Discussion Paper No. 407R

Good News and Bad News: Representation Theorems and Applications

Paul R. Milgrom*

November 1979
October 1980

*J. L. Kellogg Graduate School of Management
Northwestern University
Evanston, Illinois 60201

*I am pleased to acknowledge the many useful suggestions of Sanford Grossman, Bengt Holmström, Alvin Klevorick, Roger Myerson, Mark Satterthwaite, Robert Weber, and Ward Whitt. This research was partially supported by the Center for Advanced Studies in Managerial Economics and by NSF grant SES-8001932.

ABSTRACT

This is a paper about modeling methods in information economics. A notion of "favorableness" of news is introduced, characterized, and applied to three simple models. In the equilibria of these models, (1) the arrival of good news about a firm's prospects always causes its share price to rise, (2) more favorable evidence about an agent's effort leads the principal to pay a larger bonus, and (3) buyers expect that any product information withheld by a salesman is unfavorable to his product.

1. Introduction

Information economics is the study of situations in which different economic agents have access to different information. Many kinds of institutions and patterns of behavior have been treated as attempts to cope with such informational asymmetries. For example, Spence [1973] has treated higher education as an attempt by talented workers to signal their talents to employers. Akerlof [1976] has offered a similar analysis of the "rat race," in which employees work faster than the socially optimal pace in order to distinguish themselves from less talented co-workers. Milgrom and Roberts [1979] offer a signaling analysis of the phenomenon of limit pricing, in which an established firm sets its price below the monopoly price in an attempt to discourage potential competitors. In each of these signaling models, the analysis is driven by a monotonicity property: more talented workers buy more education (Spence) or work faster (Akerlof) than their less talented counterparts, and lower cost firms set lower prices.

Monotonicity also plays a key role in models of adverse selection. For example, in the insurance market models of Rothschild and Stiglitz [1976], C. Wilson [1977], and Pauly [1974] (in which individuals know their probability of suffering a loss but the insurers do not), the individuals with the greatest likelihood of loss buy the most comprehensive insurance coverage. Similarly, in Akerlof's [1970] famous "lemons" model, higher prices in the used car market result in a higher average quality of the cars available, since owners of good cars will simply keep them if the prevailing prices are too low.

Additional examples of the role of monotonicity can be found in the literature on search, advertising, and bidding. In bidding, for example, the typical analysis proceeds on the basis of the intuition that a buyer's bid should be an increasing function of his true reservation price. (This price, of course, is known only to the buyer.) For example, see Vickrey [1961, 1962] and Ortega-Reichert [1968].

In view of the role of monotonicity in so much of information economics, it is surprising that many studies of rational expectations equilibria and of the problem of moral hazard make no use of any such property. One might guess, for example, that in a rational expectations model the arrival of good news about a firm's prospects would cause the price of its stock to rise. Such results have, unfortunately, been out of reach because no device has been available for modeling "good news." The purpose of this paper is to introduce such a device.

In the formal model treated in Section 2, there is a single, unknown, real-valued parameter θ which is of interest to a decision maker. The variable θ might represent "quality" or "intrinsic value" in a rational expectations or adverse selection model. The decision maker observes an informative signal x. Depending on the nature of θ, an appropriate signal might be an array of experimental data, a financial or geological report, a road map, a satellite photograph, or a television news show. In the absence of extra assumptions, the form that a signal takes is theoretically irrelevant to its ability to convey information.

Thinking of θ as "effort" or "ability" or "quality," I shall say that observation x is <u>more favorable than</u> observation y if for every non-degenerate prior distribution on θ the posterior corresponding to x dominates that corresponding to y in the sense of strict first order stochastic dominance. Proposition 1 characterizes this "more favorable than" relation.

If the decision maker's signal happens to be real-valued, then there is a standard concept of monotonicity which arises in the theory of estimation and hypothesis testing: the monotone likelihood ratio property (MLRP). (For example, see Ferguson [1967]). Proposition 2 asserts that a family of densities has the strict MLRP if and only if for every pair of signals x and y, $x > y$ implies that x is more favorable than y. Proposition 3 asserts that in an arbitrary information system, if signals are always comparable, then there exists a real valued function H such that $H(x)$ is a sufficient statistic for x and has the strict MLRP.

