. After all, the worst type $\underline{t}$ prefers $x(\underline{t}), y(\underline{t})$ to the rejection point $(L, 0)$, and by single-crossing all other types have even stronger preferences that way. Also, by the monotonicity of u all types also dislike any distinct point (x, y) for which $x \geq x(t)$ and $y \geq y(t)$.

Finally, for any point (x, y) for which $x < x(t)$ and $y > y(t)$, by single-crossing the type who is just indifferent between these two points must be higher than the median type, and the types who strictly prefer the offer (x, y) must be even higher still. But then this set of types could not be a majority. A similar argument applies for (x, y) for which $x > x(t)$ and $y < y(t)$.

4.3.3 Voting Equilibria

For any mechanism where the collective choice of offer (x, y) implies a pivotal type t for whom this (x, y) is best, knowing this pivotal type t gives one information about the expected average group type $s(t)$ where $p(s(t)) = E_J[\bar{p}(\vec{t})|t \text{ pivotal}]$. For example, with a 2D median jury, if the t_i are i.i.d. (independently identically distributed), with $J = \prod_i F(t_i)$, we have

$$p(s(t)) = \frac{1}{n}p(t) + \frac{n-j}{n}p(\hat{t}) + \frac{j-1}{2n}\left(E_F[p(t')|t' \leq t] + E_F[p(t')|t' \geq t]\right). \qquad (4.4)$$

That is, knowing that one juror is the median tells one only that half of the other jurors are worse than that median, that half of the jurors are better, and tells nothing about non-jurors.

Given an average type function $s(t)$ such as equation 4.4, we can again find an interim-Pareto-dominant separating equilibria to this voting game, where again $x'(t) < 0 < y'(t)$, and equation 4.1 is satisfied. (Note that in this section we index equilibrium points by the type of the pivotal voter t, rather than the group average type s.) The difference from individual insurance is that the insurer now makes her accept/reject decision based on $V(x(t), y(t), s(t))$, implying again that $y(t) = y^*(s(t), x(t))$. (This of course reduces to the basic signaling game when $s(t) = t$.)

If we define an (under) insurance ratio $r = u'(-x)/u'(-y)$, then we can describe an agent of type t's favorite point along the (x, y) line $y = y^*(s, x)$ in terms of that type's favorite insurance ratio $r^*(s, t) = s/t$. This shows that voters will prefer other than full insurance when $s(t) \neq t$, i.e., when they do not pay for insurance according to their type's actual risk level. We can also re-express the conditions for a voting equilibrium which is fully-separating in the type of the pivotal voter as $r(\bar{t}) = r^*(s(\bar{t}), \bar{t})$, $y = y^*(s, x)$, and

$$r = \frac{s}{t}\left(1 + \frac{\ln' s}{\ln'(L - x)}\right) \qquad (4.5)$$

where for any $a(t)$, $\ln' a(t) = a'(t)/a(t)$.

These same equations should determine the $x(t)$ for the 2D median vote, the 1D median vote for $x < \tilde{x}$, or for the two candidate prospective election.

4.3.4 Juries Can Do Best

When s and $L - x$ are both increasing in t, the form of equation 4.5 indicates that, except for the worst type, insurees get less insurance (a larger ratio r) than they would prefer. This equation also suggests, however, that this problem goes away if s changes very slowly with t relative to x, i.e., when knowing the pivotal type t tells one very little about the average type s.

Since with a random jury and independent types, the pivotal juror tells one nothing about the types of non-jurors, it turns out that we can get the ex-ante optimum in the limit of large groups with fractionally small juries.

Let us define the prior median type t_m as solving $1/2 = F(t_m)$. This median will of course equal the mean $\hat{t}$ when the distribution F is symmetric. Let us also define $X(r,s)$ as solving $r = u'(-x)/u'(-y^*(s,x))$, and re-express insuree utility as $U(x,y,t) =$

$$\hat{U}(r,s,t) = p(t)\left(tu(-X(r,s)) + u(-y^*(s,X(r,s)))\right).$$

Given these definitions, we can express the following theorem.

Theorem 4.2 *When types are i.i.d. ($J = \prod_i F$) and when juries are large but fractionally-small (i.e., $j \to kn^\alpha$ as $n \to \infty$ for $\alpha \in (0,1)$), the interim-Pareto dominant separating equilibrium of both 1D and 2D median vote juries converges in probability to the utility outcome $\hat{U}(\hat{t}/t_m, \hat{t}, t_m)$. When the prior mean $\hat{t}$ equals the prior median t_m, this implies the ex-ante optimum $(x,y) = (\hat{x},\hat{x})$.*

Proof: First consider 2D median vote juries. We must have

$$\hat{U}(\frac{s}{t}, s, t) \geq \hat{U}(r(t), s, t) \geq \hat{U}(\frac{\bar{s}}{t}, \bar{s}, t)$$

where $s = s(t)$ and $\bar{s} = s(\bar{t})$. The first inequality applies because in a separating equilibria the type t can do no better than to get his favorite ratio $r = s/t$, and the second inequality applies because t always has the option of offering his favorite point along the line corresponding to the worst type $\bar{t}$; such an offer should never be rejected by the insurer in a sequential equilibrium. If $s(t)$ converges to $\bar{s}$ as n grows without bound, as it does for equation 4.4 given our assumption of fractionally-small juries, then the two outside terms have the same limit, which therefore must equal the limit of the inside term.

Since j grows without bound with n, and since in general sample medians converge in probability to prior medians, the pivotal type t must converge in probability to t_m. Equation 4.4 has all $s(t)$ converging to the mean type $\hat{t}$. Thus the outcome converges to $\hat{U}(\hat{t}/t_m, \hat{t}, t_m)$.

If the prior median equals the prior mean, so that $\hat{t} = t_m$, the actual choices x, y must converge to give the ex-ante optimal utility $U(1, \hat{t}, \hat{t})$, which by the strict concavity of u can only happen at $x = y = \hat{x}$. Thus the 2D median jury converges to the ex-ante optimum.

Regarding 1D median juries, we know that $\tilde{x} > x_1(\underline{t}) > \hat{x}$. The first inequality follows from the definition of $\tilde{x}$ and the concavity of U, as in

$$U(x^*(\bar{t}), x^*(\bar{t}), \bar{t}) = U(\tilde{x}, y(\hat{t}, \tilde{x}), \bar{t}).$$

The second inequality follows because both are full insurance points, and $\hat{t}$ is less riskier than $\bar{t}$. Since $\hat{x} < \tilde{x}$, the above proof applies to 1D juries as well, since 2D and 1D equilibria are the same for $x < \tilde{x}$. QED.

A corollary is that the same result applies for the two candidate prospective election model, since we've shown the equilibria offers are the same.

4.4 An Example

Figure 4.1 shows numerically computed separating equilibria for exponential utility $u(x) = -e^{-x}$ with an accident loss of $L = 2$, and with the risk probability of types, $p(t)$ distributed uniformly on $[.53, .77]$. For any uniform distribution on $p(t)$, equation 4.4 becomes

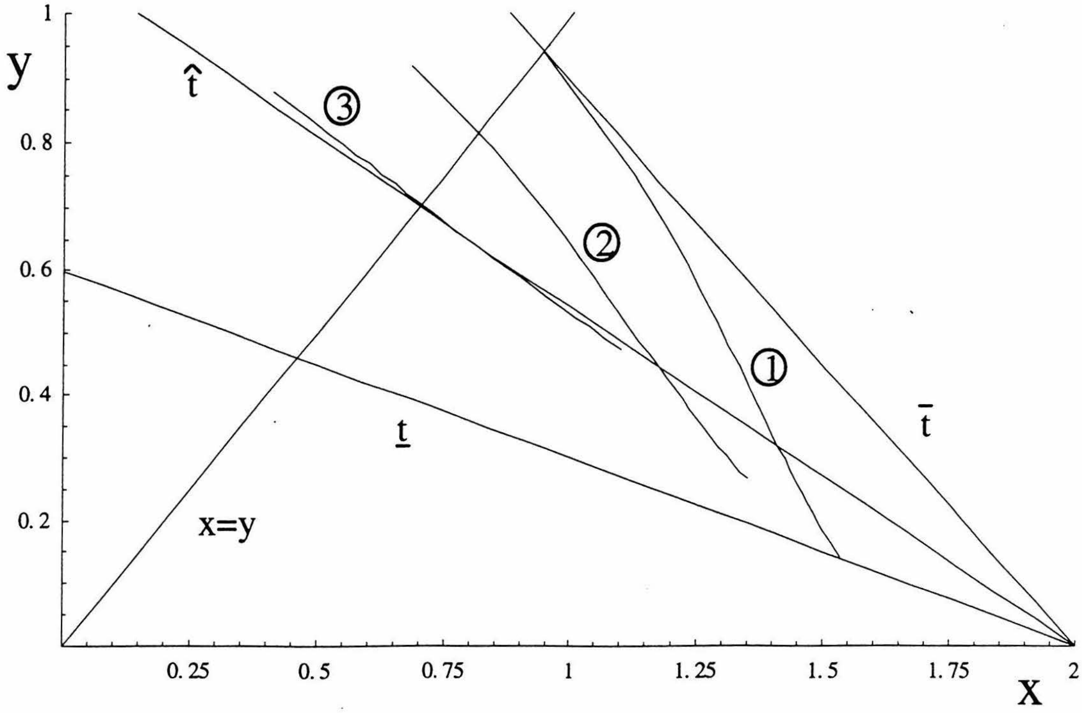

Figure 4.1: Three Examples of Equilibria

$$p(s(t)) = \frac{j+1}{2n} p(t) + \left(1 - \frac{j+1}{2n}\right) p(\hat{t}).$$

For $j = n$ and $n \to \infty$, the range of $p(s)$ goes to $[\frac{1}{2}(p(\underline{t}) + p(\hat{t})), \frac{1}{2}(p(\bar{t}) + p(\hat{t}))]$, which remains wide. Uniform distributions are symmetric, so that $\hat{t} = t_m$, and so theorem 4.2 implies small juries asymptotically reach the ex-ante optimum.

