

A Note on the Incentive to Reveal Information

by Steven Shavell *

1. Introduction and summary

There are many situations in which a party will obtain a reward or, more generally, will be treated in a way that is affected by information that he reveals, assuming that the information can be verified as true. This is ordinarily the case in goods markets; a seller (say, of a used automobile) will expect to receive a higher price if he can establish favorable facts to the buyer (for example, by showing records proving that the automobile has been regularly serviced) and to receive a lower price were he to reveal unfavorable facts. A related phenomenon occurs in insurance markets; a buyer (for example, of life insurance) will pay a lower premium if he can demonstrate (perhaps with medical records) that his risk is low. Similarly, the evidence that a person (an accident victim; a criminal defendant) substantiates before a court will influence the legal outcome (a money judgment for the accident victim; a criminal sanction for the defendant); the demonstrable reasons that a person offers for his actions (breaking a lunch engagement) in normal social intercourse will affect the attitude of others toward him. The variety of situations in which the verifiable information that a party conveys will matter to his treatment is plainly large.

Such situations will be examined here in a model in which the following assumptions are made. First, some parties possess verifiable information – they can prove that the information is valid – while other parties do not. (Some sellers of automobiles have service records and others do not; some purchasers of life insurance have reasonably complete medical records, others do not; some criminal defendants can bring forward good alibi witnesses, others cannot.) Second, the reward (hereafter, most broadly interpreted) that a party receives if he reveals verifiable information is influenced by that information. Third, the reward that a party receives if he does not reveal verifiable information – either because he does not possess it or because he possesses it but chooses to keep silent – is determined by the characteristics of the entire group of parties who do not reveal verifiable information or by rational inferences that can be drawn about them.

Under these assumptions the incentive of parties who possess verifiable information to reveal it and the reward given to the group of parties who do not reveal verifiable information are determined. To understand the conclusion, consider initially the special case

* Professor of Law and Economics, Harvard Law School. I wish to thank Lucian Bebchuk, Louis Kaplow, and A. Mitchell Polinsky for comments and the National Science Foundation (grant SES-8420226) for support.

where all parties possess verifiable information. In this case all parties will be induced to reveal their information due to a familiar “unraveling” phenomenon first discussed by Grossman [1981] and Milgrom [1981]. Namely, if a group of parties were not to reveal their information, there would have to be within this silent group a subgroup with above-the-group-average characteristics. Members of the subgroup could therefore obtain a higher reward by not keeping silent and revealing their information. They would do so, causing an unraveling of the silent group.¹ This conclusion that all parties will reveal their information seems extreme. Common observation suggests that parties with verifiable information often decide to keep silent even though there are others who also keep silent who are significantly worse than they.

That, indeed, is the nature of the conclusion in the general case of the model, where it is assumed that only some parties possess verifiable information. In this case all parties with verifiable information less favorable than a certain threshold will keep silent. In particular, the unraveling phenomenon will not lead the silent parties with verifiable information just below the threshold of favorableness to reveal their information. The reason is that they (along with other, worse silent parties) will be able to enjoy a “free ride” with the silent parties whose characteristics are better than the threshold but who are unable to prove their identity. (For example, sellers of used automobiles who possess records showing mechanical problems probably will not display these records, even if the problems indicated are relatively minor, for the price of used automobiles sold by sellers who show no records will be raised by the presence of sellers whose automobiles have no problems but who cannot prove this.²) Of course, the fact that parties with unfavorable information will choose to keep silent will itself lower the reward that the silent group receives.³

The next section of the paper demonstrates the conclusions just described and makes several remarks about them.

2. The model

Let there be a population of parties, each of which is identified by a characteristic; the characteristics are described by a probability distribution. A party has a value⁴ to a rewarding agent that depends on the party’s characteristic. Given the probability distribution over

¹ After the unraveling, only the very worst parties would be left in the silent group. They would be indifferent whether to reveal their identity since it would be deduced if they did not reveal it.

² Or a person who breaks a lunch engagement may choose not to supply a verifiable excuse (that he happened to meet and have lunch with a friend he hadn’t seen in a long time), even if it is far from the worst explanation for his behavior, preferring to give an unconfirmable but good excuse (he felt very sick) and relying on this being interpreted as possibly valid.

