Settlement Delay as a Sorting Device

THOMAS J. MICELI
University of Connecticut, Storrs, Connecticut
E-mail: Miceli@UConnvm.UConn.edu

I. Introduction

This paper examines the use of sorting by defendants as a strategy for lowering their overall costs of litigation. The effectiveness of this strategy relies on the defendant’s ability to bundle settlement offers with some other variable that is valued differently by plaintiffs. The example I examine in this paper is settlement delay. For example, if plaintiffs have different time costs of pursuing a case, then the defendant can induce those plaintiffs with higher time costs to accept lower settlement offers now, while promising higher offers at a later date. The analysis thus extends the basic asymmetric information model of settlement by allowing the defendant to dictate not only the dollar amount of the settlement offer, but also the time at which it will be paid, where the latter can range anywhere from the time the dispute arises up to the (fixed) date of the trial. The model uses standard self-selection techniques\(^1\) to derive the conditions under which sorting is profitable for the defendant.

The analysis also contributes to the growing literature, both theoretical and empirical, that examines the factors affecting the timing of the settlement of legal disputes.\(^2\) Interest in this topic is motivated in part by a desire to find better policies for facilitating settlement, thereby reducing the social costs associated with delay. The first model to introduce time explicitly into the economic model of the settlement-trial decision was by Spier (1992).\(^3\) Her model showed that the highest probabilities of settlement are at the beginning and the end of the negotiation process, a pattern that is consistent with observations of the actual settlement process [Spier (1992), p. 93].

Fournier and Zuehlke (1996) developed an empirical model, loosely based on Spier’s theoretical framework, of the causes of settlement delay. They found that several of the comparative statics that arise from static models of settlement continue to hold in a

---

\(^1\)See, e.g., Chiang and Spatt (1982), Sappington (1983), Cooper (1984), Maskin and Riley (1984), and Salant (1989).
\(^3\)Gravelle (1990) introduced time as a way of modeling the costs of delay in \textit{trial}, but in his model, all cases that settle do so immediately.
dynamic context. In particular, (1) an increase in the stakes of a case reduces the likelihood of settlement, (2) an increase in the costs of litigation increases the likelihood of settlement, and (3) fee shifting (i.e., a switch from the American to the English rule for allocating legal costs) causes a greater delay in settlement, though this effect declines over time. Kessler (1996) also empirically examined the causes of delay in settlement, with a focus on institutional factors. He found that (1) greater delays in trial courts translate into greater delays in settlement, (2) comparative negligence increases delay in settling tort cases, and (3) state laws aimed at reducing delays in settlement (specifically, the imposition of prejudgment interest on awards to victorious tort plaintiffs) may actually have the effect of increasing delay.

The analysis in this paper proposes sorting as a complementary explanation for delays in settlement. A prediction of the sorting model is that settlement amounts are an increasing function of the length of the delay in payment. Ross (1970) described such a pattern in his study of the settlement of automobile accident claims by insurance companies. In particular, he noted that “in selected cases delay may well be a tool of considerable power, and on occasions it may well be used consciously to lower the settlement” [Ross (1970), p. 85]. For example, one claims adjuster made the following threat to a claimant: “You can settle with me right now . . . , or if you don’t want to accept this [offer] you have the right to go further . . . and eventually you might get more” [Ross (1970), p. 155]. The purpose of this paper is to derive the conditions under which this strategy is cost minimizing for the defendant and to examine its impact on overall litigation costs. An important conclusion is that delay does not necessarily result in higher social costs when compared to a world in which delay is not possible. Specifically, delay is beneficial if it allows cases to settle that the defendant otherwise would have taken to trial.

Section II sets up the basic asymmetric information model that will be used to examine delay as a sorting device, and Section III derives the conditions under which sorting is optimal for the defendant. Section IV then asks whether sorting increases or decreases the social costs of litigation compared to a world in which delay in settlement can be prevented. Finally, Section V concludes.

