May 2004

THE (LEGAL) VALUE OF CHANCE: DISTORTED MEASURES OF RECOVERY IN PRIVATE LAW

Omri Ben-Shahar and Robert A. Mikos[*]
University of Michigan Law School
Ann Arbor, MI 48109
(734) 763-4608

Citation information:
Omri Ben-Shahar & Robert A. Mikos, *The (Legal) Value of Chance: Distorted Measures of Recovery in Private Law,* 7 AM. L. ECON. REV. 484 (2005)

[*] Ben-Shahar is Professor of Law and Economics, University of Michigan Law School (omri@umich.edu); Mikos is Acting Professor of Law, University of California, Davis (ramikos@ucdavis.edu). We would like to thank Hanoch Dagan, Tomotaka Fujita, Alon Harel, Jim Krier, Andrew Kull, Ariel Porat, an anonymous referee and workshop participants at Michigan, Stanford, Yale, and the ALEA 2001 Meeting in Georgetown for helpful comments. Financial support from the John M. Olin Center for Law and Economics at the University of Michigan Law School is gratefully acknowledged.

ABSTRACT

Parties who make investments that generate externalities may sometimes recover from the beneficiaries, even in the absence of contract. Previous scholarship has shown that granting recovery, based on *either* the cost of the investment *or* the benefit it confers, can provide optimal incentives to invest. However, this article demonstrates that the law often awards recovery that is neither purely cost-based, nor purely benefit-based, and instead equals either the *greater-of* or *lesser-of* the two measures. These hybrid approaches to recovery distort incentives to invest. The article demonstrates the prevalence of these practices, and explores informational and related reasons why they emerge. It argues that they generally are ill-suited to promote rational policies.

INTRODUCTION

Private parties often make investments that benefit others. Such investments are usually made under contract with the beneficiaries. The contract determines the investing party's right to recover and the measure of that recovery. Sometimes, however, a party considering making an investment is unable to contract with its potential beneficiaries. In these situations, the investing party must rely upon the law to create an obligation on the part of others to pay for the service.

Indeed, private law is replete with doctrines enabling an investing party to recover in the absence of contract. For example, a co-owner may recoup the costs of repairs she makes to co-owned property, and a doctor may recover a fee for treating an unconscious accident victim. In measuring the recovery, the law normally utilizes one of two approaches. In some instances, recovery is measured by the benefit from the investment: the obligor has to pay in accordance with the actual benefit she enjoyed. In other instances, recovery is measured by the investment's cost. A great deal of legal order has been created along this cost versus benefit distinction (Atiyah (1979, pp. 149-152, 184-189). For example, the law of torts defines obligations that are cost-based, whereas the law of restitution defines obligations that are benefit-based.

The right to recover in the absence of contract has been rationalized from an economic (that is, incentive-oriented) perspective (Landes and Posner 1978; Levmore 1985). In particular, it has been defended on the grounds that it encourages parties to make desirable investments that they would otherwise forego, usually because of the difficulty of contracting with the beneficiaries. This paper does not directly take issue with the economic literature demonstrating the desirability of imposing liability in such circumstances. Rather, it explores a systematic and puzzling inconsistency in the way the law actually determines the *magnitude* of liability. Using

economic analysis, it exposes confusion concerning the use of cost versus benefit to measure recovery, and the resulting distortion in incentives.

The common situation in which this inconsistency arises involves an investment that is expected to yield an uncertain benefit to another party. This is an investment that confers a *chance*, or probabilistic value, as opposed to certain benefit. Examples are abundant: owners of land make investments to repair or improve co-owned property that might, but are not certain to, increase the market value of the property; insured parties take non-contractible precautions that might, but are not certain to, reduce the losses that their insurers have to cover; attorneys invest litigation effort that might result in favorable judgments or settlements for their clients. By the time the law has to determine the recovery *ex-post*, the actual benefit—or lack thereof—becomes known and (usually) can be verified by the court.

If the investing party is entitled to recover, it might be expected that courts would measure the recovery either on the basis of the recipient's benefit, or on the basis of the investor's costs. The benefit-based measure would depend on (and is potentially equal to) the actual benefit that materialized. This is an *ex-post* recovery regime: the investing party will enjoy a high recovery when the court observes that the benefit is high, and a low recovery when the court observes a low benefit. The cost-based measure, alternatively, would *not* depend on any *ex-post* realization of benefit. Instead, and irrespective of whether the actual benefit is high or low, this measure would award a recovery that is fixed, equal to the reasonable economic cost of undertaking the investment. Under either the benefit-based or the cost-based regimes, if appropriately applied, the investment would be taken if and only if it is cost-justified.[1]

[1] Whether the benefit-based approach is superior to the cost-based approach (e.g., for reasons of fairness, information and administration costs, or risk) is beyond the scope of this analysis. *See, e.g.,* Polinsky and Shavell (1994) and Wittman (1985).

It turns out, however, that in many circumstances, the law takes neither a pure benefit-based nor a pure cost-based approach to measuring recovery. Instead, it often uses one of two "hybrid" recovery approaches. Under one approach, the investing party can recover either the *ex-post* benefit enjoyed by the beneficiary, or the cost of the investment, and *can elect the greater of the two*. This approach, which we label the "*greater-of*" regime, permits the investing party to recover the full benefit when it is high, or recover the cost of the investment when the benefit is low (or zero). The expected recovery under this approach is greater—potentially far greater—than the expected benefit of the investment, creating excessive incentives to invest.

Under a second hybrid approach, which we label the *"lesser-of"* regime, the investing party can again recover either the *ex-post* benefit enjoyed by the beneficiary, or the cost of the investment, but this time she is *limited to the lesser of the two*. The investing party can effectively recover the full benefit only when it is low; when the benefit turns out to be high, recovery is capped at the cost of the investment. The expected recovery under this approach is lower than the expected benefit of the investment, in fact, it is lower than the cost of the investment, thus creating insufficient incentives to invest.

The argument that hybrid regimes are distortive is straightforward and can be fully captured by the following lottery metaphor. Suppose party A owns a lottery ticket that provides a 1% chance of winning $1000, and a 99% chance of winning 0. The *ex-ante* value of such a ticket—its actuarial cost—is $10. Party A mistakenly loses her lottery ticket at party B's home and discovers the loss after the lottery draw was announced. Under the pure benefit-based recovery regime, party A can recover from party B (who found and cashed the ticket) either 0 or $1000, depending on the ticket's actual draw. Under the pure cost-based recovery regime, party A can recover the *ex-ante* value, or the cost of the ticket, $10, independent of the actual draw.

The expected recovery under both regimes is $10, correctly reflecting the value of the ticket at the time it was lost. Consider, in contrast, the two hybrid approaches described above. Under the *greater-of* approach, party A can recover $1000 if the ticket wins, and can recover $10 if the ticket's draw is 0. The expected recovery is approximately $20, twice the *ex-ante* expected value of the ticket (the ticket is worth more if lost; party A would have an incentive to lose it!). Under the *lesser-of* approach, party A can recover only $10 if the ticket wins, and recover 0 when the ticket's draw is 0. The expected recovery is 10 cents, well below the *ex-ante* expected value of the ticket.

Recognizing the distortive nature of hybrid regimes is, of course, not the main focus of this paper. In fact, given how obviously contorted and internally inconsistent these regime are, the paper will have little more to say in terms of their incentive effects. Rather, the main purpose to be pursued here is twofold, first, to demonstrate that hybrid regimes are a surprisingly common feature in the law, despite their distortive effects, and second, to explain this puzzling legal phenomenon.

To this first end, Section II of the paper (which follows a brief description of the structure of hybrids, in Section I) surveys a variety of recovery doctrines that incorporate the hybrid approaches. The analysis in that Section is more than doctrinal. The recognition that a particular hybrid regime exists usually requires more than reading of the text of the rule. It requires a more nuanced understanding of the way various rules work together, and the way plaintiffs (or defendants) can elect the recovery measures. With careful attention to such details, we identify a host of examples of prominent recovery doctrines incorporating the hybrid approach.

To illustrate, consider a party who invests in a project potentially adding value to a neighbor's land, in a setting that gives rise to a restitutionary right of recovery. A benefit-based

regime would set the recovery equal to the actual enhancement value enjoyed by the neighbor. A cost-based regime would set the recovery equal to the cost of the investment, if it is adjudged reasonable. A *greater-of* regime would permit the investing party to recover the full benefit when the enhancement value is high, and recover her costs when the enhancement value is low. Section II.E of the paper will demonstrate that this is a recovery strategy available to investing co-owners under the repairs and improvements doctrine in property law. On the other hand, a *lesser-of* regime would limit the investing party to recover the full benefit when this benefit is low and only her costs when the benefit is high. Section II.F of the paper will demonstrate that under the Restatement of Restitution this is the recovery schedule available to mistaken improvers, e.g., persons who mistakenly built an improvement on their neighbors' side of the property boundary.