³ Similar conclusions to those described in this paragraph were reached by Farrell [1986] (which came to my attention after I circulated the initial version Shavell [1987] of the present paper) under the assumption that there is a fraction of parties who are ignorant of their own characteristics (rather than, as here, who are unable to establish them). Apart from this difference in assumption, the main difference between the papers is that the equilibrium value obtained by the silent group is proved here to be the minimum value of the conditional mean over that group (the minimum of $m(y)$ in the notation to be defined).

⁴ For simplicity, I shall speak of parties having values even though, under the interpretation that a party is selling an object, it is the object, not the party, that has a value.

the characteristics, the probability distribution over the values can be calculated.⁵ This latter distribution is all that we will need to consider. Define

$$\begin{aligned} v &= \text{value of a party; } v \text{ in } [a, b], a < b; \\ f(v) &= \text{probability density at } v; f \text{ is positive and continuous on } [a, b].^6 \end{aligned}$$

While all parties know their own values, only some of the parties are able to reveal verifiably their values; let

$$q(v) = \text{proportion of parties of type } v \text{ who can verifiably reveal } v; q \text{ is continuous.}$$

A party who verifiably reveals v will obtain v from the rewarding agent. The rewarding agent cannot distinguish among parties who do not and those who cannot verifiably reveal their values; these two types of parties – to be called the silent group – will thus receive the same reward.⁷ Assume that

$$y = \text{reward received by parties in the silent group.}$$

An equilibrium will be defined by two conditions. First, parties who can verifiably reveal their values will rationally decide whether to do so; that is, a party who can verifiably reveal his value will do so if $v > y$ and will not do so otherwise. Hence, the silent group given y is comprised of those who can verifiably reveal their values but for whom $v \leq y$, and of those who cannot verifiably reveal their values. The second equilibrium condition is that there is consistency in the sense that the mean of v conditional on membership in the silent group is in fact y ;⁸ that is,

$$(1) \quad y = \frac{\int_a^y v q(v) f(v) dv + \int_y^b v (1 - q(v)) f(v) dv}{\int_a^y q(v) f(v) dv + \int_y^b (1 - q(v)) f(v) dv}.$$

The two terms in the numerator correspond, respectively, to the subgroup who can verifiably reveal their values but for whom $v \leq y$ and to the subgroup who cannot verifiably reveal their values. The two terms in the denominator are analogous (the denominator appears to make the probability densities conditional densities). An equilibrium exists if and only if (1) is satisfied for some y .

⁵ Specifically, let x denote the characteristic of a party (x need not be a scalar; it could be a vector, or even a probability distribution over a set of attributes); let P be the probability measure over the possible x ; and let $v(x)$ be the value of a party given x . Then the probability that v is in a set S equals $P[x \mid v(x) \text{ is in } S]$.

⁶ The assumption that f is positive is not essential but allows us to avoid tedious details. The assumption that v is continuously distributed is made only for simplicity; the Proposition to be shown is also true in the discrete case; the proof for that case is available on request.

⁷ Members of the silent group may not literally be silent in that they may profess to have high values; but in the absence of ability to establish their claims, their claims will be assumed meaningless.

⁸ This definition of equilibrium is conventional and is similar to that employed by, for example, Spence [1974].

The interpretation of the consistency condition is that in the long run the rewarding agent will learn the mean of v in the silent group – or the rewarding agent can calculate the mean (infer it) if he knows the functions q and f – and if this conditional mean deviated from the reward he was giving to members of the silent group, he would alter the reward.

To investigate equilibrium, consider the conditional mean of v in the silent group given reward y , that is,⁹

$$(2) \quad m(y) = \frac{\frac{\int_a^y v q(v) f(v) dv}{\int_a^y q(v) f(v) dv} + \frac{\int_y^b v (1 - q(v)) f(v) dv}{\int_y^b (1 - q(v)) f(v) dv}}{\frac{\int_a^y q(v) f(v) dv}{\int_a^y q(v) f(v) dv} + \frac{\int_y^b (1 - q(v)) f(v) dv}{\int_y^b (1 - q(v)) f(v) dv}}.$$

An equilibrium will correspond to a y such that $m(y) = y$. Let us now demonstrate several facts about $m(y)$, using Figure 1 (see the graph in the typical case).