II. The Model

Consider a risk-neutral plaintiff who has an expected value of trial equal to \( V = Pj - C_p \), where \( P \) is the probability of victory at trial, \( j \) is the expected judgment in the event of victory, and \( C_p \) is the plaintiff’s cost of trial. Let the pretrial period be of fixed length \( T \), commencing with the occurrence of an accident at \( t = 0 \) and ending on the eve of a trial at \( t = T \). To keep the model simple, I assume that all negotiations between the plaintiff and defendant take place immediately after the accident (i.e., at \( t = 0 \)), even though the result can be a settlement under which the defendant agrees to make a payment to the plaintiff in any period from \( t = 0 \) up to \( T \). In effect, the defendant offers a set of “contracts” at time \( t = 0 \), consisting of \((S, t)\) pairs, where \( S \) is the dollar payment and \( t \) is the time in which it will be paid, \( 0 \leq t \leq T \). Further, the contracts are made on a take-it-or-leave-it basis. This assumes that defendants can credibly commit not to reopen

4The fixed trial date reflects an implicit assumption that plaintiffs file suit immediately on occurrence of an accident. See Spier (1992).
bargaining after plaintiffs have selected their contracts. Plaintiffs then select their utility-maximizing contract at \( t = 0 \), and no subsequent bargaining occurs.

The defendant’s use of delay as a potential sorting mechanism is made possible by the assumption that plaintiffs incur a cost of waiting for the receipt of a settlement, which varies across plaintiffs. This cost can include the plaintiff’s opportunity cost of time plus the cost of retaining a lawyer during the bargaining period. Given the cost of delay, a plaintiff who settles for an amount \( S \) to be paid in period \( t \) receives a payoff of \( S - c_{jT} \) evaluated as of time zero, where \( c_{j} \) is the unit time cost for a plaintiff of type \( j \). For simplicity, assume that there are two types of plaintiffs, those with high costs, \( c_{H} \), and those with low costs, \( c_{L} \), where \( c_{H} > c_{L} \). Thus, for any \( t > 0 \), total waiting costs are higher for type \( H \)’s compared to type \( L \)’s.

A type \( j \) plaintiff (\( j = H, L \)) who goes to trial receives a payoff of \( V - c_{jT} \) evaluated as of \( t = 0 \). (Note that \( V - c_{j} T \geq 0 \) for the plaintiff to pursue the case; I assume this holds for both types.) \( V - c_{j} T \) thus corresponds to the reservation price of the plaintiff as of \( t = 0 \). That is, any settlement offer \( (S, \delta) \) that is acceptable to the plaintiff must satisfy \( S - c_{j} \delta \geq V - c_{j} T \), or \( S \geq V - c_{j}(T - \delta) = S_{jT}^{0} \). Figure 1 graphs the reservation price lines of the two types of plaintiffs, where \( c_{H} > c_{L} \) implies that the line for low-cost plaintiffs is above that for high-cost plaintiffs for all \( T < T \); but they intersect at \( t = T \). Any

---

\( ^{5} \)On the commitment problem, see Ross [(1970), pp. 156–158]. The need to assume that contracts will not be renegotiated is a characteristic of all asymmetric information models, because the informed party’s choice of a contract reveals information to the uninformed party. I will comment on this more specifically below. (See text at note 14 infra.)

\( ^{6} \)In general, the analysis would go through if plaintiffs differ in any known way that makes their present value of trial differ (while keeping the same \( V \)). An obvious alternative to different time costs would be different discount rates.

Appendix A examines this specification in more detail.

\( ^{7} \)The analysis would be unaffected if waiting costs were convex in \( t \).

\( ^{8} \)These two characteristics of the \( S_{jT}^{0} \) curves are sufficient to allow sorting in the current model. Specifically, note that
settlement offer \((S, t)\) that lies on or above a given plaintiff’s line will be acceptable, and any offer below it will result in a trial.