Some readers might remain skeptical even after reading Section II and reviewing the various examples we provide. Are courts really so oblivious to common sense? How could the law apply such distortive recovery schemes? To address this skepticism, the paper turns in Section III to explore why the hybrid approaches are used so frequently in the law, despite the obvious distortions they create. This inquiry demonstrates that courts occasionally employ hybrid rules inadvertently, due to systematic problems in adjudication, such as the information structure and the difficulty of drawing boundaries between related causes of action. Other times, courts recognize the fact that they are applying a hybrid regime, and nevertheless do so deliberately, to adjust the magnitude of recovery in pursuit of aims that may be unrelated to investment incentives. In these cases, however, our analysis shows that hybrid regimes add an arbitrary component to the recovery amount, and are therefore a clumsy instrument to serve their stated purposes.

As the title of the paper—The (Legal) Value of Chance—suggests, this paper can be interpreted beyond the concrete problem it identifies (and in the Conclusion Section we briefly explore possible interpretations). The paper can be read more broadly, as an observation on how the law values uncertain or stochastic value. Here, costly actions that are identical from an *ex-ante,* cost-benefit perspective, can appear dissimilar *ex-post*, once the stochastic benefit from them materializes. This appearance can lead courts to apply an inconsistent treatment of the right to recovery, bouncing in an arbitrary fashion between cost-based and benefit-based liability. Hence, a computation problem that is elementary for an economist can become difficult to resolve consistently under the decision making framework in which courts operate.

I. A Simple Model of Investment with Uncertain Benefit

A. The Framework of Analysis

Suppose that party A can spend a cost C that would yield a random benefit to party B (and 0 benefit to party A). For simplicity and without loss of generality, assume that the benefit to party B might be either high or low, with the high benefit denoted by V and the low benefit fixed at 0. The exogenous probabilities of V and 0 are p and $1-p$, respectively. Assume also that the parties are both risk-neutral and are unable to form a contract governing this investment. (Some of these assumptions are relaxed later.)

It is socially desirable for party A to spend C if and only if:

$$C \leq pV.$$

Party A's private decision whether to incur C would depend, however, on the legal regime governing her right to recover for her efforts.

B. The Pure Benefit-Based and Cost-Based Regimes

Under the benefit-based recovery regime, party A can recover the full *ex-post* benefit conferred upon party B. Thus, party A can recover V with probability p and 0 otherwise, for an expected recovery of pV. Party A will invest if and only if $C \leq pV$, which is socially optimal.[2]

Under the cost-based regime, party A's recovery is independent of the *ex-post* realization of benefit, and equals the *ex-ante* value of her investment. We consider two versions of an *ex-ante* regime. In theory, a "pure" *ex-ante* regime is one in which the measure of recovery is the average, or expected benefit pV. If party A is entitled to recover pV regardless of the realization of benefit, she will invest if and only if $C \leq pV$, which guarantees socially optimal investment. The more practically significant version of the *ex-ante* regime is the pure cost-based regime, in which the recovery is equal to party A's cost of investment, C. To guarantee that party A will invest C only when socially optimal, recovery should be conditional on $C \leq pV$; namely, on the cost being "reasonable." Like the benefit-based regime, the cost-based regime generates optimal incentives to exert effort.

C. The Hybrid Regimes

1. *The* greater-of *regime*

Under one type of hybrid regime, the *greater-of* regime, party A receives the *ex-post* benefit measure when the realization of benefit is high, and receives her cost when the realization of benefit is low. Thus, party A receives a recovery of V when the benefit is V and receives a recovery of C when the benefit is 0. Party A's net expected payoff is:

$$pV + (1 - p)\, C - C = p(V - C),$$

[2] Other hypothetical "*ex-post*", or benefit-based recovery rules which make recovery dependent on—but not exactly equal to—the actual benefit, still provide optimal incentives, as long as the expected recovery equals the expected benefit. For example, optimal incentives are provided under an entire family of "multiplier rules", under which, for all $m \geq 1$, party A can recover mV from a fraction $1/m$ of the V-type beneficiaries, and 0 otherwise. The expected recovery under such rules is again pV. The m multiplier can represent, for example, a recovery enhancement in situations where the likelihood of successful suit is only $1/m$. *See, e.g.*, Polinsky and Shavell (1998).

which is greater than $pV - C$, the net social gain from investment, for all $p < 1$. Party A will over-invest: whenever $pV < C \leq V$, party A will invest although it is too costly from a social point of view. Intuitively, the distortion in this case arises from the fact that, with some likelihood, the investing party will externalize her cost. If a high benefit occurs, the investing party internalizes both the cost and the benefit; but if the benefit is low, the cost of the investment is externalized.[3]

To illustrate, consider an investment of $120 that yields an expected benefit of $100. Socially, it is inefficient. If, however, the benefit is probabilistic, with 50% chance of $150 and 50% chance of $50, party A will take it. Under the *greater-of* regime, party A will get either $150 (when the benefit is $150) or $120 (when the benefit is $50), for an expected recovery of $135, well exceeding the cost of the investment. The greater the variance of the benefit, the greater are party A's excessive incentives to invest (e.g., if the benefit is either $200 or $0, equally likely, the expected recovery balloons to $160.)

This illustration can be generalized: the *greater-of* regime is equivalent to a benefit-based regime compounded by a put option for party A – an option to "sell" the benefit for C. If the *ex-post* value arising from the investment falls below C, party A will exercise the option and receive C; and if the *ex-post* value realizes above C, party A will not exercise the option and will receive instead the full *ex-post* value, V. Thus, the distortion under the *greater-of* approach depends on factors similar to those that affect option prices, such as the variance of the distribution of

[3] A similar distortion arises when the *ex-ante* measure of recovery equals pV, rather than C. In this case, a *greater-of* regime entitles party A to recover the actual benefit or the expected benefit whichever is higher. Accordingly, party A's expected recovery under this hybrid rule is $pV + (1 - p)pV = pV(2 - p)$, which exceeds the expected social benefit, pV, whenever $p < 1$.

benefits,[4] and the time period that party A has to exercise the option (Jackson 1978).

2. *The* lesser-of *regime*

Under the *lesser-of* hybrid regime, party A receives the *ex-post* benefit when the realization of benefit is low, but receives only her cost when the realization of benefit is high. Thus, party A receives a recovery of 0 when the benefit is 0 and a recovery of C when the benefit is V. Under the *lesser-of* regime, party A's expected net recovery is:

$$pC + (1-p)0 - C = -(1-p)C,$$

which is negative, and in particular it is less than $pV - C$, the social gain from investment, for all $1 > p > 0$. The *lesser-of* regime generates no incentives to invest: whenever $0 \leq C \leq pV$, party A will not invest although the investment is socially desirable. Intuitively, the under-investment arises from the fact that the investing party internalizes the entire cost but does not get to enjoy the entire benefit.[5]

To illustrate, consider an investment of $80 that yields an expected benefit of $100. Socially, it is an efficient investment. If, however, the benefit is probabilistic, with 50% chance of $150 and 50% chance of $50, party A will *not* take it. Under the *lesser-of* regime, party A will get either $80 (when the benefit is $150) or $50 (when the benefit is $50), for an expected recovery of $65, well below the cost of the investment.

[4] For example, consider a perturbation of the benefit values to $\{a, V-b\}$ for some small a,b, such that the mean of the distribution remains unchanged, pV (namely, if the probability of the high value $V-b$ remains p, $b = a[p/(1-p)]$. For this reduced-variance distribution, the expected net recovery under the *greater-of* regime is $p(V-a)+(1-p)C - C = p(V-C) - pa$, which is smaller (by an amount pa) than the expected recovery under the higher-variance distribution. The smaller the variance (a higher a), the lower the expected net recovery.

[5] Here, as in the *greater-of* regime, the magnitude of the distortion is equal to the value of an option, this time a call option given to party B, to "buy" the benefit at a price of C. If the *ex-post* benefit realizes above C, party B will exercise the option and pay only C; and if the *ex-post* benefit realizes below C, party B will not exercise the option and instead will pay the benefit. The value of this option is the amount by which party A is under-compensated.

II. THE DOCTRINAL PREVALENCE OF HYBRID REGIMES

Given the apparent shortcomings of the hybrid approaches, the reader may be left wondering whether these devices are of any practical concern. Accordingly, the objective of this Section is to demonstrate the prevalent use of the hybrid approaches across a broad range of legal doctrines, and thereby to dispel any notion that hybrids are a mere esoteric phenomenon. Understanding the context in which hybrid regimes operate will also help us develop, in Section III, a more general discussion of the reasons--good or bad--why hybrids are used with such surprising frequency.

