ISSN 1045-6333

THE OPTIMAL STRUCTURE OF CONTINGENT FEES IN A WORLD OF SETTLEMENT

Bruce L. Hay

Discussion Paper No. 172

11/95
(Revised 3/96)

Harvard Law School
Cambridge, MA 02138

The Center for Law, Economics, and Business is supported by a grant from the John M. Olin Foundation.

JEL Class K41

THE OPTIMAL STRUCTURE OF CONTINGENT FEES IN A WORLD OF SETTLEMENT

Bruce L. Hay*

ABSTRACT

This article examines the design of contingent fees for plaintiffs' lawyers in a legal system that gives parties the choice between going to trial and settling out of court. Using a simple principal-agent model with attorney moral hazard, the article shows that the client generally benefits from a bifurcated fee structure in which the attorney gets a large fraction of the recovery in the event of trial, but a small fraction in the event of settlement; this structure maximizes both the size of the recovery and the client's distributive share. The article also examines the limits on the use of this fee structure that are imposed by two aspects of the settlement bargaining process: (1) the allocation of settlement authority between lawyer and client, and (2) the relative bargaining power of plaintiff and defendant.

---

*Assistant Professor, Harvard Law School.

THE OPTIMAL STRUCTURE OF CONTINGENT FEES IN A WORLD OF SETTLEMENT

Bruce L. Hay*

I. INTRODUCTION

This paper examines the operation of contingent fees in the settlement of litigation. It addresses the following question: What contingent fee maximizes the plaintiff's welfare, given that the parties can choose between going to trial or settling instead?

The problem is important because of the sheer number of cases in which it comes up. For certain common types of damage claim, two facts stand out: most are brought under contingent fees giving the lawyer a fixed percentage of the award; and most are settled before trial, often early in the litigation. What characterizes the right fee in such cases? How does it differ from the fee that would be appropriate if cases went to trial?

The crux of this problem is that the amount a case settles for reflects, in some measure, the expected judgment at trial. In choosing a fee, the client thus needs to worry about its effect on the expected judgment at trial, since this will determine what she gets if the case settles.[1] Even if settlement is much more likely than trial -- indeed, even if no case ever goes to trial -- the fee must be designed with an eye on trial as much as on settlement.

In exploring the client's problem, I consider two basic fee structures: *unitary* fees, which give the lawyer the same percentage whether the case settles or goes to trial; and *bifurcated* fees, which (may) give the lawyer one percentage if the case settles, another if it goes to trial. Our

---

*Harvard Law School. I thank Louis Kaplow and an anonymous referee for their helpful comments. I am also grateful to participants in the Harvard Law School Seminar on Law and Economics.

[1]For clarity of reference, I apply the female pronoun to the client and the male pronoun to the lawyer.

objective will be to characterize the optimal fee under *both* types of structure. Generally the bifurcated fee structure is preferable to the unitary structure; however, since unitary fees are so frequently encountered in actual practice, I will examine their optimal value even in settings where the client would do better with a bifurcated fee.

The basic findings of the paper are as follows. The optimal bifurcated fee often couples a relatively *high* trial percentage for the lawyer (one that would excessive if the case were actually *going* to trial) with a a relatively *low* settlement percentage. The rationale of the large trial percentage is that it generates a large settlement; the rationale of the small settlement percentage is that it avoids paying the lawyer for (trial) work he does not perform. In contrast, under the optimal *unitary* fee, these rationales in effect cancel each other; there is no reason, in general, to expect the optimal fee to be lower or higher than would be optimal if the case were going to trial.

Previous work has shown how the use of contingent fees may affect the client's welfare in the settlement process. Geoffrey Miller has analyzed the diverging interests of lawyer and client on the question of how much to settle for; his analysis implies that if the lawyer controls the settlement decision, the client may sometimes be better off using something other than a contingent fee (such as an hourly fee).[2] More recently, Lucian Bebchuk has shown that if the client controls the settlement decision, he is generally better off using a contingent fee instead of an hourly fee.[3] These works focus, however, on the choice between simple unitary contingent

---

[2]See Geoffrey P. Miller, Some Agency Problems in Settlement, 16 J. Legal Stud. 189 (1987). The divergence in interest stems from the fact that under a contingent fee, the lawyer bears all the costs of going to trial.

[3]See Lucian A. Bebchuk, How Would You Like to Pay for That? The Strategic Effects of Contingency Fees and Retainer Arrangements on Settlement Negotiations (unpublished manuscript, 1994). The intuition behind Bebchuk's result is simple: under an hourly fee, the client bears his side's cost of going to trial; under a contingent fee, the client does not bear any costs of going to trial. All else equal, then, the client's reservation price for settling will be higher under the contingent fee. The upshot is that he can extract a greater settlement amount from the defendant.

fees and non-contingent fees. They do not analyze either the optimal value of a contingent fee (what percentage the lawyer should get) or its optimal structure. That is the attempted contribution of this paper.

Sections II and III of the paper analyze the client's problem without explicitly modeling the process of settlement; this enables us to identify the basic properties of the optimal fee without introducing unnecessary complexity. Section IV refines the analysis by building the client's optimization problem into an explicit model of the settlement process. I consider there how the optimal fee depends on the allocation of settlement authority between lawyer and client, and depends on the relative bargaining power of plaintiff and defendant in settlement. Section V discusses applications.

II. THE CLIENT'S PROBLEM

A. The Objective

Consider a hypothetical case in which the plaintiff seeks money damages from the defendant. He hires the lawyer under a linear contingent fee. At some point early in the litigation, the parties have an opportunity to settle the case. If the case fails to settle, it goes to trial. Define the following notation:

p = Probability the case settles ($p > 0$).

s = Amount of the settlement, conditional on the case settling;

w = Expected judgment in the case, conditional on the case going to trial.

The client's objective, in choosing a fee, is to maximize her net expected return from the case.[4] The fee consists of a schedule (r_s, r_t), where

r_s = Lawyer's fee percentage if the case settles ($0 \leq r_s \leq 1$);

r_t = Lawyer's fee percentage if the case goes to trial ($0 \leq r_t \leq 1$).