The problem for a defendant facing a population of plaintiffs of unknown type is to offer a menu of \((S, t)\) pairs to minimize his total expected cost, including settlements and trials, subject to his uncertainty about individual plaintiffs’ types and the constraint that plaintiffs choose the settlement offer from among those available that maximizes their expected return (self-selection). Before examining the characteristics of the optimal menu of offers, it is necessary to specify the defendant’s costs. First, consider the cost of settlement. If the defendant makes a settlement offer of \((S, t)\) that is accepted by a single plaintiff, the defendant’s cost for that plaintiff is \(S + kt\), where \(k\) is the defendant’s (constant) marginal “cost” of delay. A positive value of \(k\), implies that the defendant would rather settle sooner than later, whereas a negative value of \(k\) implies that the defendant prefers delay. There are offsetting effects. On the one hand, if the defendant discounts the future, he would prefer delay [Spier (1992), p. 94], but on the other, the cost of carrying cases creates pressure to dispose of cases in an expeditious manner.\(^9\) In what follows, I will assume that \(k \geq 0\), though it will be seen below that this assumption is not essential to obtain the results. Note that for cases that settle, the defendant has iso-cost curves that have a slope of \(-k\) in \((S, t)\) space. Thus, if \(k > 0\), the curves are negatively sloped with costs decreasing to the lower left in Figure 1, and if \(k < 0\), they are positively sloped with costs decreasing to the lower right.

Consider next the cost of a trial to the defendant. As of time zero, the cost of trial with a plaintiff whose value of trial is \(V\) equals \(V + x + kT\), where \(x\) is the defendant’s “excess” cost of taking a case to trial,\(^10\) and \(kT\) is his cost of waiting \(T\) periods for trial. Note that the defendant’s costs do not depend on plaintiffs’ costs of delay, except insofar as they permit sorting.

III. The Defendant’s Optimal Use of Delay

If the defendant could observe plaintiffs’ types, he would offer to settle immediately with both types for their reservation prices. If \(\alpha\) is the fraction of low-cost plaintiffs, the defendant’s resulting full information cost would be

\[
TC^F = \alpha(V - c_L T) + (1 - \alpha)(V - c_H T)
= V - [\alpha c_L + (1 - \alpha) c_H]T, \tag{1}
\]

which represents the lowest possible cost of resolving disputes. This outcome is not attainable under uncertainty, however, because high-cost plaintiffs would always prefer the offer intended for low-cost types given that \(V = c_L T > V = c_H T\). The result in this case would be a “pooling” equilibrium under which all plaintiffs would settle for the

\(^9\)For example, Ross [(1970), p. 85] notes that “in the ranks of field claims men and supervisors as well as among home office executives there is a great concern for closing files expeditiously.”

\(^{10}\)To be more specific, if the defendant’s expected cost of trial on the eve of trial be \(PJ + C_p\), where \(C_p\) is his litigation cost and \(P\) and \(J\) are defined as above. Adding and subtracting \(C_p\), the plaintiff’s litigation cost, to this expression yields \(PJ + C_p + C_p - C_p = V + C_p + C_p\), where, recall, \(V = PJ - C_p\). It follows that \(x = C_p + C_p\) the joint costs of trial to the two parties.
reservation price of low-cost types (point \( A \) in Figure 1). The cost for the defendant in this case would be

\[ TC^P = V - c_L T. \]  

(2)

In standard asymmetric information models of settlement,\(^{11}\) the defendant chooses between this pooling equilibrium and a "separating" equilibrium under which he settles immediately with high-cost plaintiffs for the latter’s reservation price (point \( B \)) and takes low-cost plaintiffs to trial. In this case, the trial serves as a sorting device. The defendant’s cost under this strategy is

\[ TC^S = \alpha(V + x + kT) + (1 - \alpha)(V - c_H T) \]

\[ = V + \alpha(x + kT) - (1 - \alpha)c_H T. \]  

(3)

The separating strategy is preferred to the pooling strategy if \( TC^S < TC^P \), or if

\[ \alpha < \left( \frac{(c_H - c_L)T}{x + (k + c_H)T} \right) = \alpha_1, \]  