A. Remedies for Breach of Contract

A *greater-of* approach is embodied in the choice of remedies available for breach of an explicit contract. There are two typical situations in which the election-of-remedy rules entitle the aggrieved party to a *greater-of* recovery schedule. The first situation involves total breach or repudiation of a contract after one party has partially performed. The typical example is a contractor who was discharged after performing part of the service. The aggrieved contractor may seek either expectation damages or restitution. That is, he can either enforce the bargain and sue for "make whole" damages, calculated in accordance with the contract price, or disaffirm the materially breached bargain—employ a legal fiction that the contract does not exist—and recover damages equal to the benefit conferred on the breaching party (Restatement § 373; Kull, 1994).

This is a *greater-of* hybrid regime. When the benefit the breaching party enjoys from partial performance is low, the aggrieved party would seek the standard expectation remedy, and will recover the costs actually incurred, plus the profit that would have been made had the contract been fully performed. But when the benefit to the breaching party from the partial

performance is high, the aggrieved party can choose to recover more than the cost-plus-profit measure, by recovering instead the *ex-post* value of the partial performance. For example, a contractor who was discharged after partially performing, was allowed to recover the value of a structure to the client, over $250,000, rather than the contractual fee, which was only $20,000.[6] In fact, procedural rules enable the aggrieved party to join in the complaint a claim for restitution recovery (in quantum meruit) and a claim for expectation damages, thus postponing the election of the remedy until it becomes clear, at trial, which of the two measures is greater. [7]

This *greater-of* regime is reinforced by the way restitution damages are calculated. Under the Restatement of Contracts, the benefit to the breaching party may be measured by either the market price for furnishing a service, *or* the extent to which the beneficiary's property has been enhanced in value by the service. When the enhancement-in-value measure of benefit is low, the aggrieved party is encouraged to seek the more generous market-price (cost-based) measure of restitution,[8] and when the enhancement-in-value measure is high, the aggrieved party is entitled to seek this larger sum. As a leading case summarizes: "the rule has evolved that the proper measure of damages in unjust enrichment should be the *greater of the two measures*."[9]

A second typical situation in which the *greater-of* damage measure applies is in an action for breach of warranty of title. A buyer who purchases an asset from a seller who is not the true

---

[6] Boomer v. Muir, 24 P.2d 570, 577 (Cal. App. 1933) ("[U]pon prevention of performance the injured plaintiff may treat the contract as rescinded and recover upon a quantum meruit without regard to the contract price."); *see also* Kull (1994, pp. 1477, 1498).

[7] Matarese v. Moore-McCormack Lines, Inc., 158 F.2d 631 (2nd Cir. 1947) (permitting amendment of the complaint from a suit based on express contract to one based on the theory of unjust enrichment); Frontier Management Co. v. Balboa Ins. Co. 658 F. Supp. 987, 994 (D. Mass. 1986) (holding that plaintiff may plead unjust enrichment merely on the possibility that its contractual claims will prove inadequate at trial).

[8] REST. 2D CONTRACTS §371 cmt. b ("[T]he reasonable value to the party from whom restitution is sought is usually greater than the addition to his wealth. If this is so, a party seeking restitution for part performance is commonly allowed the more generous measure of reasonable value. . .").

[9] Robertus v. Candee, 670 P.2d 540, 543 (emphasis added).

owner and later has to surrender the purchased asset to its true owner, can recover from the seller either the purchase price or the *ex-post* value of the asset at the time it was surrendered, whichever is greater. Thus, when the asset depreciates in value below the price paid, the buyer can recover the price. And when the asset's value increases, the buyer can recover the full value, uncapped by the contract price.[10] Effectively, the aggrieved buyer is granted the right to recover the actual value of the asset at the time of breach, along with a put option to sell this right for the original price. The excess recovery under the hybrid regime equals the value of this put option. The more volatile the price of the asset, the more valuable is the option, and the greater the excess recovery.

It might be perceived that the right to elect restitution damages even when they exceed expectation damages is restricted to situations in which breach is so egregious that greater deterrence is necessary. Indeed, as we will see below, the egregiousness of the breach can explain the emergence of *other greater-of* recovery rules, particularly those involving fiduciary and agency relationships. There, embezzlement by a fiduciary is the very reason that gives the victim the enhanced recovery rights. But in the general doctrine of contract remedies, case law does not conform with the "egregiousness" hypothesis: while the right to restitution damages does depend on breach being "total", it is not dependant on the breaching party's *mens rea*. The only restriction the Restatement imposes on the power to elect the greater remedy applies in a narrow set of circumstances, when the aggrieved party has performed in full and the breaching party's contractual obligation is a "definite sum of money."[11] Then, the aggrieved party is limited

[10] Menzel v. List, 246 N.E.2d 742 (N.Y. 1969); WILLISTON (1960, §1395A) (limiting damages to the purchase price "virtually confines the buyer to rescission and restitution, a remedy to which the injured buyer is undoubtedly entitled *if he so elects*, but it is a violation of general principles of contracts to deny him in an action on the contract such damages as will put him in as good a position as he would have occupied had the contract been kept").

[11] Restatement (Second) of Contracts § 373(2).

to expectation damages. Otherwise, the aggrieved party is generally unrestricted in his freedom to elect the greater of the two measures.[12]

B. Implied Contracts

In drafting the provisions of their contract, parties are free to determine whether the beneficiary's obligation to pay is to be based on effort (as are most contracts), or whether it is to be contingent on measurable benefit. In the absence of an explicit contract between the parties, however, it is up to the law to determine the recovery for services rendered.

Normally, the absence of an explicit contract indicates that the beneficiary did not wish to be charged, and thus there is no obligation to pay. However, when the absence of an explicit contract is due to high transactions costs, the doctrine of implied contracts may be invoked and an obligation to pay may be imputed. A typical situation in which an implied obligation arises is when one party performs a service during negotiations over a contract. The party providing the service may do so in anticipation of striking a deal, at the encouragement of the other party, or in an attempt to convince the other party that a deal is desirable. For example, an advertising agency might develop an idea for an advertising campaign and, in bidding for the client's account, share it with the client. If negotiations eventually break down, the investing party might seek to recover its costs or the value it created for the other party

Courts are careful not to impose an implied obligation where an explicit one could have been negotiated. But when the liability hurdle is cleared—when an implied contract is recognized—courts distinguish between two types of obligations that might be imputed, labeled implied-in-fact and implied-in-law contracts. An implied-in-fact contract may be found where

[12] Corbin (1964, § 1113) ("[t]he generally prevailing rule is that the plaintiff's recovery of the value of the consideration received by the defendant is neither measured by nor limited by the contract price or rate.")

actions other than an express promise indicate that the beneficiary intended to pay for the service. Here, as in many other areas of contract law, the parties' expectation is determined not merely from the text of their agreement (or lack thereof), but from the context as well. An implied-in-law contract, in contrast, arises even in the absence of any reliable indication of the parties' intentions. It is based, instead, on the benefit, and is intended to strip the beneficiary of this gain if the acquisition is deemed unjust under the established principles of the law of restitution (Farnsworth 1999, pp. 499-501).

The two types of implied contracts also lead to different measures of recovery. An implied-in-fact contract, once inferred, is supplemented by courts to include a provision mimicking the fee that an express contract would have stipulated, which is usually (though not necessarily) calculated on a per-effort basis. In the advertising contract example above, it would require the client to pay the firm for the billable hours it spent on the project. In contrast, an implied-in-law obligation, once constructed, often leads to restitution of the full benefit enjoyed by the beneficiary.[13] The client would have to pay the value it actually derived from the advertising campaign, which can potentially differ from the contract fee.

Put in terms of the analysis in Section I, the implied-in-fact doctrine embodies a cost-based (or fee-based) recovery approach, whereas the implied-in-law doctrine embodies a benefit-based recovery approach. Either regime, if applied consistently and in the appropriate situations, can lead to optimal recovery for pre-contractual effort. A distortion arises, however, when the plaintiff can elect the *greater of* the two recovery measures. In particular, when courts allow a

[13] In rewarding the value the benefit conferred, courts use one of two possible measures, equal either to the "net enrichment", namely, the increase in total wealth to the beneficiary, or to the "cost avoided", namely, the saving to the beneficiary in obtaining the service (Farnsworth 1999, p. 107; RESTATEMENT (SECOND) OF CONTRACTS, §371). It is only when a "net enrichment" measure is applied that the recovery under an implied-in-law claim differs from the recovery under an implied-in-fact claim.

party who conferred a high benefit to seek the restitutionary implied-in-law recovery for the entire benefit, and a party who conferred a low (or zero) benefit to seek the implied-in-fact recovery for the per-effort fee, excessive recovery results. She will recover the full benefit when her effort succeeded, and her reasonable fee otherwise, which is more than the expected benefit from her effort.