The straight lines in figure 4.1 are insurer indifference curves $V(x, y, t) = 0$ for the minimum, maximum, and average types $\underline{t}, \bar{t}, \hat{t}$. The three curves labeled circle $1, 2, 3$ are separating equilibria for three different cases.

Curve 1 describes x_1, y_1, the equilibrium for both individual insurance, with each insuree and insurer playing separately, and also for profit-maximizing group insurance. The worst type $\bar{t}$ gets full insurance with $x = y$, and better types rapidly get much less than full insurance. For large groups, the typical outcome is where the $V(x, y, \hat{t}) = 0$ line intersects curve 1, where insurance is less that half of full insurance.

Curve 2 describes the voting equilibrium, when there are 121 group members who all vote. When the pivotal voter is the worst possible type, insurers can infer that just over half of the group is the worst type, but can infer nothing about the rest of the group. Since the average inferred type in this case is better than that of the pivotal voter, the pivotal voter faces relatively cheap insurance, and chooses to overinsure. Thus the curve begins with $y > x$. Adverse selection quickly makes better groups underinsure, however, and most outcomes are likely to be near where the $V(x, y, \hat{t}) = 0$ line intersects curve 2.

This voting equilibria does typically induce substantially more insurance than under profit-maximizing group insurance. This improvement can be thought of as due to the fact that the curve starts at a better point for the worse type, since knowing that the worse type is the median type tells insurers nothing about almost half of the electorate.

Curve 3 describes the equilibrium of voting using a random jury of 11 of the 121 voters. Here the worst possible jury signals nothing about the other 110 group members, so it can buy insurance even more cheaply, and hence it chooses to overinsure even more. The most common (i.e., modal) outcome is where the $V(x, y, \hat{t}) = 0$ line intersects curve 3, which gives nearly full insurance, and is close to the ex-ante optimum.

4.5 Discussion

Signaling games have been used for several decades to model a wide variety of economic phenomena, including most social insurance and limits to freedom of contract. Most of these analyses have noted that the "market failure" of inefficient signaling is correctable by exogenous signal restrictions. Most such analyses have suggested, explicitly or implicitly, that some sort of collective choice, such as employer-based group insurance or democratic government intervention, should or does deal with this problem by imposing such restrictions.

Many commentators have observed that such normative conclusions must be tempered by the possibility of "democratic failure." For example, in one-dimensional policy spaces the preferred policy of a pivotal median voter is typically not efficient (given the possibility of transfers), and in higher-dimensional spaces democracy can induce policy cycling. These observations can seem too abstract to offer much of a guide to practical policy, however.

The explicit model of democratic failure in a signaling context presented above may, in contrast, be specific enough to illuminate policy. For example, this model suggests that beyond a dozen or so, the size of a collective is not especially important. What matters more is that member risks are independent, and so not correlated via the process by which this collective was formed, and that the collective choice process be capable of ignoring the opinions of many members, the more the better. Adverse selection is worse, for example, when all opinions are considered, as with profit-maximizing group insurance.

Another mechanism, not explicitly analyzed above, by which a collective choice process could solve an excessive signaling problem is via a special ability to make non-renegotiable commitments. Large fixed costs of invoking a collective choice process might, for example, deter later attempts at renegotiation. Of course those fixed costs would constitute another form of democratic inefficiency.

These observations suggest that, for example, the use of social insurance to mitigate adverse selection might be most successful when managed by administrative agencies who are the least responsive to public input, perhaps by being the most cumbersome and expensive to change. Similarly, judge-made law regarding limits to freedom of contract might best mitigate adverse selection to the extent when judges are relatively unresponsive and difficult to influence.

In general, we might expect substantial agency costs and inefficiencies to be associated with unresponsive administrative agencies and judges. A medium-sized random jury, however, seems able to hold the potential to avoid such a tradeoff. Perhaps labor unions should consider using random juries to propose benefits packages. And perhaps a national jury should be considered to propose a national health care reform.

4.6 Conclusion

While it has long been suggested that group insurance mitigates adverse selection by reducing the variance of the distribution of types, a precise analysis reveals that this is simply not true for separating equilibria. Such equilibria depend only on the support of the distribution of types. For example, since profit-maximizing group insurance considers the preferences of all group members, it has the same distribution of types as individual insurance, and hence offers no adverse selection advantages.

Democratic collective choice has also long been considered as solution to adverse selection and other excessive signaling problems, via government-imposed limits on signals, forced-pooling, or signal taxes. This paper demonstrates that while a democracy where everyone can vote can improve on the problematic equilibria, it also suffers a "democratic failure" and fails to achieve the ex-ante optimum.

This failure is due to the fact that voters can signal their types via the democratic process. At the very least the resulting choice can signal something about the population of types in the electorate, as in the models presented here. And presumably voters could signal even more when signal receivers can observe more detail about the democratic process. Voters might, for example, be able to publicly donate to a particular candidate.

The improvement of democracy over individual choice can be thought of as due to the fact that majority rule can ignore the opinions of up to half of the population. Further improvements can be obtained by narrowing participation even further, such as with a random jury. This suggests a fundamental tradeoff between democratic participation and the ability of governments to solve excessive signaling problems.

If the most effective mitigation of excessive signaling can come from government agents, such as administrative agencies or legal judges, who are the least responsive to the influence of public opinion, then we may face a vexing tradeoff between the agency costs of unresponsive government agents and losses from excessive signaling.

Chapter 5 Disagreements Are Not About Information

5.1 Introduction

Theory and observation seem to be in conflict.

On the one hand, persistent disagreement on matters of fact seems to be ubiquitous in the world. In such disagreements, two or more groups have differing opinions, and seem well aware of this fact, including which side of the issue each group falls on. Consider, for example, the O.J. Simpson trial, where two identifiable communities appeared to persistently disagree on the probability that O.J. killed his wife.[1] Or consider the apparent ubiquity of speculative trade, and the apparent ubiquity of longstanding disagreements in academia, industry, and politics.

On the other hand, we also have some theory that suggests that rational agents cannot agree to disagree in this manner. Bayesians with common priors cannot so disagree [Aum76, SG83, MP86, NBG+90], even approximately [MS89, Nee96a, Son95] nor can agents who satisfy some weaker rationality assumptions [Gea, Gea94, Sam90, Mor94, RW90].

There are three natural resolutions of this puzzle. First, contrary to widespread appearance and belief, most apparent disagreements may not be real. People may not be aware of their disagreement, they may be trying to signal association or ability, or they may be trying to persuade an audience. Second, people may be seriously irrational, so irrational that it would be profitable and feasible for them to modify their behavior to become more in line with these theoretical results. Third, the existing theoretical results may be fragile, and not hold up under more reasonable

[1] For future readers unfamiliar with this "trial of the century," opinion on the guilt of this black football star seem to be correlate with race.

concepts of rationality.

One example of this third type of resolution is the claim that it is not irrational for Bayesians to have different priors. Another example is the suggestion that rational agents need not have the heroic computational abilities that the existing literature on disagreement typically assumes of its agents. For example, Bayesians as well as agents characterized by less constrained possibility correspondences are assumed to exactly calculate expected values over what are typically truly immense sets of possible states. And the few explicitly computational papers which go beyond simulating very specific computational strategies allow their agents to know all theorems which can be generated in any finite time by a Turing machine, regardless of practical limits on the time a real agent can devote to theorem proving [Meg89, SW94, Lip95].

This paper, in contrast, considers Bayesian wannabes, who can have severe constraints on their computational and other abilities, and who may use most any computational strategy to deal with those constraints. Specifically, regarding most every variable of interest, agents in this model can have arbitrary state-dependent errors, i.e., differences between the agent's estimate of that variable and the estimate a Bayesian would make with the same information.

These errors will be subject to only a few easy-to-compute consistency relations, which express the idea that an agent is minimally *savvy* in the sense of being aware of certain of her limitations, and of a few important but easy-to-compute implications of these limitations.[2] Specifically, a savvy agent will make simple broad calibration adjustments to try to correct for any overall biases she perceives in her estimates, and she is aware of certain relations among such biases.

Can such agents disagree? Bayesian agents can disagree, that is, differ in their estimate of some real-valued random variable, either due to differing priors or differing information. And error prone Bayesian wannabes can, in addition, disagree due to the fact that they make different computational errors. But can savvy Bayesian wannabes nearly agree to disagree, so that they are well aware of their disagreement?

[2]This concept of error-prone but savvy agents was inspired by, and generalizes, noisy game theories such as quantal response [MP95, MP96].

And if so, to what can we attribute such disagreement: priors, state information, or computational errors?

While there seems to be some dispute about whether rational[3] agents can really have substantially different priors, it seems clear that given suitably divergence priors, Bayesians can agree to disagree purely due to such priors. Thus we should consider this possibility in our analysis. Similarly, we should consider the possibility of agents purely agreeing to disagree due to computational differences, even if we have reservations about the rationality of such behavior.

An example of a pure computational disagreement would be where one agent estimates $\pi \approx 3.14$ while the other estimates $\pi \approx 22/7$, where neither agent is willing to calculate π more exactly, and where both agents are well aware that they each use these different approximations. Such a dispute could not be attributed to external state information, if both agents were aware that the value of π is the same in all self-consistent state descriptions. Nor could such a dispute be attributed to internal state information, if the agents had no relevant uncertainties about the computational approach being used by each agent.