(i) $m(a) > a$ unless $q(v) = 1$ for all v ; in the latter case the denominator of (2) is zero at $y = a$, but the limit of $m(y)$ at a from the right is a , and $m(a)$ will be defined to be a in that case. From (2), we have

⁹ It may be helpful to observe that $m(y) = p(y)m_I(y) + (1 - p(y))k$, where

$$p(y) = \frac{\frac{\int_a^y q(v) f(v) dv}{\int_a^y q(v) f(v) dv + \frac{\int_y^b (1 - q(v)) f(v) dv}{\int_y^b (1 - q(v)) f(v) dv}}{\frac{\int_a^y q(v) f(v) dv}{\int_a^y q(v) f(v) dv + \frac{\int_y^b (1 - q(v)) f(v) dv}{\int_y^b (1 - q(v)) f(v) dv}},$$

$$m_I(y) = \frac{\frac{\int_a^y v q(v) f(v) dv}{\int_a^y q(v) f(v) dv}}{\frac{\int_a^y q(v) f(v) dv}{\int_a^y q(v) f(v) dv + \frac{\int_y^b (1 - q(v)) f(v) dv}{\int_y^b (1 - q(v)) f(v) dv}}.$$

and

$$k = \frac{\frac{\int_y^b v (1 - q(v)) f(v) dv}{\int_y^b (1 - q(v)) f(v) dv}}{\frac{\int_y^b (1 - q(v)) f(v) dv}{\int_y^b (1 - q(v)) f(v) dv}}.$$

Thus $m(y)$ may be expressed as a weighted average of the conditional mean $m_I(y)$ of values among those who can verifiably reveal their values and for whom $v \leq y$, and of the conditional mean k of values among all those who cannot verifiably reveal their values.

$$(3) \quad m(a) = \frac{\int_a^b v(1-q(v))f(v)dv}{\int_a^b (1-q(v))f(v)dv}.$$

which clearly exceeds a if it is well defined, that is, if the denominator is positive. The denominator will be zero only if $q(v)$ is identically one. In that case, however, $m(y)$ for $y > a$ equals

$$(4) \quad \frac{\int_a^y vf(v)dv}{\int_a^y f(v)dv},$$

whose limit from the right as y approaches a is, by L'Hôpital's rule, the limit of $yf(y)/f(y) = y$, or a .

(ii) $m(b) = E(v) < b$, where E is the expectation operator. From (2), we have

$$(5) \quad m(b) = \int_a^b vf(v)dv < b.$$

(iii) If $m(y) > y$ and $q(y) > 0$, then $m'(y) < 0$; if $m(y) < y$ and $q(y) > 0$, then $m'(y) > 0$; and if $m(y) = y$ or if $q(y) = 0$, then $m'(y) = 0$: Since the denominator of the derivative of (2) with respect to y is positive, the sign of the derivative is determined by the sign of its numerator. The numerator of the derivative of (2) is

$$(6) \quad yq(y)f(y) \left[\int_a^y q(v)f(v)dv + \int_a^b (1-q(v))f(v)dv \right] \\ - q(y)f(y) \left[\int_a^y vq(v)f(v)dv + \int_a^b v(1-q(v))f(v)dv \right].$$

But this can be verified to equal

$$[y - m(y)] [q(y)f(y)] \left[\int_a^y q(v)f(v)dv + \int_a^b (1-q(v))f(v)dv \right],$$

from which the assertions in (iii) follow. The assertions are intuitively obvious: if lower values than the current conditional mean are introduced in the set over which the conditional mean is computed (that is, if $y < m(y)$ and $q(y) > 0$), the conditional mean must fall; and conversely if higher values are introduced.

Using (i) – (iii), we will show

Proposition. There exists a unique equilibrium. In this equilibrium

(a) any party who can verifiably reveal his value will do so if his value exceeds y^* , the reward to those who are silent; other parties will be silent.

(b) The reward to the silent y^* equals the minimum value of the conditional mean $m(y)$ defined by (2).

(c) The reward to the silent y^* will be less than or equal to the unconditional mean $E(v)$, so that there will always be some parties who will reveal their values (assuming there are some parties capable of doing so with values in $(y^*, b]$).