Sections 3-5 treat a series of model applications. The first of these is a simple security market model in which the announcement of good news about a security's future returns causes its price to rise.

The second application is made to a model in which a principal must design a fee schedule for his agent in an uncertain venture. The principal is unable to observe directly the effort expended by the agent, but he can observe the random profit of the venture which is influenced by the agent's effort. The agent is assumed to be risk averse and to have a reservation level of utility, reflecting his other opportunities. The principal's problem is to design a fee schedule (in which the agent's fee may depend on the profit of the venture) which trades off the necessity of providing the agent with appropriate work incentives against the desire to provide some risk-sharing.

It has been something of a puzzle in the earlier analyses of this model that the resulting fee schedule may not be increasing in the venture's profits. A condition derived by Holmström [1979] to guarantee this monotonicity is shown to be a differential calculus characterization of the MLRP. Thus, non-monotonicity in the fee schedule can arise only when higher profits can be evidence of lower effort on the part of the agent. Ruling out this implausible case resolves the puzzle.

In section 5, the representation theorems of section 2 are applied to a simple game of persuasion, in which an interested party (such as a salesman or a regulated firm) tries to influence a decision maker (such as a consumer or regulator) by selectively providing data relevant to the decision. In some variations of the model, the interested party will report only the information that is most favorable to his case. However, Proposition 6 in section 5 shows that if the decision maker knows how much information is available to the interested party and if some information is withheld, then, in equilibrium, the decision maker will suspect the worst about any withheld information. Hence, in equilibrium, the interested party reports all of his information.

2. Representation Theorems

Let Θ be a subset of $\mathbb{R}$, representing possible values of the random parameter $\tilde{\theta}$. The set of possible signals about $\tilde{\theta}$ is denoted by X which, for expositional simplicity,[1] is taken to be a subset of $\mathbb{R}^n$. Let $f(x|\theta)$ denote the conditional density (or probability mass) function on X when $\tilde{\theta}$ takes the particular value θ. With this set-up, let us say that a signal x is <u>more favorable than</u> another signal y if for every non-degenerate[2] prior distribution G for θ, $G(\cdot|x)$ dominates $G(\cdot|y)$ in the sense of strict first-order stochastic dominance.[3]

Suppose, for example, that the prior distribution G assigns probability $g(\theta)$ to some parameter value θ and $g(\theta')$ to some other value $\theta' > \theta$. By Bayes Theorem,

$$\frac{g(\theta|x)}{g(\theta'|x)} = \frac{g(\theta)}{g(\theta')}\frac{f(x|\theta)}{f(x|\theta')} \tag{2.1}$$

and a similar expression describes the posterior given y. In particular, if $g(\theta) = g(\theta') = 1/2$ and if x is more favorable than y, it follows that[4]

$$\frac{f(x|\theta)}{f(x|\theta')} < \frac{f(y|\theta)}{f(y|\theta')} \ . \tag{2.2}$$

<u>Proposition 1</u>. x is more favorable than y if and only if for every $\theta' > \theta$,

$$f(x|\theta')f(y|\theta) - f(x|\theta)f(y|\theta') > 0 \ . \tag{2.2'}$$

<u>Proof</u>: Equation (2.2') generalizes (2.2) by allowing for the possibility that $f(y|\theta') = 0$, a possibility that I shall henceforth ignore. The derivation of (2.2) constitutes the proof that it is necessary.

For sufficiency, fix some non-degenerate G and choose θ^* for which $0 < G(\theta^*) < 1$. For $\theta' > \theta^*$, it follows from (2.2) that

$$\int_{\theta \leq \theta^*} [f(x|\theta)/f(x|\theta')]dG(\theta) < \int_{\theta \leq \theta^*} [f(y|\theta)/f(y|\theta')]dG(\theta) \tag{2.3}$$

Dividing each side of (2.3) into one, and then integrating over θ' yields

$$\frac{\int_{\theta' > \theta^*} f(x|\theta')dG(\theta')}{\int_{\theta \leq \theta^*} f(x|\theta)dG(\theta)} > \frac{\int_{\theta' > \theta^*} f(y|\theta')dG(\theta')}{\int_{\theta \leq \theta^*} f(y|\theta)dG(\theta)} \tag{2.4}$$

This is equivalent to the expression:

$$[1-G(\theta^*|x)]/G(\theta^*|x) > [1-G(\theta^*|y)]/G(\theta^*|y),$$

so that $G(\theta^*|x) < G(\theta^*|y)$.