While Bayesian wannabes can apparently agree to disagree purely due to divergent priors, or purely due to computational errors, they cannot agree to disagree, even approximately, purely due to different information. ("Purely" here means with common priors and zero errors.) An important open question, however, is whether such agents can agree to disagree due to an intrinsic combination of differing information and computational errors, or whether all such situations can be traced to a pure case of agreeing to disagree over computation.

This paper answers this question by showing that if any two error-prone Bayesian wannabes nearly agree to disagree regarding any random variable, are both aware of a certain result of this paper, and nearly agree that both are self-respecting in the sense of considering themselves well-calibrated, and that one of them is savvy in the sense of realizing a particular implication of calibration attempts, then these agents must nearly agree to disagree about the other agent's average calibration bias. A

[3]By "rational" I mean consistent with feasible-to-implement normative criteria.

savvy Bayesian wannabe thinks she is likely well-calibrated, so if she nearly agrees to disagree with another such agent, she must think the other agent is very likely miscalibrated.

Though such average calibration bias would be extremely difficult to compute exactly, private state information, even regarding the internal state of other agents, is irrelevant to its computation. Thus this is a case of a purely computational disagreement, which can not be attributed to any uncertainty agents might have about either the external state of nature or about the internal state of other agents.

We thus conclude that any persistent near disagreement between savvy but error-prone Bayesian wannabes with common priors can be attributed to a near disagreement purely about computation. Even for severely computationally constrainted agents, disagreements cannot be about information. And to the extent that divergent priors and pure computational disagreements are irrational, rational agents simply can not agree to disagree. In this case human disagreements would have to either be mostly illusory, or they would have to be attributed to correctable irrationality.

5.2 The Model

5.2.1 Bayesian Wannabes

We consider a finite universe containing agents who would be perfect Bayesians, if it were not for the fact that they are usually subject to computational resource constraints,

That is, given a finite (but typically *very* large) set of possible states Ω, at each state $\omega \in \Omega$ agent i would, in the absence of computational constraints, estimate any real-valued random variable $X(\omega) \in [\underline{X}, \bar{X}]$ via the exact Bayesian expected value

$$X_i(\omega) = \mathrm{E}_{\mu_i}[X \mid I_i(\omega)] = \frac{\sum_{\omega' \in I_i(\omega)} X(\omega')\mu_i(\omega')}{\sum_{\omega' \in I_i(\omega)} \mu_i(\omega')} \tag{5.1}$$

where $\mu_i(\omega) > 0$ is a prior and the *information* set $I_i(\omega)$ is an element of the partition I_i which embodies all implications of the information available to agent i. We assume a common prior, so $\mu_i = \mu$ and $\mathrm{E}_{\mu_i} = \mathrm{E}$.

These are typically sums over very large sets, however, so an agent i with limited computational resources must typically settle for an approximate *estimate*

$$\tilde{X}_i(\omega) = \tilde{\mathrm{E}}_{i\omega}[X_i(\omega)] = X_i(\omega) + e_{i\omega}[X],$$

also within $[\underline{X}, \bar{X}]$, which differs from the ideal Bayesian estimate $X_i(\omega)$ by an error $e_{i\omega}[X]$. More generally, for any random variable $Y(\omega)$, *Bayesian wannabe* i at state ω makes do with an error-prone estimate $\tilde{\mathrm{E}}_{i\omega}[Y] = Y(\omega) + e_{i\omega}[Y]$.

We consider a Bayesian wannabe i to be characterized by (I_i and) her estimation operator $\tilde{\mathrm{E}}_{i\omega}$, or equivalently by her error operator $e_{i\omega}$. This characterization of an agent is limited in the sense that the objects of beliefs are well-defined random variables $Y(\omega)$; Bayesian wannabes do not suffer from ambiguity regarding the meaning or referents of their estimates.

Beyond this limitation, however, we consider a very general class of computationally-limited agents, and attempt to make only the weakest restrictions required to obtain our results. Thus unless we explicitly assume otherwise, we assume no necessary relations between our agent's estimates of related random variables. For example, even if $c = a + b$ it need *not* be true that

$$\tilde{\mathrm{E}}_{i\omega}[c] = \tilde{\mathrm{E}}_{i\omega}[a] + \tilde{\mathrm{E}}_{i\omega}[b].$$

Such a relation may hold, however, when agent i is aware that $c = a + b$, and when $a + b$ is not a difficult computation. More generally, we will informally say that an agent is *aware* of certain simple easy-to-compute consistency conditions on relations between various random variables Y if she makes her estimates $\tilde{\mathrm{E}}_{i\omega}[Y]$ also satisfy these consistency conditions and relations.

States ω identify complete self-consistent descriptions of possible realities, and so encode information about both the external world and the internal reasoning of agents. Since one could in principle calculate the errors $e_{i\omega}[X]$ by calculating both the approximate $\tilde{X}_i(\omega)$ and the exact $X_i(\omega)$, these errors $e_{i\omega}[X]$ cannot give more information than the information I_i available to an ideal Bayesian agent. Thus for all $\omega' \in I_i(\omega)$, we must have $e_{i\omega'} = e_{i\omega}$ and $\tilde{\mathrm{E}}_{i\omega'} = \tilde{\mathrm{E}}_{i\omega}$. Of course for i an ideal Bayesian, $e_{i\omega}[Y] = 0$ and $\tilde{\mathrm{E}}_{i\omega}[Y] = Y$ for all ω, Y.

5.2.2 Calibration

While we may not allow agents to fine tune all their errors $e_{i\omega}[X]$ in the light of exact calculations of expectations (since then we'd get all $e_{i\omega}[X] = 0$), we may allow an agent to *calibrate* her estimates. That is, we consider $e_{i\omega}[X] = m_{i\omega}[X] - c_i[X]$, where $m_{i\omega}[X]$ would be the mistake with zero calibration, and $c_i[X]$ is a calibration adjustment.

Choices of calibration $c_i[X]$ would be made at *calibration sets* $D_i^X(\omega) \supset I_i(\omega)$, the set over which the agent cannot help but to make the same calibration choice of X, given her general computational strategy. (We now typically suppress X in the notation of c_i and D_i.)

$D_i(\omega) \neq I_i(\omega)$ for agents who cannot afford to make their calibration choices contingent on all available information. An agent cannot typically compute $D_i(\omega)$, but she may be aware that $D_i = \{D_i(\omega) \mid \omega \in \Omega\}$ must be a partition that coarsens I_i.

To make more precise the idea that what we mean by $\tilde{X}_i$ is agent i's best computationally-feasible approximation to her ideal expected value $X_i(\omega)$, we might let $u_i(\omega) = -(\tilde{X}_i(\omega) - X(\omega))^2$ be i's utility, and have i "seek to" adjust her calibration c to minimize

$$\bar{U}(c, \omega) = \mathrm{E}[(\tilde{X}_i(c) - X)^2 \mid D_i(\omega)],$$

the expectation of her squared error at this calibration set.

Agent i at state ω does not, however, typically have access to the exact values of this function $\bar{U}(\cdot, \omega)$ of c. She must instead make do with an estimate of $\tilde{\mathrm{E}}_{i\omega}[\bar{U}(c, \omega)]$. Furthermore, her choice of c may not exactly minimize this estimate. Under these circumstances, it may not be clear what exactly it means for u_i to be agent i's utility.

We will attempt to side-step this conceptual muddle by focusing in the sections which follow on the following two definitions. First, let us define an agent's *bias* as $\bar{e}_i[X|S] = \mathrm{E}[e_{i\omega}[X] \mid S]$, i.e., average error over some set S of possibilities. Second, let *self-respect* be a bound on a self-estimate of bias.

Definition 1 *Agent i at ω displays δ-*self-respect *about X on E if* $|\tilde{\mathrm{E}}_{i\omega}[\bar{e}_i[X|E]]| \leq \delta$.

This notion of self-respect can be motivated as follows. First, we can show the following (non-trivial proofs in the Appendix).

Lemma 5.1 *The $c_i[X]$ which minimizes* $\mathrm{E}[(\tilde{X}_i - X)^2 \mid D_i(\omega)]$ *sets* $\bar{e}_i[X|D_i^X(\omega)] = 0$.

It seems plausible that an agent i who is aware of lemma 5.1 and who desires to minimize her squared-error of $\tilde{X}_i$ should estimate that she is well-calibrated, so that

$$\tilde{\mathrm{E}}_{i\omega}[\bar{e}_i[X|D_i(\omega)]] = 0.$$

After all, if she estimated that she was biased, it seems she should expect to do better by changing her calibration c_i. Thus such an agent i at state ω should display perfect ($\delta = 0$) self-respect on $D_i(\omega)$.

Furthermore, if this agent didn't think she could identify any miscalibrations among her alternative selves at other states in an event E, then she should also satisfy

$$\tilde{\mathrm{E}}_{i\omega}[\bar{e}_i[X|D_i(\omega')]] = 0 \ \ \forall \omega' \in E.$$

If this agent also conditioned her calibration on this event, so that E was a union of D_i members (meaning $D_i(\omega) \subset S$ for all $\omega \in E$), then this would imply that $\tilde{\mathrm{E}}_{i\omega}[\bar{e}_i[X|E]] = 0$.

Thus we can think of the self-respect parameter δ as describing the degree to which agent i estimates her alternative selves to be on average miscalibrated, or to fail to condition their calibration on the event E.

5.2.3 Agreement

A Bayesian agent i *knows* event E if $E \supset I_i(\omega)$, and an event E is *common knowledge* among a set N of such agents at state ω if $E \supset I(\omega) = \wedge_i I_i(\omega)$, where the common information partition I is the meet (or finest common coarsening) of the partitions $(I_i)_{i \in N}$.