Thus, the client's objective is to choose r_s and r_t to maximize the expression

$$(1-r_s)ps + (1-r_t)(1-p)w. \qquad \textbf{(1)}$$

We will denote the solution as follows:

r_s^{**}, r_t^{**} = The optimal fee schedule in a world in which cases may settle.

If the fee is unitary, so that $r_s = r_t$, we will simply denote the solution r^{**}.

B. Constraints

The central assumption in our analysis is as follows: in choosing a fee, *the client faces the task of encouraging the lawyer to invest effort if the case goes to trial, but not if it settles.* In particular, if the case fails to settle, we will assume that the expected award at trial is significantly affected by the lawyer's choice of pretrial effort level. More precisely, we assume that additional work by the plaintiff's lawyer produces positive (but diminishing) returns at trial;[5] and that the lawyer is free to choose whatever level of effort he wishes in preparing for trial.

In contrast, we will not make any corresponding assumptions about the lawyer's efforts before settlement negotiations occur; instead, we will assume that negotiations occur sufficiently

[4]All actors in the model are risk-neutral.

[5]Thus, if x_t is the lawyer's post-negotiation investment in the case, then $dw/dx_t > 0$, and $d^2w/dx_t^2 < 0$. We assume $w(x_t)$ is continuously differentiable.

early in the case that the lawyer's prior investment has little bearing on the outcome of the negotiations. We may suppose that the lawyer incurs some cost (talking to witnesses and the like) before settlement negotiations occur; however, we will assume this cost is exogenously fixed.

The motivation here is simple. The client's problem is most interesting in a situation where settlement occurs early, before much litigation cost has been incurred. The later settlement occurs, the more settlement resembles trial for purposes of setting the fee; in the limiting case, where settlement occurs on the eve of trial, the client's problem presumably differs relatively little from (though it is not identical to) what it would be in a trial-only world. We will accordingly keep our attention on the case in which encouraging lawyer effort is only an issue if the case fails to settle.

This assumption implies that the client's maximization problem is subject to the following constraints. Define the following additional notation:

x_s = Cost incurred by lawyer prior to settlement negotiation;

x_t = Lawyer's investment of effort in trial preparation, if the case fails to settle.

Per our assumption, x_s is an exogenously fixed amount; in contrast, x_t is chosen by the lawyer. The lawyer's *incentive compatibility* constraint is that in the event of trial, he will choose the level of effort that (given the fee) maximizes his expected return from trial; thus, x_t will be chosen to maximize

$$r_t w - x_t. \qquad \textbf{(2)}$$

The lawyer's *participation constraint* is that the fee schedule must assure him (given his anticipated costs) a nonnegative expected return from taking the case; thus, the fee schedule must satisfy

$$[pr_s s - x_s] + (1-p)[r_t w - x_t] \;\geq\; 0. \tag{3}$$

III. BASIC PROPERTIES OF THE OPTIMAL FEE

A. The Model of Settlement

To begin our analysis of the optimal fee, we will not explicitly model the process of settlement bargaining. Instead, we will simply express the settlement amount as some multiple of the expected judgment, as follows:

$$s = q \cdot w,$$

where q is some nonnegative factor (not necessarily a constant). We make no assumptions about q; it may be greater or less than one. In addition, we make no assumptions about the value of p, the probability of settlement. However, we assume for clarity's sake that the values of p and q are exogenously given, rather than being a function of the fee. (Proofs of all propositions are in the Appendix.)

In characterizing the optimal fee, the following notation will be helpful:

r^* = The optimal linear fee in a trial-only world (where settlement is impossible);[6]

$\hat{r}$ = The maximum feasible percentage the client can give the lawyer;

$\check{r}$ = The minimum feasible percentage the client can give the lawyer.

The value of $\hat{r}$ may be dictated by legal restrictions on the contingent fee (which often prohibit fees exceeding, say, 50 percent). The value of $\check{r}$ is dictated by the lawyer's participation constraint (expression **(3)**); the fee must be large enough to ensure that the lawyer's expected

[6]Thus, r^* maximizes $(1\text{-}r)w$.

return from taking the case is nonnegative.

B. The Optimal Fee

1. Unitary Fees

Begin by assuming the client must use the same fee in the event of either trial or settlement, so that $r_s = r_t$. Then we have

PROPOSITION 1. *If the client is compelled to use a unitary fee, then* $r^{**} \geq r^*$.

The reason for this perhaps counterintuitive result is that the amount of settlement reflects (more precisely, varies in a linear fashion with) the expected judgment in the case. As the fee drops, so does the expected judgment, because the plaintiff's lawyer has less incentive to invest in the claim should it go to trial; as a result, the defendant has less to lose from going to trial, and thus will pay less to settle the case.

The upshot is that *even if settlement is a sure thing, and costs the lawyer nothing, the client will want to give the lawyer as high a fee percentage he would give if trial were unavoidable.* And if settling is costly for the lawyer ($x_s > 0$), it may even be that that the client will want to give the lawyer a *higher* percentage than he would in a trial-only world; this may be necessary to satisfy the lawyer's participation constraint (expression **(3)**).[7] Provided that p and q are independent of the fee r^{**} is never lower than r^*.

[7]As an example, suppose p=1, and that $x_s > r^*qw^*$. Then r^* will not satisfy the lawyer's participation constraint, and the client must give the lawyer a higher fee.