The implied-contracts doctrine is structured in a way that falls into this *greater-of* trap. While acknowledging the difference between the two types of claims, courts often accord plaintiffs the power to choose between them. In the casebook favorite *Hill v. Waxberg*, for example, a contractor who was negotiating a building project invested in "plans, ideas, and efforts" that benefited the landowner after negotiations broke down. In allowing a recovery, the court didactically distinguished between the two types of implied contracts and their associated recovery measures, and confirmed the plaintiff's right to choose between them.[14] Accordingly, it was suggested that, even when no actual benefit materializes, an implied-in-fact claim for the reasonable fee should lie (Farnsworth, 1987, p. 232). Thus, when the benefit conferred upon the other party is low, the investing party is generally encouraged to seek a recovery of her cost or hypothetical fee, based on an implied-in-fact contract claim.[15] And when the benefit is high, the investing party is not precluded from making an implied-in-law contract claim for the full benefit conferred.

One may wonder whether cases in which the plaintiff is permitted to select the greater of the two implied contract recovery measures are all that common, and thus whether the problem

[14] The court noted that "the elements of either theory could be satisfied, but since counsel has declined *to choose between them*, we are not prepared to make the choice for him." 237 F. 2d 936, 939 (9th Cir. 1956) (emphasis added).

[15] *Earhart v. William Low Company,* 600 P.2d 1344 (Cal. 1979) (in the absence of actual benefit to the defendant, the recovery of expenses incurred is allowed).

we identify here is of much practical significance. Specifically, if the circumstances under which each type of claim could arise are easily distinguished, the problem of overlap would arise only in exceptional cases. Our own understanding of the law of implied contracts suggests, however, that the problem we identified is not uncommon. For one, whenever a court is willing to recognize an implied contract, the plaintiff is generally able to satisfy the elements of both types of implied-contract claims. That is, while courts have repeatedly distinguished the two types of obligations and their respective measures on conceptual grounds, the circumstances under which each obligation arises are common, if not identical. Both claims arise when there is intent to charge for a benefit that is desired by the beneficiary, and when high transactions costs interfere with the drafting of an explicit contract (Posner 1998, pp. 151-2). Indeed, even when denying implied contract claims, courts invoke identical tests for each of the two types of obligations, focusing on why an explicit contract was not made.[16] Thus, it is not the confusion of an outlier court that creates this *greater-of* regime, but rather the core feature of the implied contract doctrine, which allows two types of claims to coincide.

Moreover, there are strong reasons to believe that a *greater-of* regime is regularly implemented in this context. First, courts explicitly permit it. When an implied-in-law action fails due to the absence of proof concerning the magnitude of the benefit, courts allow recovery to be based instead on an implied-in-fact claim for the standard fee.[17] Similarly, plaintiffs are permitted to offer several alternative theories of recovery and to delay their commitment to any particular remedy until the stage of trial at which it will become clear which recovery measure is

[16] Nowhere is this identity of grounds so obvious as in the casebook favorite Bailey v. West, 249 A.2d 414 (R.I. 1969). In deciding that the plaintiff, who maintained the defendant's horse for four years, cannot recover any fee, the Supreme Court of Rhode Island explained that an implied-in-fact contract does not exist because defendant indicated that he "would not be responsible for boarding the horse"; similarly, an implied-in-law contract does not exist because the defendant notified that he "would not be responsible for [the horse's] keep."

[17] Bastian v. Gafford, 563 P.2d 48 (Idaho 1977)

higher, and even to amend the complaint if they originally stated only the lower theory of recovery. Second, the existence of a *greater-of* regime here is less puzzling in light of the explicit adoption of a *greater-of* regime in the recovery of restitution damages. The implied-in-law/implied-in-fact dichotomy is a close parallel to the restitution/expectation dichotomy in computing damages for breach of an explicit contract, where—we saw—the doctrine openly embraces a *greater-of* regime.

In the end, the question remains: how distortive is the *greater-of* regime in measuring the recovery for failed contracts. We have focused on some elements of the doctrine, but the full picture is more nuanced than sketched here. Other limitations on recovery may at times offset the over-compensation effect of the simple *greater-of* regime. In bouncing between cost-based, fee-based, and benefit-based measures of recovery, courts are trying to tailor damage remedies that serve what they perceive to be the ends of justice. Thus, alongside the rule allowing plaintiffs to elect the greater-of several recovery measures, courts apply other rules such as damage caps to temper the risk of over-compensation. Still, in the interface between restitution and contract remedies, pockets of hybrid regimes occupy a significant domain of the doctrine. Recognizing the pattern of their existence indicates that a more systematic solution, not reliant on ad-hoc adjustments, is desirable.

C. Attorney Fees

Recovery of attorney fees is commonly governed by a hybrid regime. This Part considers three prominent examples. The first involves a trial attorney's right to recover from her client after being discharged without cause prior to the conclusion of litigation. The second example concerns a defendant's right to seek indemnification of litigation expenses that run to the benefit

of third parties. The third and final case regards a litigant's right to recover attorney fees from *opposing* parties.

1. *Discharge of an attorney-client contract*

Trial attorneys are typically compensated using one of two possible formulae. Under one approach—the billable hours contract—the attorney is paid the same fee regardless of the outcome of the litigation. This fee is calculated by multiplying the number of hours the attorney worked on the case by a pre-agreed hourly rate. The alternative approach is the contingency-fee contract, under which the attorney is paid a portion of the client's award. If the client's claim is denied, the attorney recovers nothing, but if the client's claim prevails, the attorney receives a substantial premium *vis a vis* the billable hours contract.

To protect the interests of clients to be represented by attorneys they trust, courts have traditionally held that a client has an "unfettered" right to discharge an attorney working under either type of contract. When the client exercises this right before the conclusion of litigation and dismisses her attorney without cause, the question arises as to whether and how the dismissed attorney is to be compensated for the services she has already provided the client. If the attorney worked under an hourly fee contract, he is of course paid for the hours billed, regardless of the outcome of the litigation (Annotation, 1957, p. 616). Here, in mimicking the contract price, the law provides a pure cost-based recovery. But if the attorney worked under a contingency fee contract, figuring out her compensation is more complex. This situation fits well into the framework of this paper, because the recovery of the attorney has to be determined at the time the lawyer is discharged, before the client's case is resolved, while the benefit from his efforts is still probabilistic.

Strikingly, in surveying the different state rules governing discharge of the attorney

working under a *contingency-fee* contract, we can identify all possible approaches to recovery of probabilistic benefit (Annotation, 1998). Some jurisdictions apply a pure benefit-based recovery approach that simply enforces the contingency-fee agreement. As soon as the underlying litigation concludes or settles—that is, as soon as the "benefit" to the client, if any, becomes known—the dismissed attorney recovers her full contingency fee, minus any expenses not incurred by the attorney in performing the balance of the contract.[18] If the suit is ultimately successful, recovery is high; otherwise, the attorney recovers nothing.

Other jurisdictions apply a variant of a pure cost-based approach, which permits the dismissed attorney to recover, under a damages remedy known as a quantum meruit claim, the reasonable value of her services (i.e., her costs), but not the contingency fee. Neither the attorney's right to recover, nor the amount of that recovery, are affected by the outcome of the litigation.[19] Thus, recovery often resembles what the attorney would have received under a guaranteed billable-hours contract.

In theory, a *lesser-of* regime might still emerge in these "pure" cost-based recovery jurisdictions, if the client is permitted to dismiss the attorney after the client obtains new information on the likelihood of success in her suit. Under such a scenario, when the client learns that the suit is about to succeed (or to reach a favorable settlement), the client would dismiss the original attorney and pay the attorney her costs, and when the client learns that the suit is about

[18] Tonn v. Reuter, 95 N.W.2d 261 (Wis. 1959). A few jurisdictions that permit an attorney to recover in quantum meruit the benefit received by the client, as opposed to the cost incurred by the attorney, likewise employ a benefit-based approach. [This sentence you added would be difficult for reader to understand, because we haven't explained yet what "quantum meruit" means. Instead, it might be better to beef up this FN with the names of states that actually use the stated regime.]

[19] *See, e.g.*, Lai Ling Cheng v. Modansky Leasing Co., Inc., 539 N.E.2d 570 (N.Y. 1989) (holding that recovery in quantum meruit may be "more or less than the amount provided in the contract"); Trenti, Saxhaug, Bergen, Roche, Stephenson, Richards, & Aluni, Ltd., v. Nartnik, 439 N.W.2d 418 (Minn. Ct. App. 1989) (dismissed attorney may recover in quantum meruit a sum based on hours worked multiplied by an hourly rate, regardless of the outcome of

to fail, the client would retain the attorney and pay her the contractual contingency fee – nothing. However, courts recognize the danger of such manipulation and, when detected, allow the discharged attorney a recovery equal to the full contingency fee.[20] A similar *lesser-of* approach may also arise whenever the client has limited financial resources. In that case, if the client loses her suit, the attorney will not get paid, and if the client wins her suit, the attorney will recover only the hourly fee figure. As will be explained in Section III, this type of the "lesser-of" scheme is more difficult for courts to avoid without switching to a pure benefit-based recovery approach, because they would be required to take into account counter-factual scenarios, that is, to adjust recovery in one state-of-the-world ("client wins") to make up for the potential of non-recovery in another, hypothetical state-of-the-world ("client loses").