Bayesians are also said to *p-believe* event E at states within belief sets

$$B_i^p(E) = \{\omega \mid \mu(E \mid I_i(\omega)) \geq p\},$$

and are said to have *common p-belief* of event E at states within any p-common event C where $C \subset B_i^p(C)$ and $C \subset B_i^p(E)$ for all $i \in N$. There is a unique such p-common event $C^p(E)$ which contains all other such p-common events of E. It is found by requiring p-belief at all meta belief levels, as in

$$C^p(E) = \bigcap_{n \geq 1} E^n \qquad E^n = \bigcap_{i \in N} B_i^p(E^{n-1}),$$

for $E^0 = E$ [MS89].

Let us also say that such agents *p-agree* that E at states in any p-agreement event C where $C \subset B_i^p(C \cap E)$ for all $i \in N$. Note that either common p-belief or p-agreeing at level $p = 1$ implies common knowledge, and that these concepts are closely related via their p values.

Lemma 5.2 *Common p-belief implies* $2p$–1*-agreeing, and p-agreeing implies common p-belief.*

Thus the choice between these two concepts seems largely a matter of convenience.

Regarding Bayesian wannabes, let us generalize these definitions. Let us say that an agent *q-estimates* an event E within the estimation set

$$\tilde{B}_i^q(E) = \{\omega \mid \tilde{\mathrm{E}}_{i\omega}[\mu(E \mid I_i(\omega))] \geq q\},$$

and say that the *accuracy* of agent i on this estimation set is $\mu(E \mid \tilde{B}_i^q(E))$. (Bayesians q-estimate an event if and only if they q-believe it, and they have an accuracy of q in their q-estimation.)

We could say that a set N of agents had "common q-estimation" of event E at states within any event C which satisfied $C \subset \tilde{B}_i^q(C)$ and $C \subset \tilde{B}_i^q(E)$ for all $i \in N$. It will, however, be more convenient to focus on a generalization of p-agreeing.

Definition 2 *Agents N* q-agree *that E within any C where*

$$C \subset \cap_{i \in N} \tilde{B}_i^q(C \cap E). \tag{5.2}$$

We will call such an event C a *q-agreement* event of E, and call equation 5.2 its agreement equation.

Let us also define awareness.

Definition 3 *Agent i is* q, q'-aware *that $S \subset T$ if $\tilde{B}_i^q(S) \subset \tilde{B}_i^{q'}(T)$.*

One expects $q' \leq q$. Similarly, let us say about predicates P, Q that i is q, q'-aware that P implies Q if i is q, q'-aware that $\{\omega | P \text{ at } \omega\} \subset \{\omega | Q \text{ at } \omega\}$. And let us say that i is q, q'-aware that P implies Q relative to $\mathcal{A}$ when i is q, q'-aware that the conjunction of P and $\mathcal{A}$ implies the conjunction of Q and $\mathcal{A}$.

Note that the union $C \cup C'$ of any two q-agreement events of E is itself a q'-agreement event of E, if each agent is q, q'-aware that $(C \cap E) \subset (C \cap E) \cup (C' \cap E) \supset (C' \cap E)$. Note also that q-agreement C of P, together with q, q'-awareness that P implies Q relative to C, implies that C is also a q'-agreement of Q.

Finally, note that while, for Bayesians, p-agreeing that E at ω implies constraints on higher order beliefs such as $\omega \in B_1^p(B_1^p(E) \cap B_2^p(E))$, Bayesian wannabes who q-agree need not satisfy analogous relations like $\omega \in \tilde{B}_1^q(\tilde{B}_1^q(E) \cap \tilde{B}_2^q(E))$. This is because we have not made any assumptions relating beliefs of varying orders. If one wanted to additionally require agents to be aware of their agreement, one might formally require q, q'-awareness of the agreement equation 5.2. This would be satisfied, with $q = q' = 1$, for any agent who in all states took equation 5.2 as part of the definition of the set C.

5.2.4 Disagreement

We can informally say that two agents disagree regarding their estimates of X when $\tilde{X}_i \neq \tilde{X}_j$. There are three possible reasons for such disagreement. Bayesians, for whom $\tilde{X}_i = X_i$, can disagree purely due to divergent priors, as when $\mu_i \neq \mu_j$ even though $I_i = I_j$. Bayesians can also disagree purely due to differing information, as when $I_i \neq I_j$ even though $\mu_i = \mu_j$. Furthermore, Bayesian wannabes can also disagree purely due to bounded computation, such as when $e_i \neq e_j$ even though $X_i = X_j$. And of course disagreements can be due to combinations of these three causes.

While it may be easy to understand how disagreement might arise in general, it seems harder to understand how two agents could repeatedly interact in ways which inform them about their difference of opinion, and yet end up with stable but differing opinions. That is, how could two rational agents both know that they disagree, without at least one of them wanting to adjust her estimate $\tilde{X}_i$ in the direction of the other agent's estimate $\tilde{X}_j$?

Definition 4 *Agents i, j are said to ϵ-disagree about X when $\tilde{X}_i \geq \tilde{X}_j + \epsilon$.*

Note that this definition is not symmetric between the agents i, j. When j, i ϵ-disagree about X we instead have $\tilde{X}_j \geq \tilde{X}_i + \epsilon$.

Let

$$E = \{\omega | \tilde{X}_i(\omega) \geq \tilde{X}_j(\omega) + \epsilon\}$$

be called the i, j ϵ-disagreement event about X. Then we can say that i, j *q-agree to ϵ-disagree* about X if they q-agree regarding i, j's ϵ-disagreement event about X.

Such an agreement to disagree would allow each of them to adjust their estimate in the direction of the other agent's estimate.

It is well known that if Bayesians have differing priors, they can disagree, and agree to disagree, purely due to those differing priors. It is less clear, however, that it is rational for agents to have differing priors. It also seems possible for error-prone Bayesian wannabes to agree to disagree purely due to computational errors, though it is also not clear whether this behavior is rational.

Let us say that agents i, j *q-agree to ϵ-disagree about the computation* of Y when they q-agree to ϵ-disagree about a random variable Y which is state-independent, so $Y(\omega) = Y$. There are many possible examples of such computational disagreement.

- Agents may use differing approximations to π to compute the volume of a coin. (See Examples section.)
- Agents may use different computational strategies to search for and estimate the coordinates of the minimum of some complex but state-independent function.
- Agents with access to the same photos and other relevant information about jars and jelly beans may still compute different estimates of the number of jelly beans in a jar.

In such cases, it seems possible, though not necessarily fully rational, for agents to be fully aware of the alternative computational strategies used by other agents, without being very tempted to change their choice of their own strategies. An agent's choice of computational strategy may, for example, be particularly suited to that agents computational hardware. In such situations, disagreement cannot be attributed to uncertainty about other agent's computational strategies or other internal state such as randomization choices.

In contrast, it is known that Bayesians *cannot* agree to disagree, even approximately, purely due to information [Aum76, MS89, Son95, Han94]. It may still be considered an open question, however, whether Bayesian wannabes with common priors can agree to disagree due to a combination of differing information and computational limits. That is, can computationally-limited agents who have different information but the same priors agree to disagree without also agreeing to disagree about some relevant computation?

5.3 Analysis

To address this question, we focus attention in the remainder of this paper on the implications of two agents who nearly agree to disagree at ω about X.

That is, assume that agents $1,2$ q-agree to ϵ-disagree about X. So there is a $1,2$ ϵ-disagreement event $E = \{\omega \,|\, \tilde{X}_1(\omega) \geq \tilde{X}_2(\omega) + \epsilon\}$, with $\epsilon > 0$, and a q-agreement event C satisfying equation 5.2. Also, if agents imagine there could be more than one q-agreement event of E which satisfies equation 5.2, assume that in all states both agents focus attention on the same unique q'-agreement event, such as the one with the largest prior weight.

To further simplify our notation, let $A = C \cap E$ be the set our analysis will focus on, let $B_i = \tilde{B}_i^q(A)$ be its estimation sets, let $\bar{e}_i = \bar{e}_i[X|B_i]$ be the agent's bias on those sets, and let $p_i = \mu(A \,|\, B_i)$ be the agent's accuracy on those belief sets. Note that all of these variables, E, C, A, B_i, p_i, and $\bar{e}_i$, are state-independent random variables. Thus agreeing to disagree about any one of them is agreeing to disagree about computation.

A little algebra reveals the following.

Lemma 5.3 $\bar{e}_1/p_1 - \bar{e}_2/p_2 = \mathrm{E}[\tilde{X}_1 - \tilde{X}_2 \,|\, A] + \mathrm{E}[\tilde{X}_1 - X \,|\, B_1 \backslash A](1-p_1)/p_1 - \mathrm{E}[\tilde{X}_2 - X \,|\, B_2 \backslash A](1-p_2)/p_2$.

Let us define $\Delta X = \bar{X} - \underline{X}$, $p_0 = \min(p_1, p_2)$ and

$$\hat{\epsilon}(p, \delta) = p\epsilon - 2(1-p)\Delta X - \delta/p,$$

which is positive for ϵ positive, δ not too large, and p close enough to 1. Then lemma 5.3 implies the following.

Lemma 5.4 *$\epsilon \geq 0$ and $\bar{e}_2 \geq -\delta_2 \leq 0$ imply $\bar{e}_1 \geq \hat{\epsilon}(p_0, \delta_2)$.*

For lemma 5.4 and the analysis which follows, results are given only for one of the two agents $1, 2$. Symmetric results for the other agent can be found by simultaneously switching agent labels $1 \leftrightarrow 2$ and variables $X \leftrightarrow -X$, $-\tilde{X} \leftrightarrow (\tilde{-X})$, since we can also write $E = \{\omega | -\tilde{X}_2(\omega) \geq -\tilde{X}_1(\omega) + \epsilon\}$.