(d) The reward to the silent y^* will exceed the minimum possible value a unless all parties are able to reveal their values, so that, with this exception, there will always be some parties who can verifiably reveal their values but who choose not to (assuming that there are some parties who can verifiably reveal their values with values in $[a, y^*)$).

Proof. From (i) and (ii) we know that $m(y) - y$ is non-negative at a and negative at b ; this and the fact that $m(y) - y$ is continuous imply that the graphs of $m(y)$ and y cross at some y^* in $[a, b]$. Thus, an equilibrium exists, at which (a) is true by definition. Also, (i) implies that (d) is true.

To show that the equilibrium is unique, assume otherwise. Clearly, there cannot be any interval over which $m(y) = y$: by (iii), $m'(y) = 0$ would be true over the interval, implying that m was constant over the interval, a contradiction. Hence, there must exist two points $y_1 < y_2$ at which the graphs of $m(y)$ and y cross and with no other such points between them.¹⁰ Since, by (iii), $m'(y_1) = 0$, and since there are no crossing points between y_1 and y_2 , the graph of $m(y)$ must lie underneath that of y in (y_1, y_2) . Since, therefore, $m(y)$ crosses y at y_2 from below, $m'(y_2) \geq 1$. But this is a contradiction since, by (iii), $m'(y_2) = 0$.

To show (b), that $m(y)$ is minimized at y^* , observe that since $m(b) < b$ and y^* is the unique crossing point, the graph of $m(y)$ must lie underneath y in $(y^*, b]$; and by (iii), $m'(y) \geq 0$ in this interval. Similar reasoning implies that $m'(y) \leq 0$ in $[a, y^*)$. Thus, $m(y)$ must be minimized at y^* .

Also, since the minimum of $m(y)$ occurs at y^* , we know in particular that $m(y^*) \leq m(b)$. But since $y^* = m(y^*)$ and, by (ii), $m(b) = E(v)$, we have $y^* \leq E(v)$, proving (c). Q.E.D.

3. Remarks

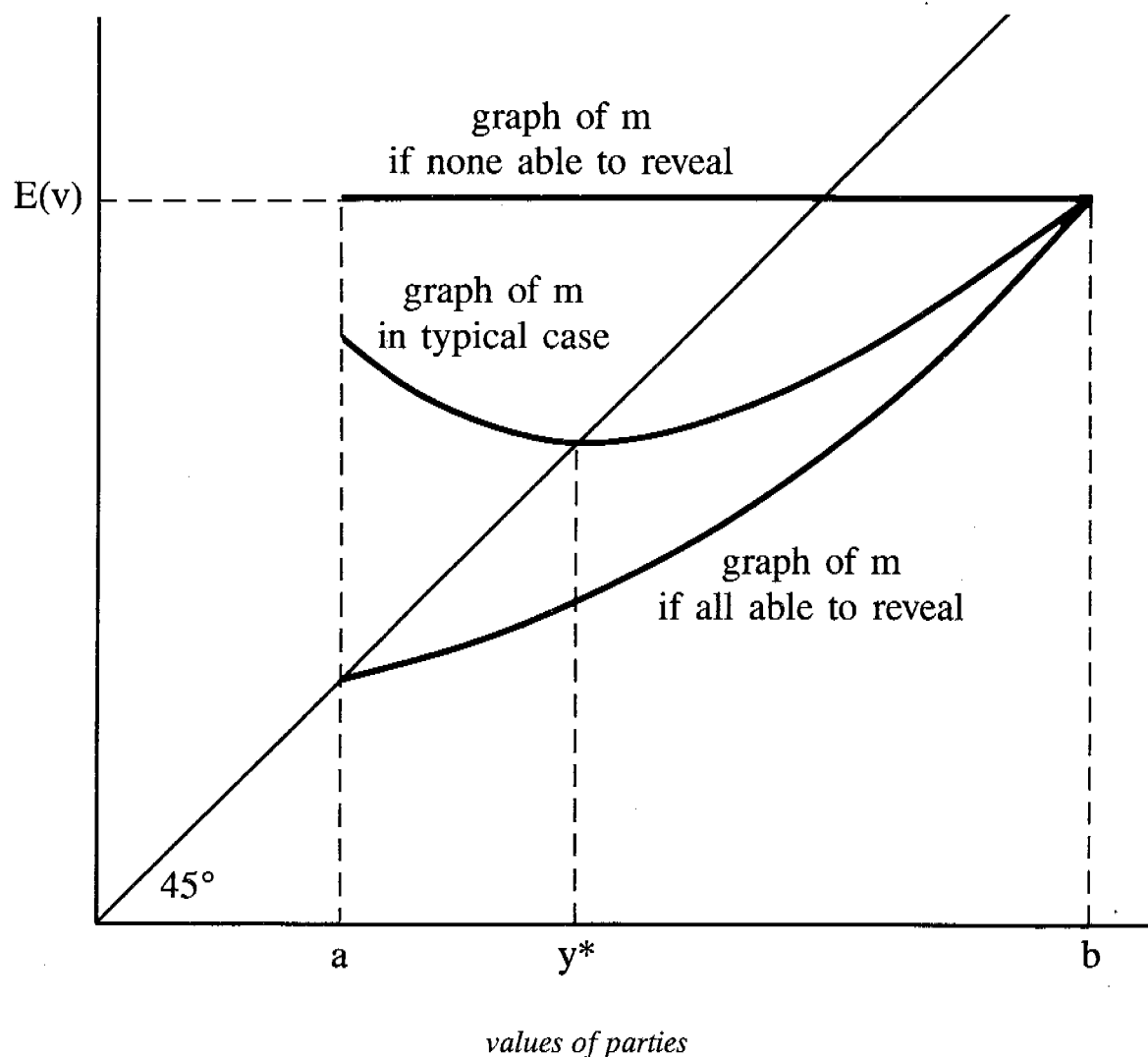
1. That the equilibrium occurs at the minimum of $m(y)$ and that $m(y)$ falls and then rises allows the equilibrium to be calculated readily by numerical methods. In addition, it may allow closed-form solutions to be obtained. For example, suppose that v is uniformly