Q.E.D.

Definition. Let $X \subset \mathbb{R}$. The densities $\{f(\cdot|\theta)\}$ have the strict monotone likelihood ratio property (strict MLRP) if for every $x > y$ and $\theta' > \theta$, (2.2') holds. If the strict inequality in (2.2') is changed to a weak inequality, then the adjective "strict" is dropped from the definition.

The monotone likelihood ratio property takes its name from the fact that the likelihood ratio $f(x|\theta)/f(x|\theta')$ is monotone in x, increasing if $\theta > \theta'$ and decreasing otherwise. This property plays a major role in statistical theory, as described in most basic textbooks on the subject. Among the families of densities and probability mass functions with this property are the normal (with mean θ), the exponential (with mean θ), the Poisson (with mean θ), the uniform (on $[0,\theta]$), the chi-square (with non-centrality parameter θ), and many others.

Proposition 2. The family of densities $\{f(\cdot|\theta)\}$ has the strict MLRP iff $x > y$ implies that x is more favorable than y.

Two signals x and y are called equivalent if for every θ and θ',

$$(2.5) \qquad f(x|\theta')f(y|\theta) - f(x|\theta)f(y|\theta') = 0$$

In view of (2.1), it is apparent that equivalent signals lead to identical posterior beliefs about $\tilde{\theta}$, starting from any prior. Two signals are called comparable if they are equivalent or if one is more favorable than the other.

Families of densities with the strict MLRP have the convenient property that any two signals are comparable. The question arises: Are there any other families of densities, perhaps on other spaces X, with this comparability property?

Proposition 3. Let X be general and suppose that any two signals in X are comparable. Then there exists a function $H:X \to \mathbb{R}$ such that $H(\tilde{x})$ is a sufficient statistic for $\tilde{x}$ and such that the densities of $H(\tilde{x})$ have the strict MLRP.

Proof: Let $h:\mathbb{R} \to \mathbb{R}$ be any bounded increasing function and define

$$H(x) = \int h(\theta)dG(\theta|x) \tag{2.6}$$

where G is any non-degenerate prior for θ. Since signals are comparable, $H(x) > H(y)$ if and only if x is more favorable than y. Therefore, by Proposition 2, the densities of $H(\tilde{x})$ have the strict MLRP. Also, since $H(x) = H(y)$ iff x and y are equivalent, $H(\tilde{x})$ is a sufficient statistic.

Q.E.D.

One final definition will be useful in the first application. A signal x is called neutral if for every prior distribution G, $G = G(\cdot|x)$. From (2.1), the following is immediate.

Proposition 4. A signal x is neutral if and only if for every θ and θ'

$$f(x|\theta) = f(x|\theta') \ . \tag{2.7}$$

3. Applications: Securities Markets

The first example is a simple model of a securities market in which the public announcement of good news about the future returns on a security causes its price to rise.

Let there be two securities: a riskless security for which the return will be 1 and a risky security with the random return $\tilde{\theta}$. All investors are assumed to be identical with concave, differentiable utility functions U for wealth. Investors are endowed with one unit each of the risky and riskless securities. Clearly, no trading takes place, so setting the price of the riskless security to be one, the price, p, of the risky security can be computed from the typical investor's first order condition:

$$p = \frac{E[\tilde{\theta}\ U'(1+\tilde{\theta})]}{E[U'(1+\tilde{\theta})]}.$$

Let $g(\cdot)$ denote the density (or mass function) for $\tilde{\theta}$, and define another density g' by

$$g'(\theta) = g(\theta)U'(1+\theta)/E[U'(1+\tilde{\theta})] \ .$$

Letting E' denote expectations under the primed density, the price can be expressed as:

$$p = E'[\tilde{\theta}] \ .$$

Now suppose that some news x is publicly revealed. Then reasoning as before and applying Bayes Theorem leads to the following expressions for a new market clearing price, p(x):

$$p(x) = \frac{E[\tilde{\theta}\ U'(1+\tilde{\theta})|x]}{E[U'(1+\tilde{\theta})|x]}$$

$$= E'[\tilde{\theta}|x] \ .$$

It is plain from the second expression that more favorable news leads to a higher price for the risky security, and good news (i.e., news which is more favorable than neutral news) results in $p(x) > p$.