An agent who is aware of lemma 5.4 should be savvy in the sense of keeping her estimates consistent with the inequality constraint given there.

Definition 5 *Agent 2 at state ω is* s-savvy *if* $\tilde{\mathrm{E}}_{2\omega}[\bar{e}_1] \geq \hat{\epsilon}(\tilde{\mathrm{E}}_{2\omega}[p_0], |\tilde{\mathrm{E}}_{2\omega}[\bar{e}_2]|) - s$.

Note as a corollary that if an s-savvy agent 2 estimates agent 1 to be unbiased, as in $\tilde{\mathrm{E}}_{2\omega}[\bar{e}_1] = 0$ then

$$\hat{\epsilon}(\tilde{\mathrm{E}}_{i\omega}[p_0], |\tilde{\mathrm{E}}_{2\omega}[\bar{e}_2]|)) - s \leq 0.$$

For large ϵ, this requires either poor savvy or self-respect, or a low estimate of accuracy p_0. Note also that it seems feasible for an agent i to condition her calibration on being in her estimation set, to make B_i be a union of D_i members.

If agent 2 is both s-savvy and δ_2-self-respecting about X on B_2, then

$$\tilde{\mathrm{E}}_{2\omega}[\bar{e}_1] \geq \hat{\epsilon}(\tilde{\mathrm{E}}_{2\omega}[p_0], \delta_2) - s,$$

and if agent 1 is δ_1-self-respecting about X on B_1 then $|\tilde{\mathrm{E}}_{1\omega}[\bar{e}_1]| \leq \delta_1$. These trivially imply the following.

Lemma 5.5 If $1, 2$ *are δ_i-self-respecting about X on B_i, if 2 is s-savvy, and if* $\tilde{\mathrm{E}}_{2\omega}[p_0] \geq p$, *then* $2, 1$ *$\bar{\epsilon}$-disagree about $\bar{e}_1$, for $\bar{\epsilon} = \hat{\epsilon}(p, \delta_2) - s - \delta_1$.*

Lemma 5.5 can be considered an implication, P implies Q. So if an agent is q', q''-aware of lemma 5.5, then if she q'-estimates the event P, that the assumptions of lemma 5.5 hold, she must also q''-estimate the event Q, that the conclusion of lemma 5.5 holds. Similarly, if the agents have a q'-agreement event C' of P, and they are q', q''-aware, relative to C', of lemma 5.5, then C' must also be a q''-agreement event for Q.

This brings us to our main conclusion.

Theorem 5.1 *Regarding agents* 1, 2 *possibly* q*-agreeing to* ϵ*-disagree about* X *(at agreement* C *for* $\epsilon > 0$*)*, if

1. *both agents* q'*-agree (in agreement* C'*) that*

   (a) *agent* 2 *is* s*-savvy and estimates both agents to be* p*-accurate on* B_i *(so* $\tilde{\mathrm{E}}_{2\omega}[p_0] \geq p$*), and*

   (b) *each agent is* δ_i*-self-respecting about* X *on* $B_i \equiv \tilde{B}_i^q(C \cap \{\omega | \tilde{\mathrm{E}}_{1\omega}[X(\omega)] \geq \tilde{\mathrm{E}}_{2\omega}[X(\omega)] + \epsilon\})$,

2. *and if each agent is* q', q''*-aware of lemma 5.5 relative to* C',

then *within* C' *agents* 2, 1 q''*-agree to* $\tilde{\epsilon}$*-disagree about the computation of* $\bar{e}_1$*, with*

$$\tilde{\epsilon} = p\epsilon - 2(1-p)\Delta X - \delta_2/p - s - \delta_1.$$

That is, let us consider the possibility that two agents nearly agree to disagree strongly enough regarding any real-valued random variable. Assume, regarding this possibility, that both are sufficiently aware of lemma 5.5 and have near agreement that they both are self-respecting enough (i.e., aware enough of lemma 5.1), and that one of the agents is savvy enough (i.e., aware enough of lemma 5.4) and considers both agents to be accurate enough in estimating their agreement. Then we can conclude that these agents must nearly agree to disagree about the the average calibration bias of the other agent.

Since average calibration bias $\bar{e}_1$ has been specified in a state-independent manner, private state-information is irrelevant to its computation. Thus savvy Bayesian wannabes who nearly agree to disagree about anything must nearly agree to disagree purely on computation, where state information is irrelevant.

Note that the statement of theorem 5.1 tries as far as possible to avoid fragile assumptions. For most assumptions used, a parameter is given so that one can relax that assumption by varying that parameter. A total of eight such parameters are given: $q, q', q'', p, \epsilon, \delta_1, \delta_2, s$. Note also that both E and C are allowed to be empty sets.

5.4 Examples

Let us now consider two examples which illustrate the above results.

In our first example two agents $i = 1, 2$ estimate the dollar value v of a gold coin. There is only one relevant state $\bar{\omega}$, so $\Omega = \{\bar{\omega}\}$. At this state, the agents agree that the coin has a height of $h = .1$ units and a radius of $r = 10$ units. They also agree that gold is worth $v/V = 100$ dollars per unit volume, and that the volume of a disk is given by $V = \pi r^2 h$. So they agree that $v = (v/V)\pi r^2 h$. That is, for $i = 1, 2$, we have

$$\tilde{\mathrm{E}}_{i\bar{\omega}}[v] = \tilde{\mathrm{E}}_{i\bar{\omega}}[v/V] \times \tilde{\mathrm{E}}_{i\bar{\omega}}[\pi] \times \tilde{\mathrm{E}}_{i\bar{\omega}}[r]^2 \times \tilde{\mathrm{E}}_{i\bar{\omega}}[h],$$

and $\tilde{\mathrm{E}}_{i\bar{\omega}}[h] = .1$, $\tilde{\mathrm{E}}_{i\bar{\omega}}[r] = 10$, and $\tilde{\mathrm{E}}_{i\bar{\omega}}[v/V] = 100$.

The agents disagree, however, in their estimates of the mathematical constant π, since $\tilde{\mathrm{E}}_{1\bar{\omega}}[\pi] = 3.14$ and $\tilde{\mathrm{E}}_{2\bar{\omega}}[\pi] = 22/7$. Thus the agents also disagree in their estimates of the value of the coin, with $\tilde{\mathrm{E}}_{1\bar{\omega}}[v] = 3.14$ and $\tilde{\mathrm{E}}_{2\bar{\omega}}[v] = 22/7$.

Since there is only one state $\bar{\omega}$, both agents perceive that there is only one self-consistent description of a possible reality. The agents do not perceive any state-uncertainty regarding any relevant parameters, including the estimates made by the other agent. This does not mean that such estimates are error free, however. For example it is possible that $\tilde{\mathrm{E}}_{2\bar{\omega}}[\tilde{\mathrm{E}}_{1\bar{\omega}}[\pi]]$ could equal either 3.14 (with zero error) or 3.15 (with an error of .01).

Our second example is an adaptation of an example by Neeman [Nee96a]. Let there be three states, $\Omega = \{1, 2, 3\}$, and a random variable $X = (X(1), X(2), X(3)) = (1, 1, 0)$. Two agents $i = 1, 2$ have information sets $I_1 = \{\{1, 2\}, \{3\}\}$ and $I_2 = \{\{1\}, \{2, 3\}\}$, and a common prior $\mu = \{(1-p)/(2-p), p/(2-p), (1-p)/(2-p)\}$.

If agents 1, 2 were Bayesians, they would each have $\mu(2|I_i(2)) = p$, and thus at

$\omega = 2$ have both common p'-belief and p'-agreement, for any $p' \leq p$, of any event E such that $2 \in E$. They would also have estimates $X_1 = (1,1,0)$ and $X_2 = (1,p,p)$, giving a disagreement of $X_1 - X_2 = (0, 1-p, -p)$, and p-agree at $\omega = 2$ that they $1-p$-disagree.

Since both Bayesians have zero bias, $\tilde{\mathrm{E}}_{2\omega}[\bar{e}_i] \geq \tilde{\mathrm{E}}_{1\omega}[\bar{e}_i] + \hat{\epsilon}(p)$ becomes $0 > \hat{\epsilon}(p) = p(1-p) - 2(1-p)$.

If agents 1, 2 are Bayesian wannabes, their actual estimates $\tilde{\mathrm{E}}_{i\omega}$ can differ from Bayesian estimates by errors $e_{i\omega}$. Since these errors constitute a very large number of possible parameters, we consider a very specific case.

Let $p = .97$ and $D_i = I_i$. If $e_{12}[\mu(2|I_1(2))] = -.01$ and $e_{22}[\mu(2|I_2(2))] = .01$, then at $\omega = 2$ the agents q-agree, for $q = .96$ regarding the event $\{2\}$. If $e_1[X] = (-.01, -.01, .04)$, and $e_2[X] = (-.01, -.15, -.15)$, then using $\epsilon = .15$ we have $\tilde{X}_1 - \tilde{X}_2 - \epsilon = (-.15, .02, -.93)$, and thus they also q-agree at $\omega = 2$ that they ϵ-disagree, since $\tilde{X}_1 \geq \tilde{X}_2 + \epsilon$.

Let us assume that at state 2 both agents estimate that both are exactly $(s = 0)$ savvy, exactly $(\delta = 0)$ self-respecting at state 2 regarding X on their information sets $I_i(2)$, and are .95-accurate on those sets. Let us also assume that in all states both agents are fully $(1,1)$ aware of lemma 5.5 relative to $\{2\}$.