Interestingly, a third set of jurisdictions applies a *greater-of* approach by permitting the dismissed attorney to elect her remedy. The attorney may collect either the reasonable value of her services *or* her contingency fee minus any expense saved because of the termination.[21] Therefore, if the award the client eventually collects is high, the attorney may elect the benefit-based measure of recovery, namely, a fraction of the client's award; otherwise the attorney may elect the cost-based measure of recovery, namely, her hourly fee.

Of course, application of this *greater-of* regime is controlled (one would expect) by the client, and a client will likely choose not to dismiss an attorney without cause where doing so

---

former client's case); Martin v. Camp, 114 N.E. 46 (N.Y. 1916)(under the New York rule, the attorney's right to recover the cost of services performed accrues immediately upon his discharge).

[20] *See, e.g.*, Fracasse v. Brent, 494 P.2d 9, 14 (Cal. 1972) (holding that an attorney discharged without cause under a contingency-fee contract is normally limited to recovering the reasonable value of his services, but may recover the full contingency fee when discharge occurs "on the courthouse steps").

[21] *E.g.*, Lockley v. Easley, 78 S.W.2d 573 (Ark. 1990); *In re* Downs, 363 S.W.2d 679, 686 (Mo. 1963) (a discharged contingency fee attorney "has the election to claim a reasonable fee for the work done [...] or to wait until the claim is liquidated by judgment or settlement and then sue [...] for his contract fee").

will increase the expected fee owed to the attorney. For example, if the client discovers that the suit is likely to fail, the client will have a disincentive to dismiss the attorney because she will have to pay the attorney her costs, whereas by retaining the attorney, the client will ultimately have to pay her nothing when the suit fails. Still, the "greater of" rule that looms in the background would cause the client to refrain from discharging an attorney whom she no longer trusts, and all the more so in cases that are headed towards litigation failure—the very cases in which it is more likely that the client disagrees with her attorney. Thus, the "greater-of" recovery can be avoided, but not before it would generate over-deterrence and undermine the stated policy of enhancing clients' economic freedom to discharge their attorneys. The courts that have adopted the *greater-of* approach have failed to recognize this consequence.

Lastly, many jurisdictions, including—after long judicial deliberations—California, apply a *lesser-of* approach by limiting the attorney's recovery to the reasonable value of the services and then, only if the client's suit is ultimately successful.[22] Thus, if the client receives a high award, the attorney gets the cost-based measure of recovery (her hourly fee), whereas if the client receives no award, the attorney gets the benefit-based measure of recovery (nothing). Courts are applying this *lesser-of* approach by design:

> [T]he better rule is that because a clent has the unqualified right to discharge his attorney, fees in such cases should be limited to the value of the services rendered or the contract price, *whichever is less*.[23]

It should be emphasized that in measuring the "reasonable value" of the attorney's services, courts distinctly utilize the cost-based measure—the attorney's hourly fee multiplied by the number of hours he worked—and not the benefit-based measure that reflects the magnitude of

[22] *E.g., Fracasse*, 494 P.2d at 14; *see also* Rosenberg v. Levin, 409 So.2d 1016, 1021-22 (Fl. 1982 (holding that "fees [...] should be limited to the value of the services rendered or the contract price, whichever is less").

[23] Chambliss, Bahner & Crawford v. Luther, 531 S.W.2d 108, 113 (Tenn. Ct. App. 1975) (emphasis added).

the client's award.[24] But at the same time, courts are weary not to enable clients to take advantage of this extremely client-friendly regime, by, for example, discharging the attorney on the courthouse steps, after the attorney has completed most of his work. If courts suspect the client of engaging in such tactics, they will increase the measure of "reasonable value" to equal the full (benefit-based) contingency award. Thus, the "lesser-of" regime is limited (at least in theory) to "honest" dismissals.[25]

Jurisdictions following the *lesser-of* approach have cited several reasons to reject either pure approach. The pure cost-based regime, according to some courts, is unfair to poor clients who cannot afford to pay the attorney's fees unless the client recovers in the suit. At the same time, the pure benefit-based regime, according to these courts, would place an undue burden on the client's right to dismiss her attorney, because the client might end up having to pay two attorneys a full contingency fee.[26] The lesser-of recovery scheme would seem to address both concerns.

What these courts fail to recognize is that giving clients unbridled freedom to terminate an existing contingency relationship with an attorney may deter attorneys from agreeing to represent some clients in the first instance. In the alternative, attorneys may charge their clients higher rates (to reflect the risk of dismissal) or switch to billable-hours contracts, which are

---

[24] S*ee* Annotation, Attorneys at Law, 7 Am. Jur. 2d pp. 312-13 (noting that courts rarely permit recovery that exceeds "the number of hours reasonably expended on the litigation multiplied by a reasonable hourly rate"); Boyette v. Martha White Foods, Inc., 528 So.2d 539 (Fla. App. 1988) (reversing award of $15,900 to dismissed attorney because the hourly fee value of the services was only $5,300; the district court had trebled that amount due to the contingency risk in the case).

[25] As noted above, the *Fracasse* court, which articulated the California version of the *lesser-of* rule, recognized this danger of manipulation and thus emphasized that attorneys who are dismissed at the last minute can still recover their full contingency fees, as an exception to the general rule that limits their recovery to hourly fees. 494 P.2d at 14.

[26] *See, e.g.,* Rosenberg v. Levin, 409 So.2d 1016 (Fla. 1982) (reasoning that "there is an overriding need to allow clients freedom to substitute attorneys without economic penalty"); *Fracasse*, 494 P.2d at 12-14.

governed by different rules. Like other mandatory terms imposed on contracting parties, the freedom to dismiss an attorney without cause comes at a price, one some clients might prefer not to pay.[27]

A similar issue concerning recovery by contingency fee attorneys has come to the fore in the settlement of state lawsuits against tobacco companies. The contractual arrangements between the states and their outside (i.e., private) attorneys usually entitled the attorneys to a pure benefit-based recovery measure, anywhere between 2% and 25% of the settlements. When the tobacco industry agreed to settlements involving enormous sums, the attorneys' combined fees reached billions of dollars. *Ex-post,* this translated to hourly fees reaching, in some cases, tens of thousands of dollars per hour. At that stage, lawmakers were ready to discharge the contingency fee arrangements and override the contracts.[28] Recognizing that these fees overwhelmingly exceed standard legal hourly rates, commentators, judges, the Press, and many lawmakers called these fees excessive, exorbitant, and even unconscionable (Brinkman, 1998).

Critics of the fees are effectively advocating a *lesser-of* recovery regime. If the suits had been unsuccessful—as were most tobacco suits prior to the settlement—the plaintiffs' attorneys would have recovered no fees. But now that the states have prevailed against the tobacco companies, the attorney fees have been scrutinized relative to hourly fees, and—as many critics endorse—capped not to exceed standard (i.e., guaranteed) hourly rates. What critics overlook is the enormous risk that many of these attorneys (albeit not all) had taken at the outset of the litigation. *Ex-ante*, in light of the slim chance of victory against the tobacco industry and the

---

[27] *See, e.g.*, Richard Craswell, *Passing on the Costs of Legal Rules: Efficiency and Distribution in Buyer-Seller Relationships*, 43 STAN. L. R. 361 (1991); M. J. Trebilcock, *The Doctrine of Inequality of Bargaining Power: Post-Benthamite Economics in the House of Lords*, 26 TORONTO L. J. 359 (1976).

[28] *See, e.g.,* 144 Cong. Rec. S6373-01, S6374 (remarks by Rep. Sessions) ("How can we violate contracts? We violate contracts all the time in this body. […] Everything about the tobacco business is being changed by this legislation. […] One of those aspects ought to be how much these fees should count for").

projected out-of-pocket cost to be incurred by the attorneys, the negotiated contingency fees seem less excessive. Measuring in hindsight the per-hour fee that the attorneys in fact recovered overlooks this risk factor. It is equivalent to the view that the holder of a winning lottery ticket is unjustly enriched by collecting the award and that he should recover no more than the price paid for the ticket. In this *lesser-of* regime, attorneys would be less willing to undertake risky projects under contingency fee arrangements.

2. *Indemnification of litigation costs*

Another regime applying a hybrid approach to the recovery of attorney fees involves indemnification of litigation costs. A party expending litigation costs to the benefit of others may seek reimbursement from the beneficiaries, even in the absence of an express indemnification agreement. The most common situation in which such an indemnity right is recognized occurs in products liability litigation over a defective product, where a seller who has defended a suit against the buyer seeks indemnity from the product's manufacturer for any damages awarded or legal expenditures. Recovery by the seller is based on quasi-contractual principles; the manufacturer is considered the beneficiary of the seller's defense because a successful defense would bar the buyer from re-asserting the same claim against the manufacturer.