(4)

where \( \alpha_1 \) is between zero and one. According to this condition, the separating strategy is more likely to dominate the smaller: the fraction of low-cost plaintiffs (\( \alpha \)), the cost of trials (\( x \)), and the defendant’s unit waiting cost (\( k \)). The separating strategy is also more likely to dominate as the difference in waiting costs between the two types of plaintiffs increases.\(^ {12}\)

Now consider the defendant’s use of delay rather than trial as a way to implement a separating strategy. In particular, suppose that the defendant offers two “contracts” at \( t = 0 \): one involving an immediate payment equal to the high-cost plaintiff’s reservation price, \( V - c_H T \) (point \( B \)), and the other promising a payment of \( S \) dollars at a later date \( t \leq T \). A separating equilibrium requires that the defendant set \( S \) and \( t \) so that high-cost plaintiffs choose the immediate settlement offer and low-cost plaintiffs choose the delayed offer. In addition, the delayed offer \( (S, t) \) must be at least as good as going to trial for low-cost plaintiffs as of \( t = 0 \) (otherwise, the outcome would be identical to the above separating equilibrium with trials). As Figure 1 shows, point \( C \) is the only delayed offer that satisfies these constraints. Note that this point represents a promise by the defendant to pay low-cost plaintiffs \( S = V \) dollars (the value of trial at \( t = T \)) on the courthouse steps.\(^ {13}\)

Total costs for the defendant under this separating equilibrium are given by

\[ TC^D = \alpha(V + kT) + (1 - \alpha)(V - c_H T) \]

\[ = V + \alpha kT - (1 - \alpha)c_H T. \]  

(5)

Note first that costs are lower under this strategy as compared to costs under the separating equilibrium with trials [expression (3)] because the current outcome avoids trial costs. The question is whether separation by delay saves costs for the defendant


\(^{12}\)The impact of \( T \) is ambiguous because it enters positively in both the numerator and denominator of \( \alpha_1 \).

\(^{13}\)I assume that when indifferent between settlement and trial, plaintiffs opt for settlement. It should be clear from Figure 1 that under this equilibrium, high-cost plaintiffs are indifferent between settling at \( t = 0 \) or at \( t = T \), whereas low-cost plaintiffs strictly prefer delayed settlement to settlement at \( t = 0 \). Further, points \( B \) and \( C \) together represent the lowest cost separating equilibrium attainable by the defendant.
compared to the pooling outcome. Comparison of (5) and (2) shows that separation is preferred if and only if
\[ \alpha < \left[ (c_H - c_L) T \right] / \left[ (k + c_H) T \right] = (c_H - c_L) / (k + c_H) = \alpha_2. \]  
This condition is identical to (4) except for the absence of \( x \) in the denominator of the right-hand side in (6). \( \text{(Thus, } \alpha_2 > \alpha_1) \) According to (6), separation is more desirable the smaller are \( \alpha \) and \( k \), and the larger is the waiting cost differential between the two types of plaintiffs.

Note that the settlement pattern that emerges from the separating equilibrium with delay fits the twin observations that a large percentage of cases settle close to trial [Spier (1992)], and that the settlement amount is increasing in the length of delay (all else equal) [Ross (1970)]. It is important to emphasize, however, that this pattern did not arise in the current model as a result of sequential bargaining throughout the pretrial period but, instead, reflects the optimal menu of take-it-or-leave-it contracts offered by the defendant immediately after the accident. It is now easy to see why the defendant needs to make a credible commitment not to renegotiate these contracts once plaintiffs have made their choices. As soon as low-cost plaintiffs refuse the immediate offer to settle, they reveal their type with certainty. At that point, the defendant has an incentive to settle with them immediately for their reservation price, \( V^L = c_L T \), thus achieving the perfect information outcome. If high-cost types anticipate this, however, they will refuse to accept the immediate offer of \( V^H = c_H T \), thereby upsetting the equilibrium. The theory of delay in this paper therefore relies on the defendant’s ability to make credible commitments. Thus, it is likely that such a strategy can be used only by repeat defendants or by those who can otherwise develop reputations for making binding commitments [Ross (1970), pp. 156–158].