2. Bifurcated Fees

Now let us assume that the client can bifurcate the fee, so that r_s need not be the same as r_t. We have

PROPOSITION 2. *Under the optimal bifurcated fee,*

(*a*) $r_s = \check{r}$, *and*

(*b*) $r_t = \hat{r}$ *if the value of p or q is sufficiently great.*

The essential intuition here is that the expected judgment increases with r_t but is unaffected by r_s; and since the settlement amount is tied to the expected judgment, it too increases with r_t but is unaffected by r_s. Thus, on the one hand, the client can reduce the value of r_s without reducing the (expected) amount collected from the defendant. It is therefore in the client's interest to lower r_s as much as possible.[8]

On the other hand, the client can, by increasing r_t, increase the amount collected from the defendant, without (if the case settles) increasing the fraction paid to the lawyer. This is easiest to see if we assume $p = 1$. The client may as well make r_t as large as possible; doing so will generate a very large settlement, and (precisely because the case settles) the client will *not actually have to pay* the lawyer r_t.[9]

---

[8]Observe that as r_s drops while r_t is held constant, the left-hand term in **(1)** increases while the right-hand term is unaffected.

[9]A similar argument shows that if q is sufficiently large, the client will (even if $p < 1$) make r_t as large as possible. The greater q is, the more valuable it is to the client to increase the value of w. If q is sufficiently large, the client will want to set $r_t = 100\%$; even though (if $p < 1$) that means she will get nothing in the event the case goes to trial, that prospect is more than outweighed by the enormous amount she will get in the event of settlement.

C. Advantages of Bifurcation

This discussion highlights two ways in which the bifurcated fee, optimally used, improves upon the unitary fee. It both *increases the size of the pie* (the expected payment from the defendant) to be divided), and *increases the client's distributive share of the pie*. The point again can be made most clearly if we assume that $p = 1$. Under the optimal unitary fee, the client gets

$$(1-r^*)qw^*, \qquad \textbf{(4)}$$

under the optimal bifurcated fee, she gets

$$(1-\check{r})q\hat{w}, \qquad \textbf{(5)}$$

where $\hat{w}$ denotes the expected award when the lawyer's trial fee is the maximum possible percentage. Since (in general) $\hat{w} > w^*$, the bifurcated fee generates a bigger settlement than the unitary fee. And since (in general) $\check{r} < r^*$, the bifurcated fee gives the client a bigger fraction of the settlement.

The essential insight here is as follows. In a trial-only world (where the expected judgment is a function of lawyer effort), the linear contingent fee forces the client to make a tradeoff between maximizing the size of the pie and maximizing her distributive share; the solution normally lies between the minimum and maximum feasible fee amounts.[10] A very low fee yields an unacceptably small expected judgment; a very high fee gives the lawyer an unacceptably large share of the expected judgment.

In a world of settlement, she faces the same tradeoff -- if the fee must be unitary. Since

---

[10] See Bruce L. Hay, Contingent Fees and Agency Costs, Working Paper No. 162, John M. Olin Program on Law, Economics and Business, Harvard Law School (July 1995), forthcoming, J. Legal Stud. (June 1996).

the settlement amount reflects the expected judgment, it does the client no good to lower the fee; the effect would simply be to reduce the amount of settlement. The optimal fee remains (no lower than) r^*, even though the lawyer may do little or no work before the case settles.

A bifurcated fee enables the client to separate the problems of pie-slicing and pie-enlarging. The value of r_t determines the size of the recovery; r_s determines the client's distributive share. Thus, starting at r^*, she can -- in this model -- enlarge the pie (by raising r_t) without cutting into her distributive share of the pie; likewise, she can enlarge her share of the pie (by increasing r_s) without reducing the size of the pie.

IV. OPTIMAL FEES IN A MODEL OF SETTLEMENT

An important limitation of the above analysis is that there is no obvious reason to expect p and q to be independent of the fee. Refining our results to take account of this limitation, however, requires us to use a more explicit model of the settlement bargaining process. In what follows we will therefore see what the optimal fee looks in a more precisely-specified picture of settlement bargaining.[11] (Proofs are again in the Appendix.)

A. Refined Model of Settlement

The settlement range in a case -- the set of settlement amounts that make both parties better off than going to trial -- is defined by the gap between the *plaintiff's reservation price for*

[11]The model builds on the standard works in the economic literature on settlement. See, for example, William M. Landes, An Economic Analysis of the Courts, 14 J. L. & Econ. 61 (1971); Richard A. Posner, An Economic Approach to Legal Procedure and Judicial Administration, 2 J. Legal Stud. 399 (1973); Steven Shavell, Suit, Settlement and Trial: A Theoretical Analysis under Alternative Methods for the Allocation of Legal Costs, 11 J. Legal Stud. 55 (1982); Lucian A. Bebchuk, Litigation and Settlement under Imperfect Information, RAND J. Econ. 404 (1984).

settling and the *defendant's reservation price* for settling. The *probability* of settlement depends on whether the latter exceeds the former; obviously settlement is impossible if this precondition is not met. The *amount* of settlement depends in part on the location of the settlement range, and in part on how the parties split the settlement surplus; they may settle at a point near one party's reservation price, or somewhere in between.

Our interest will be in seeing how the optimal fee depends on two overlapping features of the bargaining process, which are presumably exogenous to the fee itself: (1) *Who -- as between lawyer and client -- controls the plaintiff's settlement decision*; and (2) *who -- as between plaintiff and defendant -- dictates where on the settlement range the settlement will likely occur.*[12] To examine these issues, we will express the settlement amount as

$$s \;=\; \lambda(V_p) \;+\; (1\text{-}\lambda)V_d, \qquad \textbf{(6)}$$

where

V_p	=	Plaintiff's reservation price for settling;
V_d	=	Defendant's reservation price, and
λ	=	Exogenous parameter indicating where, on settlement range, the parties settle.

In expression **(6),** the the relative bargaining power of the plaintiff and defendant is

[12]In isolating these features, I have been influenced by the earlier studies of Miller, supra note __, and Bebchuk, supra note ___. Regarding the first feature, Miller has emphasized how settlement outcomes can differ depending on the allocation of control between lawyer and client; I follow him in considering what happens when one or the other actor controls the settlement decision. Regarding the second feature, Bebchuk's model assumes that the defendant captures the entire surplus from settling; this raises the question of how things change if we make contrary assumptions.

captured by the λ term.[13] A large value of λ indicates the case is expected to settle at a point near the plaintiff's reservation price; this corresponds to the possibility that the defendant is generally (credibly) able to make the final settlement offer on a take-it-or-leave-it basis. A small value of λ means the case is expected to settle at a point near the defendant's reservation price, meaning the plaintiff can credibly make the final offer. A moderate value of λ means neither party is generally more likely than the other to make the final offer. meaning and so on.