4. Application: Moral Hazard

In Holmström's [1979] treatment of the principal/agent problem, the agent expends effort θ to influence the profit of a venture. Let π denote the profit and let $\tilde{\alpha}$ denote the random state of nature. Realized profits depend on both θ and $\tilde{\alpha}$: $\tilde{\pi} = \pi(\tilde{\alpha},\theta)$. It is assumed that $\partial\pi/\partial\theta$ is positive, so that effort always improves profits. However, the agent dislikes expending effort; his payoff $U(x) - V(\theta)$ is an increasing function of his wealth x and a decreasing function of effort θ: $U' > 0$ and $V' > 0$. In addition, the agent is risk averse: $U'' < 0$. The principal has utility for wealth only. His payoff is denoted by $G(x)$, where $G' > 0$ and $G'' \leq 0$. A <u>fee schedule</u> or <u>sharing-rule</u> $s(\cdot)$ is a function that specifies the agent's compensation for each possible profit level of the venture. Notice that s depends <u>only</u> on π because the variables θ and $\tilde{\alpha}$ cannot be observed by the principal.

It is, of course, the fact that θ and $\tilde{\alpha}$ are unobservable that leads to the moral hazard problem. If, for example, the principal were risk-neutral (i.e., $G'' = 0$) and θ were observable, then any Pareto optimal sharing rule would involve the agent receiving a fixed fee for undertaking a specified level of effort, and the principal would bear all risks (Spence and Zeckhauser, [1971]). Since θ is assumed not to be observable, however, a contract based on a specified level of effort is not enforceable, and the agent must be given some incentive to expend effort.

In this setting, it might seem reasonable that the sharing rule should be increasing, since a rule with some decreasing segments is bad from a risk-sharing point of view, and appears to reduce the agent's effort incentive. As the following example shows, this appearance is misleading.

Let $P\{\tilde{\alpha} = 0\} = P\{\tilde{\alpha} = 1\} = .5$, let $\theta = [0, .9]$, and let $\tilde{\pi} = \tilde{\alpha} + \theta$. Then θ can be perfectly inferred from any realization π. If the principal is risk-neutral and the agent is risk-averse, then the agent's compensation in the optimal contract will depend only on θ. Thus, the agent's share when $\pi = 1.0$ can quite sensibly be smaller than his share when $\pi = 0.1$.

A plausible model in which the sharing rule is increasing results if one assumes that π has the MLRP as information about θ. To formalize this, let $f(\pi|\theta)$ denote the conditional distribution of output given effort. Assume that f is differentiable and let f_θ denote $\partial f/\partial\theta$. Holmström showed that the optimal sharing rule must satisfy the following relationship for some θ^*, b, and c $(c > 0)$.

$$\frac{G'(\pi-s(\pi))}{U'(s(\pi))} = b + c\,\frac{f_\theta(\pi|\theta^*)}{f(\pi|\theta^*)} .$$

From the concavity of U and G, it is apparent that s is increasing in π if $f_\theta(\pi|\theta^*)/f(\pi|\theta^*)$ is increasing in π. This latter condition is a local characterization of the monotone likelihood ratio property.

Proposition 5. The family $\{f(\pi|\theta)\}$ has the MLRP if and only if for every θ^*, $f_\theta(\pi|\theta^*)/f(\pi|\theta^*)$ is increasing.

Proof. Notice that $f_\theta/f = \partial \ln f/\partial\theta$. It follows that for any θ' and θ'',

$$f(\pi|\theta')/f(\pi|\theta'') = \exp\{-\int_{\theta'}^{\theta''} [f_\theta(\pi|\theta)/f(\pi|\theta)]d\theta\} .$$

The conclusion follows easily.

Q.E.D.

5. Application: The Persuasion Game

The two previous examples display routine applications of the monotone likelihood ratio property to well-known models. This powerful modeling tool can also be used to lend tractability to a whole range of new problems.