IV. The Impact of Sorting on Social Costs

Most recent studies of settlement delay have focused on its social costs. This suggests that if delay could somehow be prevented, there would be a net social gain. In this section, I ask whether the sorting model provides any basis for suggesting that delay might actually save social costs compared to a world in which delay is prohibited.

Social costs in the model consist only of the costs of delay and, in the case of the first separating equilibrium, the cost of trial. Thus, in the pooling equilibrium, social costs are zero \( (\text{SC}_P = 0) \) because all cases settle immediately. In the separating equilibrium with trials, social costs consist of the waiting and trial costs for those cases involving low-cost plaintiffs:
\[ \text{SC}_S = \alpha \left[ x + (k + c_L) T \right]. \]  

---

14On the ability of repeat defendants to develop a reputation for binding commitments, see Miceli (1993).
15Thus, trials only occur in the current model when defendants cannot credibly separate by delay, and separation by trial is preferred to settling all cases (i.e., \( \alpha < \alpha_1 \)).
17This measure of social costs treats the number of cases as fixed. A more complete analysis would recognize that the number of cases is likely to vary with different equilibrium outcomes, due, for example, to differential deterrence effects. For models that incorporate these broader issues, see Polinsky and Rubinfeld (1988), Cooter and Rubinfeld (1989), and Shavell (1997).
Finally, in the separating equilibrium with delay, social costs consist only of waiting costs, again in cases involving low-cost plaintiffs:

\[ SC^D = \alpha (k + c_L) T. \]  

(8)

The social costs incurred in equilibrium depend on the defendant’s optimal strategy. When separation by delay is possible, recall that the defendant’s choice between separation and pooling is determined by condition (6): When \( \alpha < \alpha_2 \) the defendant opts for delay, and when \( \alpha > \alpha_2 \), he pools. The resulting social costs are graphed in Figure 2 as a function of \( \alpha \). In the separation range, costs increase linearly with \( \alpha \) (the lower, positively sloped line), and in the pooling range they drop to zero.

Now suppose that delayed settlement is somehow prevented\(^1\); all cases either settle immediately or go to trial at time \( T \). Defendants now choose between the pooling solution, in which all cases settle immediately, and a separating equilibrium, in which low-cost plaintiffs go to trial with delay. We saw above that, given this choice, the defendant will separate if \( \alpha < \alpha_1 \) and pool if \( \alpha > \alpha_1 \). The resulting social costs are drawn in Figure 2. Again, costs increase linearly with \( \alpha \) up to \( \alpha_1 \) (the upper line) and then drop to zero. Social costs of separation are higher in this case compared to separation by delay because of the addition of trial costs, but offsetting this saving from delay is the fact that the possibility of delay expands the range over which the defendant chooses the separating strategy, reflected by the fact that \( \alpha_2 > \alpha_1 \).

---

\(^1\)For example, suppose that delayed settlements (but not trials) are taxed at a rate \( \tau \) per period after the occurrence of the accident. The defendant’s cost of delay in (5) therefore becomes \( (k + \tau)T \). If \( \tau \) is large enough, then the defendant will prefer to use trial as the sorting device over delay. Specifically, comparing (5) and (3) shows that \( TC^D < TC^P \) if \( \tau < \frac{x}{T} \) or \( \tau > \frac{x}{T} \).

FIG. 2. The social costs of delayed settlement.
The foregoing shows that the ability of defendants to sort by delay rather than by trials creates offsetting effects. It saves on trial costs over the range $\alpha \in [0, \alpha_1]$ where low-cost plaintiffs settle on the courthouse steps rather than going to trial, but it imposes costs over the range $\alpha \in [\alpha_1, \alpha_2]$ where the defendant delays settling with low-cost types rather than settling with them immediately. Area 1 in Figure 2 thus gives the savings from delay, and Area 2 gives the cost; either area may be larger. (Appendix B provides details of the comparison.) It follows that policies aimed at reducing delay may not have the intended effect of lowering social costs if they channel some cases to trial that otherwise would have settled.