As for the allocation of settlement authority between lawyer and client, this element is captured by the term V_p. To say that the client has control over the settlement decision is to say that she establishes the bottom line for settling. More precisely: Suppose lawyer and client are faced with the choice of settling the case for some amount $\bar{s}$ or going to trial.[14] If the client controls the decision, *the plaintiff will settle if and only if it makes the client better off than going to trial* (whether or not it makes the lawyer better off). Similarly, if the lawyer controls the decision, the plaintiff will settle if and only if it makes the lawyer better off than going to trial (whether or not it makes the lawyer better off).[15]

---

[13]The value of λ can be construed in either of two ways. First, the parties might settle somewhere in between their reservation prices, and λ indicates the location (in terms of the fraction of the distance from the defendant's to the plaintiff's reservation price). Second, the parties might always settle at one or the other's reservation price (depending on who make the final offer), and λ reflects the probability that the defendant, rather than the plaintiff, makes the final offer. It makes no difference which interpretation we adopt.

[14]We assume in this example that the defendant is prepared to settle for $\bar{s}$.

[15]Here we can imagine different plausible stories. In some settings the client will have only a very rough idea of whether a proposed settlement amount approximates the expected judgment; consider a tort case involving complex issues of proof, or highly variable damage measures such as pain and suffering. The lawyer in such settings may be willing and easily able to convince the client to accept a settlement that is (unbeknownst to the client) well below the expected judgment. (My assumption here is that the client has formal veto power over a proposed settlement. It is unlikely a court would honor any attorney-client agreement that gave the lawyer the power to settle over the client's objection.) .

On the other hand, if the client has independent sources of information concerning the expected judgment, or if the lawyer is worried about possible adverse professional consequences of "selling out" on the client, we might expect the plaintiff to refuse any amount much below the expected judgment .

For clarity of exposition, we assume that the parties are symmetrically informed about the case, and that the case definitely settles if there is a positive settlement range.

B. The Optimal Fee

Let us now characterize the optimal (bifurcated) fee in this model.[16] In general, our model does not generate the corner solution we reached in Section III, where r_t was as set high as possible and r_s set just large enough to give the lawyer a nonnegative return in the case. Rather, upon taking the mechanics of settlement bargaining into account we find that, when they are optimally chosen, the gap between r_s and r_t usually is narrower than our initial analysis suggested. This is true for slightly different reasons, depending on whether the settlement decision is controlled by the client or the lawyer.

1. Client Control

Begin with the case in which the client controls the plaintiff's settlement decision. We have the following result:

PROPOSITION 3. *If the client controls the settlement decision, then* $r_s{}^{**} = \check{r}$, *while* $r_t{}^{**}$ *varies with the value of* λ:

(*a*) *For large values of* λ, $r_t{}^{**} \approx r^*$.

(*b*) *For small values of* λ, $r_t{}^{**} \approx \hat{r}$.

[16] We will not compel the client to use a unitary fee in the analysis to follow.

This point may be seen as follows. The parties' reservation prices for settling are given by

$$V_p = \frac{(1 - r_t)w}{1 - r_s} \tag{7}$$

and

$$V_d = w + y, \tag{8}$$

where

y = Defendant's anticipated costs from going to trial.

Then if the case settles, the client's net recovery is given by[17]

$$\lambda[(1-r_t)w] + (1-\lambda)[(1-r_s)(w+y)]. \tag{9}$$

The client's problem is to choose a fee schedule that maximizes **(9)**. Consider r_s: the first bracketed term in **(9)** is unaffected by r_s, while the second bracketed term is maximized by choosing the smallest possible value of r_s; accordingly, no matter what the value of λ, the client wants to make r_s as small as possible. Now consider r_t: the first bracketed term is maximized by setting r_t equal to r^*, while the second bracketed term is maximized by giving r_t the largest possible value.[18] The client thus faces a tradeoff between setting r_t close to r^* and setting it close to $\hat{r}$; generally, she will want to set it somewhere between the two; the smaller λ is, the closer r_t should be set to $\hat{r}$. (Note that when $\lambda = 1$, trial is the same as settlement from the client's perspective, so it is unsurprising that she should use the same fee she would in a trial-only world.)

---

[17] If the case settles, the client's recovery is $(1-r_s)s$. The value of s in this setting is derived by plugging **(7)** and **(8)** into **(6)**. Simplifying yields expression **(9)**.

[18] Recall that w increases with r_t. We assume that the defendant's litigation costs (y) are nondecreasing with x_t; if so, they are nondecreasing with r_t.

2. Attorney Control

Turn now to the case in which the attorney controls the plaintiff's settlement decision. We have

PROPOSITION 4. *If the lawyer controls the settlement decision, then the optimal fee is independent of λ, and generally falls in between the extremes that are possible when the client controls the decision. In particular, the optimal fee generally sets $\check{r} < r_s^{**} < r_t^{**}$, and $r^* < r_t^{**} < \hat{r}$.*

To see this point, begin by observing that

$$V_p = \frac{r_t w - x_t}{r_s}, \tag{10}$$

while V_d remains as in **(8)** above. Examining **(8)** and **(10)**, we see that, by raising r_t and lowering r_s, the client increases V_p as well as V_d. Thus, no matter what the value of λ, *raising r_t while lowering r_s has the effect of increasing the settlement amount, while at the same time increasing the client's share of the settlement.* As a result, raising raising r_t while lowering r_s seems unambiguously desirable to the client. However, there is a limit on the extent to which the client can do this: if she makes the gap between r_s and r_t too great, then V_p will exceed V_d, meaning the case will not settle.