The model considered in this section is a simple version of what I call a persuasion game, in which one or more interested parties provide information to a decision maker in an attempt to influence his decision. This game can be used to model regulatory decisions, courtroom battles, and sales encounters. The kinds of questions that these games help to answer are: How effectively does an adversary system provide useful information to decision makers? When should a buyer rely on a salesman, and when should he incur costs to gather his own information?

Let us consider a simple sales encounter. The buyer's payoff depends on his estimate of the quality $\tilde{\theta}$ of the commodity being sold. Specifically, if the buyer purchases z units at price p, his payoff is

$$\tilde{\theta}F(z) - pz$$

where F is bounded, increasing, concave, and differentiable.

Let the salesman have N pieces of data about his product, $\tilde{x}_1,\ldots,\tilde{x}_N$.

The salesman may report or conceal the value of any of these variables, but he cannot misreport them. Such a feature might arise if the information is verifiable by a product demonstration, or if there are truth-in-advertising laws. The buyer, after hearing the report, must select z. The choice he makes may depend on how much information the seller conceals. To complete the specification of this model as a game, let the salesman's payoff be his "commission", αz.

A reporting strategy for the salesman in this game is a function r mapping points in $\mathbb{R}^N$ (representing possible data points) into the set of possible reports A. Now, let $g(x)$ denote the purchase quantity z that maximizes $E[\tilde{\theta}F(z) - pz|\tilde{x}=x]$, i.e., $g(x)$ is the quantity the buyer would purchase if he learned that $\tilde{x}=x$. We shall say that r is a strategy of <u>full disclosure</u> if: $r(x) = r(x') \Rightarrow g(x) = g(x')$. Thus, a reporting strategy is one of full disclosure if it withholds only inessential information. A buying strategy b is a function from A to $\mathbb{R}$, i.e., it maps possible reports into purchase decisions.

We shall study the perfect Nash equilibria of the persuasion game. The concept of perfectness, defined by Selten [1975, 1978] for general game settings, serves in this example to prevent the buyer from making "unreasonable" threats. Nevertheless, when the seller withholds information, the possibility that bad news is being withheld is not lost on the buyer.

<u>Proposition 6.</u>[5] In every perfect equilibrium of the persuasion game described above, the salesman uses a strategy of full disclosure.

<u>Proof</u>: Let r be the salesman's reporting strategy. Then, corresponding to any report r^*, the buyer's best quantity choice is $b(r^*) = F'^{-1}(p/E[\tilde{\theta}|r(\tilde{x}) = r^*])$, and so this is his quantity choice at equilibrium.

If there is some x such that $r(x) = r^*$ and $E[\tilde{\theta}|\tilde{x}=x] > E[\tilde{\theta}|r(\tilde{x}) = r^*]$, then the seller could increase his sales by reporting fully when $\tilde{x} = x$. So, in equilibrium, if $r(x) = r^*$, then $E[\tilde{\theta}|\tilde{x}=x] \leq E[\tilde{\theta}|r(\tilde{x}) = r^*]$. But it is an identity of probability theory that $E[E[\tilde{\theta}|\tilde{x}]|r(\tilde{x}) = r^*] = E[\tilde{\theta}|r(\tilde{x}) = r^*]$, so the inequality established above must be an equality. Q.E.D.

The argument offered above depends, in an essential way, on the assumptions that the buyer knows N, so that he can discern when information is being withheld, and that N is not "too large", so that the buyer can assimilate all of the information. It also assumes that it is costless for the salesman to transmit information. When any of these assumptions fail, the analysis becomes more complicated, and the monotone likelihood ratio property becomes a useful modeling tool.

Suppose, for example, that $\tilde{x}_1, \ldots, \tilde{x}_N$ are (conditional on $\tilde{\theta}$) independent and drawn from a common distribution $F(\cdot|\tilde{\theta})$ where the corresponding family of densities has the MLRP. Suppose that N is common knowledge, but the buyer can assimilate only $k < N$ pieces of data.

Proposition 7. The persuasion game described above has a perfect equilibrium in which the salesman always reports the k most favorable pieces of data.