V. Conclusion

Sorting represents a heretofore unexamined source of delay in the settlement of legal disputes. This paper has combined the standard asymmetric information model of litigation and settlement with techniques from the self-selection literature to derive the conditions under which an uninformed defendant who is able to commit to a menu of offers will use delay as a sorting device to lower his overall costs of litigation. The fact that this strategy can be effective in lowering a defendant’s private litigation costs, however, does not ensure that it will also lower social costs. Indeed, we saw that sorting may or may not lower social costs, depending on whether the defendant would have gone to trial or settled immediately with those cases that he settles after delay. From a policy perspective, this implies that observed delays in settlement are not necessarily undesirable from a social perspective, as conventional wisdom might suggest.

Appendix A

Sorting by Differences in Discount Rates

Suppose that plaintiffs have the same value of trial, $V$, but different discount rates, $r$. Specifically, suppose that there are two types of plaintiffs, those who discount the future at rate $r_H$ and those who discount the future at rate $r_L$, where $r_H > r_L$. Also assume that there are no other costs of delay. The present value of trial as of $t=0$ for a type $j$ plaintiff ($j = H, L$) is thus $Ve^{-r_jT}$. Similarly, the present value of a settlement offer $(S, t)$ is $Se^{-r_jt}$. Equating these expressions yields $^{19}$

$$S_j^0 = Ve^{-r_jT - \beta}, \quad j = H, L.$$  \hspace{1cm} (A1)

It follows that $\partial S_j^0/\partial t > 0$, implying that the minimum acceptable settlement amount is increasing with time. Further, $S_j^0 - S_H^0 = Ve^{-r_H(T - \beta)} - Ve^{-r_L(T - \beta)} > 0$ for $t < T$. Thus, the reservation price line for plaintiffs with a low discount rate is above that for plaintiffs with a high discount rate everywhere except at $t = T$. The situation is therefore identical to that in Figure 1. If we combine differences in plaintiff discount rates with differences in waiting costs (as in the text), the analysis would be more complicated, but separation, in principle, would remain possible as long as the reservation price curves intersect somewhere between $t = 0$ and $T$.

$^{19}$If courts award prejudgment interest on trial awards based on rate $R$ [Kessler (1996)], then the present value of trial for a type $j$ plaintiff is $Ve^{(-r_jT + \beta T)}$, and (A1) would become $S_j^0 = Ve^{-r_jT + \beta T}$. Since this has the same properties as the expression in (A1), the basic conclusions would be unaffected.
Appendix B

Condition Under Which Delay Lowers Social Costs

Recall from the text that the savings from delay are given by Area 1 in Figure 2, which is the expected trial costs with high-type plaintiffs, $\alpha x$, integrated over $[0, \alpha_1]$, or

\[ \text{Area } 1 = \int_{0}^{\alpha_1} \alpha x \, dx = \frac{x^2}{2} \bigg|_0^{\alpha_1} = \frac{\alpha_1^2}{2}. \quad (B1) \]

The costs of delay (Area 2) are the waiting costs for high type plaintiffs and the defendant over the range $[\alpha_1, \alpha_2]$:

\[ \text{Area } 2 = \int_{\alpha_1}^{\alpha_2} \alpha (k + c_L) T \, dx = (k + c_L) T (\alpha_2^2 - \alpha_1^2) / 2. \quad (B2) \]

Sorting by delay saves social costs if Area 1 > Area 2. After substituting for $\alpha_1$ and $\alpha_2$ from (4) and (6), respectively, this comparison becomes:

\[ 1 + x / [(k + c_L) T] > \left[ 1 + x / [(k + c_L) T] \right]^2. \quad (B3) \]

Because $c_L < c_H$, the expression in braces on the right-hand side is smaller than the expression on the left-hand side, and both are larger than one. Thus, this inequality may or may not hold.

References


Settlement delay