The optimal fee is therefore drawn from the set of fees that makes V_p equal to (or just below) V_d.[19] By choosing a fee from this set, the client ensures that the case settles, and extracts from the defendant the entire settlement surplus.[20] Under such a fee, the client's expected recovery is simply given by[21]

$$(1-r_s)(w+y). \tag{11}$$

We can, accordingly, write the client's optimization problem of maximizing **(11)**, subject to the new constraint that $V_p = V_d$. In general, this new constraint generates an interior solution for both r_s** and r_t** -- fixing r_t at a point below the maximum feasible fee $\hat{r}$ (though above r*), and fixing r_s somewhere above the minimum feasible fee $\check{r}$.

C. Further Considerations on the Optimal Contract

1. The Value of Bifurcation

We know, of course, that using a bifurcated fee weakly dominates using a unitary fee, in that the optimal bifurcated fee always performs at least as well as the optimal unitary fee. (One could always set r_s equal to r_t.) But when does it perform substantially better? The foregoing analysis yields

PROPOSITION 5. *The advantages of a bifurcated fee are greatest when either*

(a) the lawyer controls the settlement decision, or

(b) the client controls the settlement decision and λ *is small.*

---

[19]For this reason, the value of λ is irrelevant in determining the optimal fee.

[20]Note, however, that she is forced to share some of the surplus with the lawyer.

[21]The client's recovery is $(1-r_s)s$, where $s = \lambda(V_p) + (1-\lambda)V_d$; since $V_p = V_d$, we have $s = V_d$. Substituting **(9)** for V_d gives expression **(11)**.

In contrast, there is little to be gained from using a bifurcated fee when the client controls the settlement decision and λ is large.

This follows from the points we have made above. As expression **(9)** makes clear, if the client controls the settlement decision and λ is very large, the client's expected recovery is (approximately) given by

$$(1-r_t)w; \tag{12}$$

this expression is unaffected by the value of r_s. As a result, the client gains nothing by choosing a value of r_s different from r_t; he does about as well using a unitary fee.

Yet if λ is very small, then we know from **(9)** that the client's expected recovery when she controls the settlement decision is (approximately) given by

$$(1-r_s)(w+y); \tag{13}$$

this is also the client's expected recovery (under the optimal fee) if the lawyer controls the settlement decision. By inspection, the lower the value of r_s, the greater the value of **(13)**.[22] In this setting, therefore, the client has something to gain by choosing a relatively low value of r_s. Bifurcating the fee thus makes a difference.

2. <u>The Allocation of Settlement Control</u>

One further question suggests itself. What leaves the client better off -- keeping control of

[22]Recall that r_s has no effect on the value of either w or y.

the settlement decision, or ceding control to the lawyer?[23] Our analysis implies

PROPOSITION 6. *If λ is large, the client is generally better off (under the optimal fee) if the lawyer controls the settlement decision. If λ is small, the client is generally better off (under the optimal fee) if she herself controls the settlement decision.*

The point may be seen informally as follows. Suppose the client controls the settlement decision. If λ is large, her recovery under the optimal fee is approximately

$$(1-r^*)w^*; \tag{14}$$

if λ is small, her recovery under the optimal fee is approximately

$$(1-\check{r})\hat{w}. \tag{15}$$

In contrast, if the lawyer controls the decision, the plaintiff's recovery under the optimal fee is

$$(1-r_s)w, \tag{16}$$

where (as we have seen) r_s is typically somewhere between $\check{r}$ and r^*, while w is typically somewhere between w^* and $\hat{w}$; thus, **(16)** is usually greater than **(14)** but less than **(15)**.

The essential reason for this result is as follows. If the client controls the settlement decision, there is -- under the optimal fee -- a substantial gap between the parties' reservation prices. Thus, the client's welfare depends heavily on the value of λ. In contrast, if the lawyer controls the decision, the optimal fee eliminates the gap between the parties' reservation prices, so λ is not an issue. However, when the lawyer controls the decision, there is a binding constraint

[23] In practice, there may be no choice in the matter. (For example, the client's lack of information about the case may be such that the lawyer inevitably controls the decision.) I pose this issue purely hypothetically.

on bifurcating the fee that does not exist when the client controls the decision.[24] Thus, lawyer control is better than the worst-case (high λ) scenario of client control, but not as good as the best-case (low λ) scenario.

V. APPLICATIONS AND CONCLUDING REMARKS

A. Predicting Market Behavior

How well does the model predict or correspond to actual market behavior? Little data exists to tell us. So far as I am aware, the only data concerning the use of unitary versus bifurcated fees is contained in one of the RAND Institute for Civil Justice's studies of the tort system.[25] Using a sample of 371 personal injury cases in which contingent fees were used,[26] the RAND researchers examined both the structure of the fee and the average percentages paid to the lawyer. Though insufficiently detailed to test the model,[27] their findings give us some useful information.

Table 1 contains the RAND findings. In the first column, I use the term "unitary" to refer to fees in which the lawyer collected a simple fixed percentage of the recovery, regardless of how

---

[24]This constraint is that $V_p \leq V_d$. This constraint is present in both the client-control and lawyer-control settings; however, it is binding only in the lawyer control setting. (The reason is that, as the gap between r_s and r_t gets bigger, the constraint becomes harder to satisfy when the lawyer controls the settlement decision, but *easier* to satisfy when the client controls the decision.)

[25]Deborah R. Hensler et al., Compensation for Accidental Injuries in the United States 135-36 (RAND Institute for Civil Justice, Publication No. R-3999-HHS/ICJ) (1991).

[26]The data were compiled from all accident cases (excluding fatalities and toxic exposure cases) brought in 1988 by individuals in a random sample of roughly 18,000 households.