Proof: Let the buyer's strategy be as follows. When k pieces of data $x_1, \ldots, x_k$ are reported with $x_1 \geq \ldots \geq x_k$, estimate $\tilde{\theta}$ by $e = E[\tilde{\theta}|\tilde{x}_1 = x_1, \ldots, \tilde{x}_k = x_k, \tilde{x}_{k+1} \leq x_k, \ldots, \tilde{x}_N \leq x_k]$ and buy $F'^{-1}(p/e)$ units. When the report is $(x_1, \ldots, x_j)$ for $j < k$, "assume the worst", i.e., estimate θ by $e = \inf_z E[\tilde{\theta}|\tilde{x}_1 = x_1, \ldots, \tilde{x}_j = x_j, \tilde{x}_{j+1} = z, \ldots, \tilde{x}_N = z]$, and buy $F'^{-1}(p/e)$ units. It is routine to check that this pair of strategies forms a perfect equilibrium. Q.E.D.

6. Conclusion

This paper has been devoted to formalizing the notions of good news and bad news for use in economic modeling. Three applications were studied--a model of securities markets, a model of incentive contracts, and a model of a sales encounter--in which the formal ideas were successfully applied. An extensive application of these ideas to a model of auctions is given by Milgrom [1979].

All of these developments assume that quality can be represented as a one-dimensional attribute. In subsequent work, (Milgrom and Weber [1980]), a generalization of these ideas to the case where quality is multidimensional is developed and applied to the analysis of a very general model of competitive bidding.

Footnotes

[1]I also assume for simplicity that the densities are positive everywhere. The propositions in this section are true exactly as stated for general measurable spaces and general density functions.

[2]A distribution is degenerate if it assigns probability one to a single point γ, and non-degenerate otherwise.

[3]A distribution G_1 dominates G_2 in this sense if for every θ, $G_1(\theta) \leqq G_2(\theta)$, with strict inequality for some value of θ. An equivalent statement is that for every increasing function U,

$$\int U(\theta)dG_1(\theta) > \int U(\theta)dG_2(\theta) .$$

[4]One could also define "more favorable than" by using second order stochastic dominance. A distribution G_1 dominates G_2 in this sense if for every increasing concave function U,

$$\int UdG_1 > \int UdG_2 .$$

When G has two point support, these concepts of dominance are identical; so (2.2) is necessary to conclude that x is more favorable than y in either sense. As Proposition 1 shows, it is also sufficient.

[5]A result very like this one was proved by Grossman and Hart (1980) in a study of security disclosure laws.

References

G. Akerlof [1970], "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," Quarterly Journal of Economics, 84, 488-500.

________ [1976], "The Economics of Caste, the Rat Race, and Other Woeful Tales," Quarterly Journal of Economics, 90, 599-617.

T. Ferguson [1967], Mathematical Statistics: A Decision Theoretic Approach, Academic Press, New York.

S. Grossman and O. Hart [1980], "Disclosure Laws and Takeover Bids," Journal of Finance, 35, 323-334.

B. Holmström [1979], "Moral Hazard and Observability," Bell Journal of Economics, 19, 74-91.

P. Milgrom [1979], "Rational Expectations, Information Acquisition, and Competitive Bidding," Center for Mathematical Studies in Economics and Management Science, Discussion Paper No. 406, Northwestern University. (forthcoming, Econometrica)

P. Milgrom and D. J. Roberts [1979], "Equilibrium Limit Pricing Doesn't Limit Entry," Center for Mathematical Studies in Economics and Management Science, Discussion Paper No. 399R.

P. Milgrom and R. Weber [1980], "A Theory of English Auctions," unpublished notes.

M. Pauly [1974], "Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection," Quarterly Journal of Economics, 88, 44-54.

M. Rothschild and J. Stiglitz [1976], "Equilibrium in Competitive Insurance Markets: An Essay in the Economics of Imperfect Information," Quarterly Journal of Economics, 90, 629-650.

R. Selten [1975], "Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games," International Journal of Game Theory, 4, 25-55.

________ [1978], "The Chain Store Paradox," Theory and Decision, 9, 127-159.

A. M. Spence [1973], "Job Market Signalling," Quarterly Journal of Economics, 87, 355-374.

A. M. Spence and R. Zeckhauser [1971], "Insurance, Information, and Individual Action," American Economic Review, 61, 380-87.

C. Wilson [1977], "A Model of Insurance Markets with Incomplete Information," Journal of Economic Theory, 16, 167-207.