Assuming the agents are actually savvy at state 2, the agent's estimates will satisfy $\tilde{\mathrm{E}}_{22}[\bar{e}_1] \geq \hat{\epsilon}(.15, 0) = .0425$ and $\tilde{\mathrm{E}}_{12}[\bar{e}_2] \leq -.0425$. These consistency constraints are satisfied for errors $e_{12}[\bar{e}_2] = .04$ and $e_{22}[\bar{e}_1] = .1$, since then $\tilde{\mathrm{E}}_{22}[\bar{e}_1] = .09 \geq .0425$ and $\tilde{\mathrm{E}}_{12}[\bar{e}_2] = -.11 \leq -.0425$. The resulting disagreement about average bias is as predicted, with $\tilde{\mathrm{E}}_{22}[\bar{e}_1] - \tilde{\mathrm{E}}_{12}[\bar{e}_1] - \hat{\epsilon}(p, 0) = .0475 > 0$ and $\tilde{\mathrm{E}}_{22}[\bar{e}_2] - \tilde{\mathrm{E}}_{12}[\bar{e}_2] - \hat{\epsilon}(p, 0) = .0675 > 0$.

5.5 Conclusion

Since Bayesians with a common prior cannot "agree to disagree", to what can we attribute persistent human disagreement? One possibility is that disagreements are only apparent. A second possibility seems to be purely differing priors, while a third possibility seems to be purely differing computational errors. However since differing information appears to many to play a central role in human disagreements, many have sought a forth theoretical alternative, where information plays a central role. Can we understand persistent human disagreement as due to an intrisnic mixture of differing information and unavoidable computational constraints on human inference?

In an attempt to address this question, this paper has introduced the concept of Bayesian wannabes, which is a very general class of agents who would be Bayesians if it were not for computational limitations. The paper also introduced the concept of savvy, i.e., awareness of certain easy-to-compute implications of computational errors.

Speaking loosely, for savvy Bayesian wannabes with common priors we have shown that arbitrary situations of "nearly agreeing that both agents are savvy and yet disagree about something else" imply "nearly agreeing to disagree about the computation of average calibration bias".

Thus we can attribute persistent disagreement to not wanting to be Bayesian, to differing priors, to persistent disagreement about a matter of pure computation, or to a lack of savvy, but *not* to any other sort of differing information. That is, the only differing information which can alone explain disagreement is uncertainty about whether the other agent is aware of certain state-independent facts about disagreements in general. Beyond this, disagreements are not about information.

While this paper has introduced and illustrated the concept of agreeing to disagree about computation, it has said relatively little about this concept, which seems ripe for further analysis. Are there further simple-to-compute consistency relations, i.e,. further concepts of savvy, which can limit the situations when it is possible to rationally agree to disagree about computation?

Agents who agree to disagree regarding computation would seem to be forgoing Pareto improvements obtainable by a combination of a monetary transfer and an agreement to use some intermediate estimate. Can this intuition be formalized? Can such agreements be enforced?

Finally, what dynamics should we expect for computational-error-driven disagreements? Are there empirically-testable differences between such dynamics and the dynamics one should expect for information-driven and prior-driven disagreements?

5.6 Appendix

Lemma 5.1 *The c_i which minimizes $E[(\tilde{X}_i - X)^2 \mid D_i(\omega)]$ sets $\bar{e}_i[X|D_i(\omega)] = 0$.*

Proof: Since $e_i = \tilde{X}_i - X_i$, then $(\tilde{X}_i - X)^2 = (e_i + (X_i - X))^2 = e_i^2 + (X_i - X)^2 + 2e_i(X_i - X)$. The second term on the right cannot be affected by adjusting $\tilde{X}_i$, and the expectation of the third term on the right over $D_i(\omega)$ vanishes because e_i is constant over each $I_i(\omega) \subset D_i(\omega)$ and $E[X_i - X \mid I_i(\omega)] = 0$, which follows from equation 5.1. Thus to minimize $E[(\tilde{X}_i - X)^2 \mid D_i(\omega)]$ is to minimize $E[e_i^2 \mid D_i(\omega)]$. Write $e_i = m_i - c_i = (m_i - \bar{m}_i) + (\bar{m}_i - c_i)$, where $\bar{m}_i = E[m_i \mid D_i(\omega)]$. Then $e_i^2 = (m_i - \bar{m}_i)^2 + (\bar{m}_i - c_i)^2 + 2(m_i - \bar{m}_i)(\bar{m}_i - c_i)$. But the expectation of the third term here over $D_i(\omega)$ vanishes by the definition of $\bar{m}_i$, the first term is independent of c_i, and the second term is minimized by $c_i = \bar{m}_i$. So $\bar{e}_i = \bar{m}_i - c_i = 0$. QED.

Lemma 5.2 *Common p-belief implies 2p–1-agreeing, and p-agreeing implies common p-belief.*

Proof: Regarding the first claim, $C \subset B_i^p(C \cap E)$ implies both $C \subset B_i^p(C)$ and $C \subset B_i^p(E)$ due to the general relation that $B_i^p(S) \subset B_i^p(S')$ whenever $S \subset S'$. Regarding the second claim, for all ω in a common set C, $\mu(C \mid I_i(\omega)) \geq p$ and $\mu(E \mid I_i(\omega)) \geq p$. Defining $a_1 = \mu(C \cap E \mid I_i(\omega))$, $a_2 = \mu(C \backslash E \mid I_i(\omega))$, $a_3 = (E \backslash C \mid I_i(\omega))$, and $a_4 = 1 - a_1 - a_2 - a_3$, we thus have $a_1 + a_2 \geq p$ and $a_1 + a_3 \geq p$. This implies $a_3 + a_4 \leq 1 - p$ and $a_2 + a_4 \leq 1 - p$ which implies $1 - a_1 = a_2 + a_3 + a_4 \leq a_2 + a_3 + 2a_4 \leq 2(1 - p)$ so that $a_1 = \mu(C \cap E \mid I_i(\omega) \geq 2p - 1$ for all $\omega \in C$. QED.

Lemma 5.3 $\bar{e}_1/p_1 - \bar{e}_2/p_2 = \mathrm{E}[\tilde{X}_1 - \tilde{X}_2 \mid A] + \mathrm{E}[\tilde{X}_1 - X \mid B_1 \backslash A](1 - p_1)/p_1 - \mathrm{E}[\tilde{X}_2 - X \mid B_2 \backslash A](1 - p_2)/p_2$.

Proof: Since $e_i = \tilde{X}_i - X_i$, we have $\mathrm{E}[e_1 \mid A] - \mathrm{E}[e_2 \mid A] = \mathrm{E}[\tilde{X}_1 - \tilde{X}_2 \mid A] - \mathrm{E}[X_1 - X_2 \mid A]$. The strategy of proof is to find expressions for each of these terms, substitute them, and then solve for $\bar{e}_1/p_1 - \bar{e}_2/p_2$.

First, rearranging equation 5.1 implies that $\mathrm{E}[X_i \mid S] = \mathrm{E}[X \mid S]$ for any S which is a union of I_i members. And B_i must be a union of I_i members since $\tilde{\mathrm{E}}_{i\omega'} = \tilde{\mathrm{E}}_{i\omega}$ is the same for all $\omega' \in I_i(\omega)$. Thus $\mathrm{E}[X_i \mid B_i] = \mathrm{E}[X \mid B_i]$. Since $p_i = \mu(A \mid B_i)$, we also have

$$\mathrm{E}[f \mid B_i] = \mathrm{E}[f \mid A]p_i + \mathrm{E}[f \mid B_i \backslash A](1 - p_i) \tag{5.3}$$

for any f. Using $f = X$ and $f = X_i$, we can then solve for $\mathrm{E}[X_i \mid A] = \mathrm{E}[X \mid A] + E[X - X_i \mid B_i \backslash A](1 - p_i)/p_i$, which implies

$$\mathrm{E}[X_1 - X_2 \mid A] = E[X - X_1 \mid B_1 \backslash A](1 - p_1)/p_1 - E[X - X_2 \mid B_2 \backslash A](1 - p_2)/p_2.$$

Second, using $f = e_i$ in equation 5.3 yields $\mathrm{E}[e_i \mid A] = \bar{e}_i/p_i - \mathrm{E}[e_i \mid B_i \backslash A](1 - p_i)/p_i$. Finally, we can leave $\mathrm{E}[\tilde{X}_1 - \tilde{X}_2 \mid A] = \mathrm{E}[\tilde{X}_1 - \tilde{X}_2 \mid C \cap E]$ alone, as this must be at least ϵ by the definition of E. Substituting into the original equation, noting that $(X_i - X) + e_i = \tilde{X}_i - X$, and solving for $\bar{e}_1/p_1 - \bar{e}_2/p_2$ gives the result. QED.

Lemma 5.4 *$\epsilon \geq 0$ and $\bar{e}_2 \geq -\delta_2 \leq 0$ imply $\bar{e}_1 \geq \hat{\epsilon}(p_0, \delta_2)$.*

Proof: Since $\tilde{X}_1 - \tilde{X}_2 \geq \epsilon$ everywhere in E, and $A \subset E$, the first right side term in lemma 5.3's equation is at least ϵ. For $\epsilon \geq 0$ the most negative imaginable case for this right side is where $(B_1 \backslash A) \cap (B_2 \backslash A) = \emptyset$, with $\tilde{X}_1 = \underline{X}$ and $X = \bar{X}$ on $B_1 \backslash A$, and $\tilde{X}_2 = \overline{X}$ and $X = \underline{X}$ on $B_2 \backslash A$. This gives

$$\bar{e}_1/p_1 \geq \bar{e}_2/p_2 + \epsilon - \Delta X((1 - p_1)/p_1 + (1 - p_2)/p_2).$$

Multiplying this equation by p_1, the most negative case for the last two right side terms is $p_1 = p_2 = p_0$, and when $\bar{e}_2 = -\delta_2 \leq 0$, the most negative case for the first right side term is $p_1 = 1, p_2 = p_0$. This implies the result. QED.