[27]One major difficulty is that, concerning non-unitary fees, the study lumps together three structures: (1) fees in which the percentage paid depends on whether the case settles or is tried (what I have called a bifurcated fee); (2) fees in which the percentage paid depends on the amount of time spent on the case; and (3) fees in which the percentage paid depends on the amount recovered. See id. at 135-36.

the recovery was obtained (and regardless of the size of recovery). The term "non-unitary" fees includes not only what I have called bifurcated fees (in which the percentage paid depends on *whether the case settles or is tried*), but also two other distinct structures: fees in which the

TABLE 1: CONTINGENT FEE STRUCTURE IN
IN PERSONAL INJURY CASES

Fee structure	Fraction of cases in which structure was used	Mean fee percentage	Median fee percentage
Unitary	.67	$r = 29\ \%$	$r = 33\ \%$
Non-unitary	.2	$r_l = 20\ \%$ $r_h = 30\%$	$r_l = 25\ \%$ $r_h = 33\%$

Source: RAND Publication R-3999-HHS/ICJ (1991).

percentage paid depends on the *amount of time spent* on the case; and fees in which the percentage paid depends on the *amount recovered.*[28] Unfortunately, the study does not contain separate data for these different types of non-unitary fee. (Notice, however, that a contingent fee based on the amount of time spent should, in practice, resemble a fee based on whether the case settles or goes to trial.)

The second column indicates the frequency with which unitary and non-unitary fees were used. The third column indicates the mean percentages paid the lawyer. The term r_l refers to the

[28]See id. at 135-36.

mean low fee in non-unitary fees; r_h refers to the mean high fee. I have used this notation because the data do not distinguish between different types of non-unitary fee. To the extent the data refers to bifurcated fees, r_l corresponds to r_s; r_h corresponds to r_t.

Regarding the results in the second column, one fact stands out: fewer than one in five cases involved bifurcated fees. The RAND study does not indicate the settlement rate of the cases in the sample, but presumably the vast majority settled and were *expected* to settle at the time the lawyer was hired. Why so few bifurcated fees? Putting aside potential market imperfections that might prevent the use of the optimal fee,[29] a few possibilities suggest themselves.

One possibility, suggested by the model, is that clients effectively control the decision to settle, and that λ is large. For then, as we have seen, bifurcating the fee does the client no good; she does just as well using the optimal unitary fee. In considering this hypothesis, we do not need to suppose the client controls the litigation in any other respect; we can assume that clients often play passive roles in virtually all aspects of the litigation. This point is worth emphasizing. All we need to believe is that *lawyers do not have the ability to reject offers that the client prefers to accept*; and that *defendants, knowing this, make take-it-or-leave it offers that make the client barely better off (in expected terms) than going to trial.* If so, clients gain nothing (or little) by bifurcating the fee.

A glance back at expressions **(7)** and **(10)** will make the point clearer. As those

[29]Lawyer market power seemss an implausible explanation, if only because the average unitary fee charged -- 29% -- is well below the maximum legal contingent fee in most jurisdictions. If lawyers lack sufficient market power to extract the maximum possible fee, why would they be able to compel the use of unitary fees? Client ignorance might be an explanation; one conjectures, however, that competition for clients would lead lawyers to call prospective clients' attention to the advantages of a bifurcated fee (for fear of losing business to rivals who did so). I do not pursue these matters here.

expressions indicate, lowering r_s and raising r_t have the effect of increasing the lawyer's reservation price, but of decreasing the plaintiff's reservation price. The upshot is that, if r_t substantially exceeds r_s, the lawyer's reservation price will often exceed the client's -- meaning that *there will be some (low) settlement amounts that the client wants to accept but the lawyer wants to reject.* Suppose that the lawyer has no power to refuse such offer amounts over the client's objection;[30] and suppose the defendant can credibly threaten to go to trial if they are not accepted. This is the same as saying that the client controls the decision, and that λ is large. Whether this story holds in most cases is debatable, but not obviously implausible.

Other possibilities, not encompassed by the model but easily incorporated into it, are risk aversion or high discount rates on the part of tort claimants. As we have seen, if the lawyer controls the settlement decision, then bifurcating the fee has the effect of narrowing the anticipated settlement range. If there is uncertainty attaching to the value of V_p and V_d, then bifurcating the fee has the effect of reducing the probability that the case will settle.[31] If the prospect of a risky trial is sufficiently intolerable for the client, or if the cost of delay in obtaining compensation is sufficiently high for her, she will -- in order to ensure the case settles -- prefer to set r_s equal to r_t.[32] (Still another possibility, also excluded from the model but easily incorporated, is that the size of the settlement is a function of the amount of effort the lawyer

---

[30]Or, equivalently, is unable to talk the client out of accepting such an amount.

[31]Suppose V_p and V_d reflect the expected value of a distribution of possible reservation prices, as indicated on the horizontal axis. If The parties' realized reservation prices fall in the zone of overlap, the case will fail to settle (even though, in expected terms, there was a positive anticipated settlement range). All else equal, narrowing the distance between V_p and V_d increases the likelihood that the case will fail to settle.

[32]Indeed, she might prefer to set r_t a lot lower than r_s, so as to truly encourage the lawyer to settle. This is probably only a theoretical possibility. Since such an arrangement would make the lawyer's compensation inversely proportional to the amount of work done (and the amount of risk he sustained), it might well expose the lawyer to professional discipline of some sort.

chooses to invest in the case *before settlement*; this would require choosing a relatively high value of r_s in order to encourage pre-settlement effort.)

Regarding the results in the third column of Table 1, two points stand out. First, in the 20 percent of cases where non-unitary cases are used, there is (on average) a substantial gap between the large and small components of the fee. This is not necessarily inconsistent with the points just made about the appeal of unitary fees; these average figures may mask considerable variation among cases, as the difference between the mean and median figures suggests. Some cases in the sample may involve a bifurcated fee in which the gap between r_s and r_t is relatively small; in others, the gap may be very large, as we would suspect would be at least sometimes be optimal.