¹⁰ To prove that the claimed y_1 and y_2 exist, observe that the set, say Z , of zeros of $m(y) - y$ is closed (since $m(y) - y$ is a continuous function); and that by hypothesis Z contains no intervals but more than one point. Let $y^\alpha < y^\beta$ be two points in Z . Since Z contains no intervals, there must be some point y° not in Z with $y^\alpha < y^\circ < y^\beta$. Let y_1 be the supremum of the y in Z which are less than y° (this set is nonempty since it contains y^α) and let y_2 be the infimum of the y in Z which exceed y° (this set is nonempty since it contains y^β). Now since Z is closed, y_1 and y_2 are members of Z and $y_1 < y^\circ < y_2$. Also, by construction, there are no elements of Z in (y_1, y_2) . We have therefore shown that the claimed y_1 and y_2 exist.

distributed over $[0,1]$ and that $q(v) = q$ over the interval. In this case, from (2), $m(y) = [qy^2/2 + (1-q)/2] / [qy + (1-q)]$; $m'(y) = 0$ is equivalent to $q^2y^2 + 2(q-q^2)y - (q-q^2) = 0$; and using the formula for the positive root of a quadratic equation, we obtain that $y^* = [(1-q)^{1/2} - (1-q)]/q$. Thus, for instance, if $q = .2$, then $y^* = .47$; if $q = .4$, $y^* = .44$; if $q = .6$, $y^* = .39$; and if $q = .8$, $y^* = .31$; note that y^* decreases as the fraction q who can verifiably reveal their values increases (since there is then less free riding).

2. The special case mentioned in the Introduction is a special case of the Proposition. The case where all parties can verifiably reveal their values is that where $q(v) = 1$ for all v . In this case, $m(a) = a$, so at the equilibrium $y^* = a$, and all parties thus reveal their values; there is complete unraveling. The case in which no parties can reveal their values is also a special case of the Proposition. In this case $q(v) = 0$ for all v ; hence $m(y)$ is a constant equal to the unconditional mean $E(v)$, which is thus also the equilibrium y^* . Both cases are illustrated in Figure 1.