There are different approaches across jurisdictions regarding sellers' rights to recover legal expenses. Some jurisdictions allow a seller to recover reasonable legal expenses regardless of the outcome of the litigation with the buyer.[29] These jurisdictions take a distinctly cost-based approach to indemnification. If the seller expends a reasonable sum defending against the buyer's suit, the seller may recover its costs whether the benefit to the manufacturer is "low" (because the buyer prevailed) or "high" (because the seller prevailed).

[29] *E.g.*, Booker v. Sears Roebuck & Co., 785 P.2d 297, 303 (Okla. 1989).

Other jurisdictions allow a seller to recover its legal expenses from the manufacturer, but only—and surprisingly—when the seller *loses* in its defense against the buyer.[30] These jurisdictions follow an inverted *lesser-of* approach to indemnification. If the seller expends a reasonable sum defending against the buyer's suit, the seller may recover its costs only if the seller, and hence the manufacturer, is liable, that is, when the *ex-post* benefit of the defense to the manufacturer is low. If the *ex-post* benefit to the manufacturer is high, that is, if the seller prevails and the manufacturer thereby avoids liability for damages, the seller cannot recover its legal expenses.

This *lesser-of* approach distorts sellers' incentives to defend against product liability suits brought by buyers. The less-than-full indemnity induces sellers to act as less-than-perfect defense agents of manufacturers. Sellers may decline to assert a defense against buyers to avoid jeopardizing their indemnification rights.

3. *Reimbursement of fees under statute*

The third application of the hybrid approach involving attorney fees arises with respect to a plaintiff's right to recover attorney fees from a defendant. The general rule of American law is that each party must bear its own litigation costs. But exceptions to the rule are found in state and federal statutes that establish a right to recover litigation expenses from a defendant in a variety of causes of action (Conte, 1993). From an economic perspective, these statutes are intended to give individuals an added incentive to prosecute violations of the law, by reducing the expected cost of pursuing claims, and to persuade attorneys to represent indigent clients, by enhancing the

[30] This doctrine is based on the notion that the burden of indemnification lies only on a *liable* manufacturer. By succeeding in its defense against the buyer (thereby establishing also the absence of manufacturer's liability), the seller eliminates the basis for indemnification. *See, e.g.*, Merck & Co., Inc. v. Knox Glass, Inc., 328 F.Supp. 374, 376 (E.D. Pa. 1971) (noting that "[i]ndemnity arises where one is legally required to pay an obligation for which another is primarily liable").

prospects of getting paid for their services. Since most lawsuits involve a sure cost but confer only a chance of victory and recovery, these statutes and their incentive effects are well captured by the recovery-for-chance model, even though a plaintiff clearly does not undertake litigation for the benefit of a defendant.

To recover under most fee-shifting statutes, the plaintiff must first have prevailed in the underlying litigation.[31] The recovery is then measured using the "lodestar" approach. To calculate the lodestar, the court simply multiplies the number of hours the plaintiff's attorney worked on the successful portions of the case by a reasonable hourly rate. Importantly, this hourly rate is usually the rate the attorney would charge for non-contingent work. The court may then adjust the lodestar to take into account other factors such as the plaintiff's degree of success in the litigation. However, while the court may adjust the lodestar figure downward to account for poor results obtained in litigation, it may not adjust the figure upward to account for such factors as the risk involved in the litigation.[32]

This statutory approach to recovery resembles a *lesser-of* hybrid regime. The regime takes an element of the benefit-based approach by requiring that a party prevail in order to recover anything at all. However, the regime takes a distinct element of the cost-based approach by measuring recovery on the basis of the reasonable cost of services. Accordingly, where the value of the suit turns out to be high, the prevailing party may recover only the cost of the attorney's services. Where the value of the suit turns out to be low or nominal, however, the party recovers the *ex-post* assessment of the suit's value—nothing or a reduced cost-based figure. The statutory scheme provides less-than-optimal recovery: Whenever the probability of

[31] A party prevails where it recovers monetary damages from its opponent or where it vindicates significant non-monetary interests in the litigation. *See* Hensley v. Eckerhart, 461 U.S. 424, 433 (1983).

[32] *See, e.g.*, City of Burlington v. Dague, 112 S. Ct. 2638 (1992).

prevailing in the suit is less than 1, the party and her attorney will be under-compensated by this regime and may thus under-invest in litigation.

D. Recovery for Precautions

Another setting in which one party might invest to the benefit of another involves accident prevention. A party who takes actions aimed at preventing a harm that might be suffered by another, or for which another party might be liable, often has a claim to recovery, even in the absence of a contract with the other party, on the basis of restitutionary principles. Recovery may be measured by either the benefit conferred, or the reasonable cost of the precautions. If the precautions eliminated an imminent risk, the benefit to the party-at-risk (or the party who is liable for the risk) is readily apparent *ex-post*. Often, however, these precautions only reduce the risk and do not eliminate it, and thus situations arise in which precautions that are cost-justified *ex-ante* provide zero measurable benefit in hindsight. This might be the case if, even after the precautions are taken, the harm—which due to the precautions has become less likely to occur—nevertheless occurs. Or, it might turn out that the harm—which, without the precautions, was more likely to occur—would nevertheless not have materialized. In either case, the *ex-post* benefit from the precautions is zero.

In some situations, where the investing party is a professional performing a service that is within her occupation, the law provides a pure cost-based recovery, equal to the service provider's standard contractual fee. For instance, a doctor who treats an unconscious accident victim may recover her costs, irrespective of the actual benefit to the patient, which could be either higher (if the risk was eliminated) or lower (if the precautions failed). In other situations the law provides a pure benefit-based recovery. For instance, a salvor who comes to the aid of a sinking ship may recover a portion of the value of the salvaged cargo and vessel, but only if the

efforts prove successful; this recovery schedule mirrors the "no cure, no pay" condition commonly found in salvage contracts (Schoenbaum, 2001, §16-5).

Oftentimes, however, a *lesser-of* approach applies. One situation, which was identified by Saul Levmore (1994), involves an insured party who takes precautions to reduce loss for which she is insured. Whenever the precautions go beyond the preventive steps required under the insurance agreement and reduce the likelihood of the insured-against harm, the insured party is conferring a probabilistic benefit upon the insurer. If the precaution is determined *ex-post* to have been successful in fully eliminating the harm, the insured may be able—although this is still controversial—to recover from the insurer the costs of the precaution, even if the insurance contract does not contain a "sue-and-labor" clause requiring the insurer to cover these charges.[33] If, however, the reasonable precaution fails to eliminate the harm which eventually materializes (and which becomes part of the insured's claim), the insured is usually unable to recover the cost of the precaution, as this precaution cannot be proven to have benefited the insurer (Annotation, 1970). Unless there is a provision in the insurance contract covering the insured's prevention expenses, in which case recovery is independent of the success of the prevention effort, a quasi-contractual claim to recover costs would fail in the absence of an *ex-post* benefit. Thus, when the *ex-post* value of the precaution turns out to be high, the insured recovers only its costs—that is, less than the *ex-post* value. Otherwise, when the *ex-post* value of the precaution is zero, the insured recovers nothing. This "half-step" remedy, as Levmore calls it, or *lesser-of* approach as we call it, provides inefficiently low incentives to take precaution.

E. Repairs and Improvements by Co-tenants

Property law also governs several types of investments having probabilistic benefits,

[33] *See* Leebov v. U.S. Fidelity & Guaranty Co., 165 A.2d 82 (Pa. 1960). But see Schlosser Co., Inc. v. Ins. Co. of North America, 600 A.2d, 836 (Md. 1992) for the opposite view.

including repairs and improvements made on co-owned property. While repairs and improvements are not always easily distinguished, the courts tend to treat them quite differently, potentially creating a hybrid regime to govern these investments.

Consider first the rule governing *repairs*. In many jurisdictions, a co-tenant who *repairs* property without the consent of her co-tenants may recover a portion of the cost of those repairs from her non-contributing co-tenants in an action for partition or accounting (Dukeminier and Krier 1998, pp. 358-9). These jurisdictions apply a cost-based recovery approach; if repairs are reasonable, the investing tenant recovers the cost of the repairs (the portion commensurate with the other tenant's stake in the property) regardless of whether the repairs in fact benefit her fellow tenants. For example, a mining company was able to recover one-half of the cost of repairs to a railroad track it jointly owned with another company, even though the passive tenant used the rail far less than the investing tenant and thus derived relatively little benefit from it.[34]

By contrast, the tenant who *improves* property without the consent of her co-tenants may recover the increase in value of the property attributable to those improvements, but not their cost, in an action for partition, or, in some jurisdictions, in an accounting (Dukeminier and Krier 1998, p. 360). This rule resembles a pure benefit-based recovery regime; the investing tenant recovers the full benefit, if any, of the improvements she makes. For example, a co-tenant who invested roughly $1,000 in clearing and draining land to use as pasture and crop acreage was allowed to recover a portion of the enhancement value of such improvements, potentially totaling more than $29,000.[35] The improving co-tenant may not, however, recover her costs where the improvements do not increase the value of the property.