Second, the mean high fee in the non-unitary group is greater than the mean unitary fee, though only by a bit. In general, the model would predict r_t** to be greater (perhaps a lot greater) than the optimal unitary fee r** . In direction, then, if not in magnitude, the RAND data loosely track the model. Once again, however, the RAND study does not distinguish among different type of non-unitary fee, so we cannot read much about bifurcated fees into the data. They are at most suggestive.

B. Implications for Fee Regulation

Critics of the contingent fee often argue that fee percentages are typically too high, given that most cases settle early in the litigation.[33] A contract giving the lawyer, say, the conventional one-third of the recovery might (it is argued) be reasonable for a case in which an expensive and

[33]See, for example, L. Brickman et al., Rethinking Contingency Fees 22 (1994), which urges regulation requiring the use of bifurcated fees.

risky trial is likely; but such a contract is unreasonable for a case in which a cheap and speedy settlement is likely. Regulatory caps on contingent fees are often proposed as a solution. What does the model have to say about this matter?

Let us assume there exists some market defect that warrants such regulatory intervention on behalf of clients.[34] The model carries four basic lessons concerning the design of fee caps. First, the prospect of settlement does not, in general, justify lowering the fee, *if* the fee is unitary. Rather, the prospect of settlement justifies (in many cases) the substitution of a bifurcated fee for a unitary fee. If the fee must (for some reason) be unitary, capping the fee is no more justified when settlement is likely than when trial is likely.

Second, if the client controls the settlement decision (in the sense of establishing the bottom line for settlement), bifurcating the fee has little value unless the plaintiff has substantial bargaining power vis-a-vis the defendant. In the -- perhaps numerous -- situations where the defendant has the upper hand in bargaining (is able to credibly make final offers), no fee can much improve on r^*, the fee that would be optimal in a trial-only world.

Third, if the lawyer controls the settlement decision, bifurcating the fee is usually beneficial to the client. However, the gap between r_s and r_t cannot be too large, or the lawyer will be tempted to forgo settlement in favor of trial.

Finally, when bifurcation is desirable, the optimal fee will generally set r_t above the optimal fee in a trial-only world. That is, it will in effect overpay the lawyer in the event trial occurs.

---

[34]Three qualifications are needed for this discussion. First, since I have not explicitly modeled the workings of the contingent fee market (how clients search for lawyers, and so on), I take no position here on the need for regulation. My object here is to ask what form regulation should have, *if* it is undertaken. Second, I focus here on regulations designed to promote the welfare of plaintiffs. I put aside the question of regulations designed for other purposes, such as discouraging certain types of claim, keeping levels of liability down, and so forth. Third, I focus on fee caps on conventional linear contingent fees. I ignore alternatives such as allowing plaintiffs to sell their claims to third parties.

Regulatory efforts to cap r_t in order to prevent such overpayment may hurt its intended beneficiaries, by reducing the amounts plaintiffs collect in settlement.

APPENDIX

Proof of Proposition 1

The client's expected recovery if the lawyer works under a given fee r is

$$(1-r)[pq + (1-p)]w. \tag{A1}$$

The derivative of **(A1)** with respect to r is (since p and q are constants) simply the derivative of $(1-r)w$ with respect to r, that is,

$$(1-r)\left(\frac{dw}{dr}\right) - w\ . \tag{A2}$$

By definition, **(A2)** is zero at $r = r^*$ and is negative for $r > r^*$.[35] It follows if the lawyer's participation constraint (given by expression **(3)** in the text) is satisfied by r^*, then r^* is the optimal fee; if **(3)** does not hold at r^*, then the optimal fee is the lowest fee exceeding r^* that satisfies **(3)**.

Proof of Proposition 2

The plaintiff wants to choose r_s and r_t to maximize

$$(1-r_s)pqw + (1-r_t)(1-p)w. \tag{A3}$$

Since w is independent of r_s, differentiating **(A3)** with respect to r_s gives

$$-pqw, \tag{A4}$$

which is always negative; thus, the client should choose the lowest possible value of r_s fee satisfying the lawyer's participation constraint.

[35]This follows from our assumption that then $dw/dx_t > 0$, and $d^2w/dx_t^2 < 0$.

Differentiating **(A3)** with respect to r_t gives

$$(1-r_s)pq\frac{dw}{dr_t} + (1-p)\frac{d(1-r_t)w}{dr_t} \; . \qquad \textbf{(A5)}$$

At $r_t = r^*$, the right-hand term of **(A5)** is zero, while the left-hand term is positive.[36] To see how the value of r_t ** varies with the p and q, consider arbitrary values $\tilde{p}$ and $\tilde{q}$. Let $\tilde{r}_t$ be the optimal fee given these values; thus, **(A5)** is zero when it contains $\tilde{p}$, $\tilde{q}$, and $\tilde{r}_t$. If we replace $\tilde{q}$ with a higher value of q, the left-hand term of **(A5)** increases while the right-hand term remains constant, meaning $\tilde{r}_t$ is then dominated by a greater value of r_t. It follows that (holding p constant) r_t ** increases with q; and that for a given value of p, it must be possible to select a value of q sufficiently large that $r_t^{**} = \hat{r}$. A similar procedure shows that, for any given value of q, as p approaches1, r_t^{**} approaches $\hat{r}$.

<u>Proof of Proposition 3</u>

Observe first that the client wants the case to settle, which requires that r_s and r_t be chosen such that $V_p \leq V_d$. For consider any fee schedule ($\tilde{r}_s$, $\tilde{r}_t$) such that $V_p > V_d$: the client's expected recovery is given by $(1-\tilde{r}_t)\tilde{w}$. Suppose the client replaces $\tilde{r}_s$ with some lower value $\bar{r}_s$ such that $V_p \leq V_d$: then the client's expected recovery is (as we saw in expression **(9)** in the text) given by $\lambda[(1-\tilde{r}_t)\tilde{w}] + (1-\lambda)[(1-\bar{r}_s)(w+y)]$, which exceeds $(1-\tilde{r}_t)\tilde{w}$.