Figure 1: Graph of the function $m(y)$



3. The assumption that the reward to the silent group equals the conditional mean of values within the silent group is not necessary to the main result, that some parties with verifiable information will choose not to reveal it. It is clear that this would remain the case were the reward to the silent group a more complicated function of the conditional probability distribution than the mean, as long as the function was increasing in some sense in high values (for that is what makes free riding valuable).¹¹

4. The model studied here can easily be extended to allow for the fraction $1 - q(v)$ of parties who had been assumed completely unable to reveal verifiably their values to be able to do so at a cost c (as where, for instance, a seller of a used automobile pays a fee for a mechanic's report). In this case such parties will spend c and reveal verifiably their values when $v > y + c$. The consistency condition (2) will therefore become

$$(7) \quad y = \frac{\int_a^y v q(v) f(v) dv + \int_a^{\min(y+c, b)} v (1-q(v)) f(v) dv}{\int_a^y q(v) f(v) dv + \int_a^{\min(y+c, b)} (1-q(v)) f(v) dv}.$$

Setting $m(y)$ equal to the right hand side of (7), we have, as before, that $m(a) \geq a$ and that $m(b) < b$, so that, by continuity, there exists a y^* where $m(y^*) = y^*$; this is an equilibrium (although it need not be unique¹²). The equilibrium resembles that which we studied, since there is free riding on parties who have values above y^* but less than $y^* + c$.

5. The equilibrium outcome in the model studied here departs from the first-best outcome because, obviously, the rewarding agent does not have information from some parties, meaning that there may be resulting "efficiency losses" due to the rewarding agent's inability to use the information. (For example, if the purchaser of a used automobile does not know that its brakes are bad, he may not be properly cautious in driving; if an insurer does not know the risk class of an insured and he is low risk, he will purchase too little coverage; if the court does not know that a person is truly very culpable, the court may not impose a properly high sanction.)

The equilibrium outcome also differs, in two respects, from the following second-best outcome: that which would be commanded by a dictator who had the power to order parties to reveal information if it could be verifiably revealed, or, referring to the extension to the model noted in paragraph 4, to order parties to spend to reveal verifiable information.

¹¹ In some contexts it is not necessarily natural to assume that the reward to the silent group would equal the conditional mean. For instance, it is not obvious that it would be optimal for a court to impose on a criminal about whom there are some unknown facts a sanction equal to the mean of the sanctions it would impose on otherwise similar criminals about whom all facts are known.

¹² This is because the function $m(y)$ need not fall and then rise.

Under the second-best outcome, all parties who could costlessly reveal verifiable information would be told to do so. But, as we saw, in equilibrium parties who can costlessly reveal verifiable information will not do so if it is less favorable than a threshold. (This suggests a justification, *ceteris paribus*, for a legal rule requiring disclosure of unfavorable verifiable information.)

The other departure from second-best optimality is that parties with favorable information may have too great an incentive to spend to reveal it verifiably. By doing so they distinguish themselves from parties with unfavorable information, even though the efficiency gains from their having done so may be smaller than their private gains. (Consider the case where used automobiles will have different values to buyers only because they will last different lengths of time, not because this information can be used to advantage by buyers. Here, under the second-best outcome, no sellers would be instructed to spend on mechanics' reports since that would not result in efficiency gains. Nevertheless, sellers with very good automobiles would have a substantial incentive to spend on mechanics' reports to obtain a higher price.) This departure from optimality is connected to the first, since the more parties with unfavorable information who keep silent (the first departure), the lower is the reward for silent parties, and the greater is the incentive of a party with favorable information to spend to reveal verifiably that information.

REFERENCES

- FARRELL, J. [1986]: "Voluntary disclosure: Robustness of the unraveling result, and comments on its importance", chapter 4 in *Antitrust and Regulation*, R. Grieson (Ed.), Lexington, Lexington (Mass.), 91-103.
- GROSSMAN, S. [1981]: "The informational role of warranties and private disclosure of product quality", *Journal of Law and Economics*, 24 (December 1981), 461-484.
- MILGROM, P. [1981]: "Good news and bad news: Representation theorems and applications", *Bell Journal of Economics*, 12 (Autumn 1981), 380-391.
- SHAVELL, S. [1987]: "A note on the incentive to reveal information", Discussion Paper No. 25 (February 1987), Program in Law and Economics, Harvard Law School, Cambridge (Mass.).
- SPENCE, M. [1974]: *Market Signalling*, Harvard, Cambridge (Mass.).