Bibliography

[AD82] G.A. Ackerlof and W.T. Dickens. The economic consequences of cognitive dissonance. *American Economic Review*, 72:307–319, 1982.

[AH90] Philippe Aghion and Benjamin Hermalin. Legal restrictions on private contracts can enhance efficiency. *Journal of Law, Economics, and Organization*, 6(2):381–409, Fall 1990.

[And96] Fredrik Andersson. Income taxation and job-market signaling. *Journal of Public Economics*, 59:277–298, 1996.

[AS90] David Austen-Smith. Information transmission in debate. *American Journal of Political Science*, 34(1):124–152, February 1990.

[ASB95] David Austen-Smith and Jeffrey S. Banks. Cheap talk and burned money. Draft, April 1995.

[Aum76] Robert Aumann. Agreeing to disagree. *The Annals of Statistics*, 4(6):1236–1239, 1976.

[BA94] Kim Border and Charalambos Aliprantis. *Infinite Dimensional Analysis.* Springer-Verlag, New York, 1994.

[Bar05] John Marshall Barker. *The Saloon Problem and Social Reform.* The Everett Press, Boston, 1905.

[BCP94] Jeffrey Banks, Colin Camerer, and David Porter. An experimental analysis of nash refinements in signaling games. *Games and Economic Behavior*, 6:1–31, 1994.

[BD93] Mark J. Browne and Helen I. Doerpinghaus. Information asymmetries and adverse selection in the market for individual medical expense insurance. *Journal of Risk and Insurance*, 60(2):300–312, 1993.

[Blu96] Andreas Blume. Information transmission and preference similarity. Technical report, May 1996.

[BS93] Jeffrey S. Banks and Rangarajan K. Sundaram. Adverse selection and moral hazard in a repeated elections model. In Barnett, Hinich, and Schofield, editors, *Institutions, Competition and Representation*. Cambridge University Press, 1993.

[Bur93] Paul Burrows. Patronizing paternalism. *Oxford Economic Papers*, 45:542–572, 1993.

[CK87] In-Koo Cho and D.M. Kreps. Signalling games and stable equilibria. *Quarterly Journal of Economics*, 102:179–222, 1987.

[CL87] Jean-Paul Cresta and Jean-Jacques Laffont. Incentive compatibility of insurance contracts and the value of information. *Journal of Risk and Insurance*, 54:520–540, 1987.

[CM95] Stephen Coate and Stephen Morris. On the form of transfers to special interests. *Journal of Political Economy*, 103(6):1210–1235, 1995.

[Cou90] Peter Coughlin. Candidate uncertainty and electoral equilibria. In J. Enelow and M. Hinich, editors, *Advances in the Spatial Theory of Voting*. Cambridge University Press, Cambridge, 1990.

[Cou92] Peter Coughlin. *Probabilistic voting theory*. Cambridge University Press, Cambridge, 1992.

[CS82] Vincent P. Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, 50(6):1431–51, November 1982.

[Dah89] Robert A. Dahl. *Democracy and Its Critics.* Yale University Press, New Haven, 1989.

[Dia92] Peter Diamond. Organizing the health insurance market. *Econometrica*, 60(6):1233–1254, November 1992.

[Dow57] Anthony Downs. *An Economic Theory of Democracy.* Harper Collins, New York, 1957.

[Dow93] Anthony Downs. The origins of an economic theory of democracy. In Bernard Grofman, editor, *Information, Participation, and Choice, An Economic Theory of Democracy in Perspective*, pages 197–199. The University of Michigan Press, Ann Arbor, 1993.

[Eis95] Melvin Aron Eisenberg. The limits of cognition and the limits of contract. *Stanford Law Review*, 27:211–259, January 1995.

[ER96] Dennis Epple and Richard E. Romano. Public provision of private goods. *Journal of Political Economy*, 104(1):57–84, 1996.

[Fei85] Richard Jay Feinstein. The ethics of professional regulation. *The New England Journal of Medicine*, 312(12):801–804, March 21 1985.

[FG89] Joseph Farrell and Robert Gibbons. Cheap talk with two audiences. *American Economic Review*, 79(5):1214–1223, December 1989.

[FP96] Timothy J. Feddersen and Wolfgang Pesendorfer. The swing voter's curse. *American Economic Review*, 86(3):408–424, June 1996.

[Fra80] Joel Franklin. *Methods of Mathematical Economics.* Springer-Verlag, New York, 1980.

[FS89] Morris P. Fiorina and Kenneth A. Shepsle. Is negative voting an artifact? *American Journal of Political Science*, 33(2):423–439, May 1989.

[FS90] Morris P. Fiorina and Kenneth A. Shepsle. Negative voting: An explanation based on principal-agent theory. In John A. Ferejohn and James H. Kuklinski, editors, *Information and democratic processes*, pages 119–239. University of Illinois Press, Urbana, 1990.

[FT91] Drew Fudenberg and Jean Tilole. *Game Theory.* MIT Press, Cambridge MA, 1991.

[Gal96] Douglas Gale. Spoiled for choice: Variety and efficiency in markets with incomplete information. Technical report, April 1996.

[Gea] John Geanakoplos. Game theory without partitions, and applications to speculation and consensus. *Journal of Economic Theory.* written 1989, forthcoming.

[Gea94] John Geanakoplos. Common knowledge. In R.J. Aumann and S. Hart, editors, *Handbook of Game Theory, Volume 2*, pages 1438–1496. Elsevier Science, 1994.

[Ger95] Hans Gersbach. Information efficiency and majority decisions. *Social Choice and Welfare*, 12:363–370, 1995.

[Gie85] Dale H. Gieringer. The safety and efficacy of new drug approval. *Cato Journal*, 5(1), Spring 1985.

[Gig94] Frank Gigler. Self-enforcing voluntary disclosures. *Journal of Accounting Research*, 32(2):224–240, Autumn 1994.

[GJ95] Thomas Gehrig and Peter-J. Jost. Quacks, lemons, and self regulation: A welfare analysis. *Journal of Regulatory Economics*, 8:309–325, 1995.

[Gla87] Amihai Glazer. A new theory of voting: Why vote when millions do. *Theory and Decision*, 22:257–270, 1987.

[Gou93] Miguel Gouveia. Majority rule and public provision of health care. Technical Report 93-26, University of Pennsylvania, Center for Analytic Research in Economics and the Social Sciences, 1993.

[Gui82] Owen Guillermo. *Game Theory.* Academic, New York, 2nd edition, 1982.

[GW50] O. Gross and R. Wagner. A continuous colonel blotto game. Technical Report RM-408, RAND, Santa Monica, California, June 1950.

[GW93] Bernard Grofman and Julie Withers. Information-pooling models of electoral politics. In Bernard Grofman, editor, *Information, Participation, and Choice, An Economic Theory of Democracy in Perspective*, pages 55–64. The University of Michigan Press, Ann Arbor, 1993.

[Han85] John Mark Hanson. The political economy of group membership. *American Political Science Review*, 79:79–96, 1985.

[Han94] Robin Hanson. Consensus by identifying extremists. *Theory and Decision*, 1994. to appear.

[Hem92] David Hemenway. Propitious selection in insurance. *Journal of Risk and Uncertainty*, 5:247–251, 1992.

[Hig95] Robert Higgs. *Hazardous to Our Health? FDA Regulation of Health Care Products.* The Independent Institute, Oakland, California, 1995.

[Hin77] Melvin Hinich. Equilibrium in spatial voting: The medium voter result is an artifact. *Journal of Economic Theory*, 16:208–219, 1977.

[HL96] Hugo Hopenhayn and Susanne Lohmann. Fire-alarm signals and the political oversight of regulatory agencies. *Journal of Law, Economics, and Organization*, 12(1):196–213, April 1996.

[Hym93] David N. Hyman. *Public Finance, A Contemporary Application of Theory to Policy.* Dryden Press, New York, fourth edition, 1993.

[Ire94] Norman Ireland. On limiting the market for status signals. *Journal of Public Economics*, 53:91–110, 1994.

[Isa65] Paul E. Isaac. *Prohibition and Politics, Turbulent Decades in Tennessee, 1885-1920.* University of Tennessee Press, Knoxville, 1965.

[Jen92] Gail Jensen. Regulating the content of health plans. In R. Holmes, editor, *American Health Policy: Critical Issues for Reform*, pages 167–193. AEI Press, Washington, D.C., 1992.

[JGP+85] B.D. Johnson, P.J. Goldstein, E. Preble, J Schmeidler, D.S. Lipton, B. Spunt, and T. Miller. *Taking Care of Busines, The Economics of Crime by Heroin Abusers.* Lexington Books, Toronto, 1985.

[Kel81] Steven Kelman. Regulation and paternalism. *Public Policy*, 29(2), Spring 1981.

[Ken82] Duncan Kennedy. Distributive and paternalistic motives in contract and tort law, with special reference to compulsory terms and unequal bargaining power. *Maryland Law Review*, 41(4):563–658, 1982.

[Led84] John Ledyard. The pure theory of large two-candidate elections. *Public Choice*, 44:7–41, 1984.

[Led89] John Ledyard. Information aggregation in two-candidate elections. In Peter Ordeshook, editor, *Models of Strategic Choice in Politics*, pages 7–30. University of Michigan Press, 1989.