Accordingly, the client wants to choose r_s and r_t to maximize

---

[36]Solving the lawyer's maximization problem in the event the case goes to trial, we find (I omit the details here) that $dw/dr_t = (w')^2/r(-w'')$, which by assumption is positive.

$$(1\text{-}r_s)s, \qquad \textbf{(A6)}$$

which is given by expression **(9)** in the text. Differentiating **(9)** with respect to r_t yields

$$\lambda\left[\frac{d(1-r_t)w}{dr_t}\right] + (1-\lambda)(1-r_s)\left[\frac{dw}{dr_t} + \frac{dy}{dx_t}\frac{dx_t}{dr_t}\right]. \qquad \textbf{(A7)}$$

At $r_t = r^*$, **(A13)** is positive, since the left-hand term is zero and the right-hand term is always positive (assuming, reasonably, that dy/dx_t is nonnegative). For $r_t > r^*$, the left-hand bracketed term is negative and decreasing in r_t. As $\lambda \to 1$, **(A7)** approaches the value of its bracketed component, which is maximized by r^*. Thus, as $\lambda \to 1$, any value of r_t greater than r^* can be improved upon by a value of r_t closer to r^*. As $\lambda \to 0$, the left-hand term approaches zero, while the right-hand term is positive and increasing in r_t. Thus, as $\lambda \to 0$, any value r_t less than $\hat{r}$ can be improved upon by a value of r_t closer to $\hat{r}$.

Differentiating **(9)** with respect to r_s yields

$$\text{-}(1\text{-}\lambda)(w+y), \qquad \textbf{(A8)}$$

which is negative (provided that $\lambda < 1$). Thus, $r_s^{**} = \check{r}$ anytime $\lambda < 1$.

Proof of Proposition 4

If the lawyer controls the settlement decision, then V_p is given by expression **(10)** in the text. Consider any fee schedule ($\bar{r}_s$, $\bar{r}_t$) such that $V_p < V_d$. Suppose the client replaces $\bar{r}_s$ with a lower value of r_s such that $V_p = V_d$. The effect of doing so is to simultaneously increase the size of the settlement amount *and* increase the client's fractional share of it. Under the optimal fee, then, $V_p = V_d$, and under the resulting settlement, $s = w+y$.

The client's problem, then, is to choose a fee schedule (r_s, r_t) to maximize

$$(1-r_s)(w+y), \qquad \textbf{(A9)}$$

subject to the constraint that $V_p = V_d$, that is, subject to

$$r_s = \frac{r_t w - x_t}{w + y}. \qquad \text{(A10)}$$

Substituting **(A10)** into **(A9)**, we can rewrite the client's problem as setting r_t to maximize

$$(1-r_t)w + x_t + y. \qquad \textbf{(A11)}$$

Differentiating **(A11)** with respect to r_t gives

$$\frac{d(1-r_t)w}{dr_t} + \frac{dx_t}{dr_t}\left(1 + \frac{dy}{dx_t}\right), \qquad \textbf{(A12)}$$

which is positive at $r_t = r^*$, but may be positive or negative at $r_t = \hat{r}$. Plugging r_t^{**} into **(A10)** gives r_s^{**}. Observe that the lawyer's expected recovery is given by $r_s^{**}s^{**} = r_t^{**}w^{**} - x_t^{**} > 0$; thus, $r_s^{**} > \check{r}$.

<u>Proof of Proposition 6</u>

If the *client* controls the settlement decision, her expected recovery is given by **(9)** in the text; if the *lawyer* controls the decision, the plaintiff's expected recovery is given by **(11)** in the text. Suppose the client controls the decision: we have seen that as $\lambda \rightarrow 1$, $r_t^{**} \rightarrow r^*$, implying that her expected recovery approaches $(1-r^*)w^*$. However, if the lawyer controls the settlement decision, when $r_t = r^*$, her expected recovery is $(1-r^*)w^*+x_t^*+y^*$; the optimal bifurcated fee must yield at least as great a recovery as this. Thus -- for λ near 1 -- under the optimal fee, the client does better when the lawyer controls the settlement decision than when she herself controls the decision.

Yet when the client controls the decision and $\lambda \rightarrow 0$, we know that $r_t^{**} \rightarrow \hat{r}$ and $r_s^{**} \rightarrow \check{r}$, so her expected recovery approaches $(1-\check{r})\hat{w}$. In contrast, if the lawyer controls the settlement decision, $r_t^{**} \leq \hat{r}$, while $r_s^{**} \geq \check{r}$; so the plaintiff's expected recovery may be less than (an in any event never exceeds) the figure $(1-\check{r})\hat{w}$. Thus -- for for λ near 0 -- under the optimal fee, the client generally does better when she herself controls the settlement decision.

REFERENCES

Bebchuk, Lucian A., *Litigation and Settlement under Imperfect Information*, 14 RAND J. Econ. 404 (1984).

_____, *How Would You Like to Pay for That? The Strategic Effects of Contingency Fees and Retainer Arrangements on Settlement Negotiations* (unpublished manuscript, 1994).

Brickman, Lester, et al., *Rethinking Contingency Fees* (1994).

Hay, Bruce L., *Contingent Fees and Agency Costs*, Working Paper No. 162, John M. Olin Program on Law, Economics and Business, Harvard Law School (July 1995), forthcoming, 25 J. Legal Stud. (June 1996).

Hensler, Deborah R., et al., *Compensation for Accidental Injuries in the United States* (RAND Institute for Civil Justice, Publication No. R-3999-HHS/ICJ) (1991).

Landes, William M., *An Economic Analysis of the Courts*, 14 J. L. & Econ. 61 (1971).

Miller, Geoffrey P., *Some Agency Problems in Settlement*, 16 J. Legal Stud. 189 (1987).

Posner, Richard A., *An Economic Approach to Legal Procedure and Judicial Administration*, 2 J. Legal Stud. 399 (1973).

Shavell, Steven, *Suit, Settlement and Trial: A Theoretical Analysis under Alternative Methods for the Allocation of Legal Costs*, 11 J. Legal Stud. 55 (1982).