[34] Wagner Coal Co. v. Roth Coal Co., 267 S.W. 1096 (Ky. App. 1925) (noting that the investing tenant shipped five-times more coal on the rail than did the passive tenant).

[35] Buschmeyer v. Eikermann, 378 S.W.2d 468 (Mo. 1964).

One type of distortion arises under the *improvements* doctrine when courts limit the investing party to recovering the *lesser-of* the improvement value or its cost (Stoebuck and Whitman, 2000, p. 208). According to this approach, when the improvement value is low, the investing tenant can recover no more than the value added, which might be less than her cost; and when the improvement value is high, the tenant can recover no more than her cost, which is less than the value added.[36]

A similar *lesser-of* approach is sometimes applied under the *repairs* doctrine as well. Some courts will allow a co-tenant to recover for making repairs only if, in hindsight, the repairs actually increased the value of the land.[37] For example, courts may find a repair to be unreasonable, and hence, not reimbursable, when the repair turns out not to affect the value of the property, even though *ex-ante* the repair seemed like a good idea. As one commentator noted, "the necessity of a repair has been determined in some instances by judging the results of the mending process rather than by the nature of the repairing act" (Note, 1957).

Even more interestingly, the lack of a clear practical distinction between acts that constitute "repairs" and acts that constitute "improvements" may permit an investing tenant to create a *greater-of* regime for expenditures lying on the interface between the two categories of investments. When courts cannot easily distinguish repairs and improvements (or simply refuse to do so) a plaintiff is effectively accorded the power to choose the higher of the two measures of recovery, cost or benefit. If the benefit is low, the tenant would sue to recover her costs under the repairs doctrine; and if the benefit is high, the tenant would sue to recover the benefit under the

[36] Madrid v. Spears, 250 F.2d 51, 54 (10th Cir. 1957).

[37] Clifton v. Clifton, 810 S.W.2d 51, 54 (Ark. Ct. App. 1991) (re-characterizing repairs as improvements and denying recovery on basis of evidence that joint tenant's expenditures had not enhanced value of the property); Womach v. Sandygren, 180 P. 922, 924 (Wash. 1919) (disallowing recovery for repairs where investing tenant failed to show that the repairs enhanced the value of the property).

improvements doctrine.

It is easy to imagine how such a hybrid approach might arise in practice. Courts struggle to classify some investments as either "improvements" or "repairs" across many areas of property law, oftentimes using the two terms interchangeably despite the differing legal treatment accorded each. Indeed, in a variety of cases, courts have (wittingly or unwittingly) allowed the parties to manipulate the distinction between repair and improvement to their advantage.[38] Given the lack of a clear distinction between the two types of investment and the incentive of some parties to muddle them, the hybrid regime may emerge in borderline cases.

F. Mistaken Improvements

A related doctrine concerns recovery for *mistaken* improvement of real property. An investor might improve property she does not own when she unknowingly holds land under an invalid title, mistakes the nature of her interest, or mistakes the location of her land. When the mistake is exposed, an interesting question arises as to whether and how much the improver may recover from the true owner of the land for the improvements.

In most jurisdictions, the mistaken improver who meets certain criteria, such as acting in good faith and under the color of title, may recover from the true owner of the property (Dickinson, 1985). Recovery, however, is capped so as not to exceed the lesser of the *cost* of the investment or its value. For example, under the Restatement of Restitution (§42), the improver may recover "to the extent that the land has been increased in value by [the] improvements, or for the value of the labor and materials employed in making such improvements, *whichever is least*." Furthermore, many states have enacted betterment acts that accomplish the same result by

[38] *Compare* Gilpin v. Brooks, 115 N.E. 421 (Mass. 1917) (holding that, although mortgagee is not allowed to make permanent improvements on the property, he may finish a building if necessary to preserve its value, and the work will be found to be repairs), *with* Warwik v. Harvey, 148 A. 592 (Md. 1930) (holding that a similar completion of a building is an improvement in the context of the mistaken improver doctrine).

allowing the true owner to elect the remedy for the improver.

The mistaken improver doctrine takes a *lesser-of* approach to recovery. *Ex-post*, when the improvement turns out to be valuable, the improver recovers only her costs, but when the improvement turns out to be of little or no value, the improver recovers only that nominal sum. The *lesser-of* approach could potentially distort *ex-ante* incentives. By reducing the recovery from that which the parties would have agreed upon had they contracted (namely, a recovery equal to either the cost of the investment or a portion of the benefit it creates), the Restatement's scheme dilutes incentives to invest and induces excessive caution prior to the unilateral investment in improvements.

III. WHY ARE HYBRID APPROACHES USED IN PRACTICE?

The analysis thus far has demonstrated that hybrid regimes are distortive. Nonetheless, they underlie a wide variety of substantive legal doctrines. This Section explores more systematically why such regimes are used in practice. Does the broad existence of these regimes manifest the confusion of courts in distinguishing between *ex-post* (benefit-based) and *ex-ante* (cost-based) conceptions of value, or can they be justified from either an economic or alternative perspective?

In general, hybrid regimes are created in two ways. Some hybrid regimes are created intentionally by courts or legislatures in order to adjust recovery and thus serve purposes that are often unrelated to investment incentives. Part A explores such purposes and whether the hybrid recovery structure is capable of furthering them. Other hybrid regimes are created inadvertently. Parts B and C describe how problems of drawing boundaries between similar causes of action and problems of information transform what were designed as pure recovery regimes into

inadvertent hybrid regimes. Finally, Part D explores whether the distortion underlying hybrid regimes is eliminated when courts condition any type of recovery on the efficiency of the investment.

A. Deliberate Adjustment of the Recovery

Courts and legislatures occasionally employ hybrid recovery schemes deliberately, to adjust the expected recovery for the investing party and thus serve other instrumental goals unrelated to investment incentives. This Section considers a variety of such goals.

1. *Provide Incentives to Contract or to Avoid Unsolicited Investment*

One goal a downward adjustment might serve is to give investors incentives to contract with beneficiaries or to avoid unsolicited investment. For example, a *lesser-of* regime, by reducing an investor's expected recovery, may give her incentives to verify title to her land before making an improvement.

While providing such incentives may be desirable policy, use of the *lesser-of* hybrid regime is a misguided way to implement it, in several respects. First, if the policy is intended to induce the investing party to contract with the beneficiary rather than make a unilateral investment, or to take more care before making a mistaken improvement, better incentives might arise if *no recovery* were allowed. Indeed, the doctrines that create restitution liability already incorporate a fault standard: the right to recover is itself conditional on the investor either taking sufficient care or not having reasonable opportunities to contract. For example, a mistaken improver must show that she acted in good faith and under color of title before making her improvements. Such conditions for the *incidence* of liability provide investors with adequate incentives to take care and to contract, rendering unnecessary additional tinkering with the *magnitude* of liability. Thus, in those cases where the investor has satisfied the "due care"

requirements like the ones embodied in the mistaken improver doctrine, it is unclear what instrumental purpose, if any, a reduction of the damages award serves.

2. *Protect "Innocent" Parties*

Another stated purpose for the *lesser-of* approach is to protect "innocent" parties from burdensome liability. For example, the Restatement of Restitution asserts that forcing an "innocent" owner to pay the full value of improvements mistakenly placed on her land is harsh. An innocent beneficiary of an illiquid benefit should not be forced to liquidate her property to be able to pay for the improvements, proponents of the *lesser-of* approach assert.[39] Thus, the owner should not have to pay for the full enhancement value. Further, even if the owner could afford to pay, it seems unfair, from an *ex-post* perspective, to require an owner who received no enhancement value to compensate a mistaken investor for the costs of the failed improvement effort. The *lesser-of* regime seems to perfectly serve this dual protective goal.

Upon careful examination, however, the *lesser-of* rule of the Restatement is more difficult to justify. In the absence of any apparent wrongdoing, it is unclear why fairness necessarily favors one innocent party over another, that is, why the owner, but not the equally "innocent" and potentially cash-strapped mistaken improver, should be protected. More fundamentally, even if some sort of reduction in liability is desirable, to protect the "autonomy" of the owner who did not solicit the improvement, the reduction achieved through the *lesser-of* regime is still ill-suited to this purpose. As the economic analysis in Section I demonstrated, the *lesser-of* regime achieves a reduction in liability equal to the option value embodied in election between the cost-based and the benefit-based values. The magnitude of this reduction depends primarily on the *variance*, or the riskiness of the investment, a factor that is independent of the

[39] REST. RESTITUTION § 42 cmt a (acknowledging that, while the rule is "harsh to the one making the improvements by mistake . . . in many cases it would be still more harsh to require the one receiving the benefits to pay therefor").

reasons this reduction was deemed desirable in the first place. Thus, when the investment yields a certain—instead of a probabilistic—benefit, there is no reduction in liability although the same hardships confront the innocent landowner. Recognizing the probabilistic nature of the benefit demonstrates, therefore, that the reduction in liability attained by the *lesser-of* rule is arbitrary: it is not tailored to serve the goal motivating the reduction.