[Lef78] Keith B. Leffler. Physician licensure: Competition and monopoly in american medicine. *Journal of Law and Economics*, pages 165–186, 1978.

[Lel79] Hayne E. Leland. Quacks, lemons and licensing: A theory of minimum quality standards. *Journal of Political Economy*, 87(6):1328–46, 1979.

[Lev89] David Levy. The statistical basis of Athenian-American constitutional theory. *The Journal of Legal Studies*, 18(1):79–103, 1989.

[Lip95] Barton Lipman. Information processing and bounded rationality: a survey. *Canadian Journal of Economics*, 28(1):42–67, 1995.

[LM90] David A. Lambert and G. McGuire. Political and economic determinants of insurance regulation in mental health. *Journal of Health, Policy, and Law*, 15(1), Spring 1990.

[LM96] Arthur Lupia and Mathew D. McCubbins. *The Triumph of Reason: Knowledge and the Foundation of Democracy.* 1996.

[Loh94] Susan Lohmann. Information aggregation through costly political action. *American Economic Review*, 84:518–530, 1994.

[LOR95] Dean Lueck, Reed Olsen, and Michael Ransom. Market and regulatory forces in the pricing of legal services. *Journal of Regulatory Economics*, 7:63–83, 1995.

[LRT96] Renee M. Landers, James B. Rebitzer, and Lowell J. Taylor. Rat race redux: Adverse selection in the determination of work hours in law firms. *American Economic Review*, 86(3):329–348, June 1996.

[LS79] Seymour Martin Lipset and William Schneider. The public view of regulation. *Public Opinion*, pages 6–13, January 1979.

[LT93] Jean-Jacques Laffont and Jean Tirole. *A Theory of Incentives in Procurement and Regulation.* MIT Press, Cambridge, Massechusetts, 1993.

[Mau74] Alex Maurizi. Occupational licensing and the public interest. *Journal of Political Economy*, 82:399–413, 1974.

[Meg89] Nimrod Megiddo. On computable beliefs of rational machines. *Games and Economic Behavior*, 1:144–169, 1989.

[Mei91] Kenneth J. Meier. The politics of insurance regulation. *Journal of Risk and Insurance*, 58(4), December 1991.

[Mil91] Richard Lawrence Miller. *The Case For Legalizing Drugs.* Praeger, New York, 1991.

[MO86] Richard D. McKelvey and Peter C. Ordeshook. Information, electoral equilibria, and the democratic ideal. *The Journal of Politics*, 48:909–937, 1986.

[MOFP93] George Mailath, Masahiro Okuno-Fujiwara, and Anderew Poslewaite. Belief-based refinements in signaling games. *Journal of Economic Theory*, 60:241–276, 1993.

[Mor94] Stephen Morris. Trade with hetereogeneous prior beliefs and asymmetric information. *Econometrica*, 62(6):1327–1347, November 1994.

[Mor96] Stephen Morris. The logic of belief and belief change: A decision theoretic analysis. *Journal of Economic Theory*, 69:1–23, 1996.

[MP86] Richard McKelvey and Talbot Page. Common knowledge, consensus, and aggregate information. *Econometrica*, 54(1):109–127, January 1986.

[MP95] Richard D. McKelvey and Thomas Palfrey. Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10:6–38, 1995.

[MP96] Richard D. McKelvey and Thomas Palfrey. A statistical theory of equilibrium in games. *The Japanese Economic Review*, 47(2), June 1996.

[MR78] Michael Mussa and Sherwin Rosen. Monpoly and product quality. *Journal of Economic Theory*, 18:301–317, 1978.

[MS82] Paul Milgrom and Nancy Stokey. Information, trade and common knowledge. *Journal of Economic Theory*, 26:17–27, 1982.

[MS89] Dov Monderer and Dov Samet. Approximating common knowledge with common beliefs. *Games and Economic Behavior*, 1:170–190, 1989.

[MT90] E. Maskin and J. Tirole. The principal-agent relationship with an informed principal. I private values. *Econometrica*, 58:379–410, 1990.

[Mye93] Roger Myerson. Incentives to cultivate favored minorities under alternative electoral sytems. *American Political Science Review*, 87(4):856–869, December 1993.

[NBG+90] Lars Tyge Neilsen, Adam Brandenburger, John Geanakoplos, Richard McKelvey, and Talbot Page. Common knowledge of an aggregate of expectations. *Econometrica*, 58(5):1235–1239, September 1990.

[Nee96a] Zvika Neeman. Approximating agreeing to disagree results with common p-beliefs. *Games and Economic Behavior*, 12:162–164, 1996.

[Nee96b] Zvika Neeman. Common beliefs and the existence of speculative trade. *Games and Economic Behavior*, 16:77–96, 1996.

[Ols65] Mancur Olson. *The Logic of Collective Action.* Harvard University Press, 1965.

[Ord79] Janusz A. Ordover. Products liability in markets with heterogeneous customers. *Journal of Legal Studies*, 8:505–25, 1979.

[Ord86] Peter Ordeshook. *Game Theory and Political Theory, An Introduction.* Cambridge University Press, Cambridge, 1986.

[Par88] Johyun Park. *A General Strategic Information Transmission.* PhD thesis, University of California at San Diego, 1988.

[Pop91] Samuel L. Popkin. *The Reasoning Voter.* University of Chicago Press, Chicago, 1991.

[Pos95] Eric A. Posner. Contract law in the welfare state: A defense of the unconscionability doctrine, usury laws, and related limitations on the freedom to contract. *Journal of Legal Studies*, 24, June 1995.

[PS82] Durk Pearson and Sandy Shaw. *Life Extension: A Practical Scientific Approach.* Warner Books, New York, 1982.

[Rad96] Margaret Jane Radin. *Contested Commodities.* Harvard University Press, Cambridge, Massachusetts, 1996.

[Ree89] Robert Reed. Information in political markets: A little knowledge can be a dangerous thing. *Journal of Law, Economics, and Organization*, 5(2):355–374, Fall 1989.

[Ron91] Uri Ronnen. Minimum quality standards, fixed costs, and competition. *RAND Journal of Economics*, 22(4), 1991.

[Ros93] Alan Rosenthal. *The Third House, Lobbyists and Lobbying in the States.* Congressional Quarterly Press, Washington, D.C., 1993.

[RS76] Michael Rothschild and Joseph Stiglitz. Equilibrium in competitive insurance markets; an essay on the economics of imperfect information. *Quarterly Journal of Economics*, 90:629–649, 1976.

[RW90] Ariel Rubinstein and Asher Wolinsky. On the logic of 'agreeing to disagree' type results. *Journal of Economic Theory*, 51:184–193, 1990.

[Sam90] Dov Samet. Ignoring ingorance and agreeing to disagree. *Journal of Economic Theory*, 52:190–207, 1990.

[SG83] James Sebenius and John Geanakoplos. Don't bet on it: Contingent agreements with asymmetric information. *Journal of the American Statistical Association*, 78(382):424–426, June 1983.

[Sha86] Carl Shapiro. Investment, moral hazard, and occupational licensing. *Review of Economic Studies*, 53:843–862, 1986.

[Sha87] Steven Shavell. *Economic Analysis of Accident Law.* Harvard University Press, Cambridge, Massachusetts, 1987.

[Sob85] Joel Sobel. A theory of credibility. *Review of Economic Studies*, 52:557–573, 1985.

[Son95] Doron Sonsino. 'Impossibility of speculation' theorems with noisy information. *Games and Economic Behavior*, 8:406–423, 1995.

[Spe74] Michael Spence. *Market Signaling: informational transfer in hiring and related screening processes.* Harvard University Press, Cambridge, 1974.

[Spe77] Michael Spense. Consumer misperceptions, product failure, and product liability. *Review of Economic Studies*, 77:561–572, 1977.

[SS81] Avner Shaked and John Sutton. The self-regulating profession. *Review of Economic Studies*, 48:217–234, 1981.

[Ste89] Jeremy C. Stein. Cheap talk and the fed: A theory of imprecise policy annoucements. *American Economic Review*, 79(1):32–42, 1989.

[Sti71] George J. Stigler. The theory of economic regulation. *Bell Journal of Economics and Management Science*, 2(1):1–21, Spring 1971.

[Sto90] David Stockton. *The Classical Athenian Democracy.* Oxford University Press, Oxford, 1990.

[SW94] Hyun Song Shin and Timothy Williamson. Representing the knowledge of turing machines. *Theory and Decision*, 37:125–146, 1994.

[Tre93] Michael J. Trebilcock. *The Limits of Freedom of Contract.* Harvard University Press, Cambridge, 1993.

[Uhl89] Carole J. Uhlaner. Rational turnout: The neglected role of groups. *American Journal of Political Science*, 33(2):390–422, May 1989.

[VVH95] W. Kip Viscusi, John M. Vernon, and Joseph E. Harrington, Jr. *The Economics of Regulation and Antitrust.* The MIT Press, Cambridge, Mass., second edition, 1995.

[Wal91] Jack L. Walker, Jr. *Mobilizing Interest Groups in America, Patrons, Professions, and Social Movements.* The University of Michigan Press, Ann Arbor, 1991.

[Wei80] Murray Weiner. Should the public have the legal right to use unproven remedies? yes. In Louis Lasagna, editor, *Controversies in Therapeutics.* W.B. Saunders, Philadelphia, 1980.

[Wil] Robert Wilson. Reputations in games and markets. In Alvin E. Roth, editor, *Game-theoretic models of bargaining.* Cambridge University Press, Cambridge.

[Wit95] Donald Wittman. *The Myth of Democratic Failure.* University of Chicago Press, Chicago, 1995.

[WR80] Raymond Wolfinger and Steven Rosenstone. *Who Votes?* Yale University Press, London, 1980.