Another area in which the *lesser-of* reduction of liability is intended to ease the compensatory burden placed on the beneficiary involves recovery for breach of a contingency fee agreement. Here, courts applying the *lesser-of* approach intend to give the client greater freedom to dissolve the relationship with his current attorney and enter into a better match with a different attorney. The pure cost-based regime, according to some courts, is unfair to poor clients who cannot afford to pay the attorney's fees unless the client recovers in the suit. At the same time, the pure benefit-based regime, according to these courts, would place an undue burden on the client's right to dismiss her attorney, because the client might end up having to pay two attorneys a full contingency fee.[40] It is less often recognized, however, that while the reduced recovery accords greater freedom to the client, the added risk it places on the attorney might diminish the attorney's willingness to take on the client's case on a contingency basis (recall that billable-hours contracts are governed by different rules), or cause the attorney to raise her rates to absorb the risk, both to the client's detriment.

3. *Deter wrongdoing*

Finally, a hybrid approach may also be deliberately tailored to serve deterrent concerns. For example, fiduciary and agency doctrines entitle a principal to a *greater-of* recovery against a fiduciary or an agent who violates her duties to the principal. If the agent receives a large benefit

[40] *E.g.*, *Fracasse*, 494 P.2d at 12-14.

by violating her duty of loyalty (say, if the agent expropriates funds and invests them in her own account successfully), the principal is entitled to recover the entire *ex-post* benefit. And if the agent receives little or no benefit from the violation (say, if the agent's investment failed), the principal can alternatively recover damages equal to the value taken from the principal's account (Restatement 2d of Agency §407).

This *greater-of* rule can be rationalized on the basis of deterrence theory. Since many violations of fiduciary and agency relationships go undetected, the risk of over-recovery, which normally arises under the *greater-of* regime, is not much of a factor. By applying the greater of the *ex-post* and the *ex-ante* recovery values, an increase in deterrence is achieved, countering some of the effect of imperfect detection. That is, while the expected recovery under the hybrid rule exceeds the expected value of the funds that were taken, this premium hardly measures up to the "discount" enjoyed by the wrongdoer who goes undetected.

B. Overlap of Pure Regimes

The hybrid approaches are not always adopted deliberately. A hybrid regime might also arise accidentally, where two different "pure" recovery regimes overlap. When a particular investment can lead to recovery under two different causes of action, one employing a pure cost-based recovery approach and the other employing a pure benefit-based recovery approach, a hybrid regime might *de facto* govern this investment, for two reasons. First, courts may openly defer to the investing party to elect which of the two causes of action to apply to her investment; not surprisingly, she will elect the one that gives her the greater measure of recovery. This was shown to be the case, for example, in the choice of remedies regime governing total breach of a partially performed contract.

More interestingly, courts may not be able to prevent a party from opportunistically

pursuing one cause of action over another, as when the boundaries between two related causes of action are imprecise. For example, the co-tenant repairs and improvements doctrines overlap in some cases where an investment can be categorized as both a repair and an improvement. The use of a cost-based approach in recovery for repairs and a benefit-based approach in recovery for improvements may become a hybrid *greater-of* approach if the investing party can elect which of the two doctrines to apply. To the extent that courts cannot draw a bright line between what constitutes a repair versus an improvement, the investing party can effectively elect the greater of the two pure recovery measures. Likewise, a problem of imprecise boundaries exists within the implied contracts doctrine. To the extent that courts cannot draw a bright line between the grounds for implied-in-fact and implied-in-law claims—and, at least in the context of precontractual investment, such a line is difficult to draw—the investing party can claim the greater recovery measure.

Note that in the repair/improvement case, and to some extent in the implied contracts case as well, courts do not openly permit the investing party to characterize her investment so as to secure the higher recovery measure. In fact, if courts were aware of the problem, they might be driven to draw more precise boundaries between existing amorphous causes of action. Unfortunately, the type of sorting of claims that creates these *greater-of* regimes occurs "pre-trial", distant from the judge's scrutiny, when potential plaintiffs privately design their pleading strategies. In the usual case, a plaintiff pleads only one cause of action, either a pure *ex-post* or a pure *ex-ante* recovery claim. It is only across cases that a *greater-of* pattern emerges.

C. Information Problems

Finally, and perhaps most interestingly, hybrid approaches might also emerge inadvertently when courts lack the information necessary to apply a "pure" regime consistently

across a class of cases. In order to apply a pure benefit-based regime, courts must be able to verify the actual benefit received. In order to apply a pure cost-based regime, courts do not need to know the actual benefit, but they do need to know the cost of the investment and the *ex-ante* distribution of benefits associated with the investment (to guarantee that the cost is reimbursed only if it was reasonable). The discussion below demonstrates that when some of this information is not readily verifiable in court, pure recovery regimes might be transformed into hybrid regimes. Formally, the doctrine employs a pure approach to measuring recovery; in practice, given information problems, it operates like a hybrid.

1. *Information Regarding the Distribution of Benefits*

To apply a cost-based recovery regime, courts need to calculate the *ex-ante* distribution of benefits. In contrast to the benefit-based regime, in which courts need only measure the actual realization of the benefit, under the cost-based regime courts need to assess whether the investment was reasonable in light of its projected benefits. In order to do that, courts have to consider the range of possible benefits that were associated with the investment and their associated likelihoods. That is, courts need to be able to measure not only the actual benefit that materialized, but also hypothetical (or counter-factual) ones. When the difficulty in estimating the prior distribution of benefits is accounted for, a pure cost-based recovery regime can be transformed inadvertently into a *lesser-of* regime.

One of the factors that could—and we believe, in fact does—interfere with a court's ability to accurately assess the distribution of benefits at the time when the costly action was taken is the hindsight bias. When a court knows the *ex-post* value of an investment, but does not have enough information to determine its *ex-ante* expected value, it may draw an inference about expected value from the value that was realized. If the actual benefit from the investment turns

out to be high, it is likely to appear cost-justified, and recovery of the cost would be allowed. If, instead, the actual benefit from the investment turns out to be low or zero, the investment as a whole might seem unreasonable, and recovery of the cost would be denied. Under these conditions, a cost-based recovery regime, in which the investing party recovers her costs only if they are reasonable, may turn into a *lesser-of* regime.[41]

This hindsight bias can explain the emergence of the *lesser-of* regime in several of the areas surveyed in Section II. For example, it can explain the *lesser-of* rule that sometimes governs restitution for repairs made by co-tenants and the quasi-contractual recovery for precautions. Some courts, when evaluating the reasonableness of certain repairs, condition the right to recover repair costs on whether the repairs appear, in hindsight, to be justified. Thus, the mere fact that the repairs did not add value *ex-post* is used to justify a conclusion that they were not reasonable *ex-ante* and thereby to deny recovery of their costs.[42] Courts fail to see, in this context, that even repairs that "failed" to generate value could have been reasonable when made. Similarly, in assessing the desirability of precautions taken by an insured, courts already know whether the precautions succeeded in preventing or reducing the loss, and are susceptible to a well-documented hindsight bias (Rachlinsky, 1998). One of the effects of this bias is that when a precaution fails to reduce the loss, courts draw an inference that it was not cost-justified in the first place and refuse to award even the *ex-ante* measure of recovery. Another potential effect of this bias might occur when, in hindsight, it is clear that the loss was avoided independently of the precaution, which again might lead courts to wrongly conclude the that the precaution was

[41] Another possibility is that a court might mistakenly award recovery for an investment that was "unreasonable" *ex-ante*, because, by chance, it proved valuable *ex-post*.

[42] *See, e.g.,* Clifton v. Clifton, 810 S.W.2d 51, 54 (Ark. 1991) (no recovery of cost of repair when it added no value to the property).

unjustified *ex-ante* and deny the recovery of its cost.

This "hindsight" problem is also illustrated in the debate over the plaintiff attorneys' fees stemming from the tobacco settlement. Either an hourly fee that is not contingent or a contingency fee that is not truncated could adequately compensate the attorneys representing the states. However, conditioning the recovery on success and then limiting it to the (guaranteed) hourly fee creates a *lesser-of* regime. The rhetoric utilized by advocates of this regime suggests that they fail to consider the substantial *ex-ante* likelihood that the contingency fee attorneys could have received no recovery at all.

In theory, courts can avoid or mitigate the problems created by gaps in information by adopting the "pure" regime for which the best information is available. If it is consistently difficult for courts to assess the *ex-ante* value of any given type of investment, the courts could instead apply an *ex-post* approach to govern those investments, assuming, of course, that information about the actual benefit is relatively more obtainable.

2. *Verifiability of the Actual Benefit*

A different type of information problem might arise if courts cannot easily verify the actual benefit. Because it is the defendant/beneficiary who usually possesses the best information regarding the benefit enjoyed, the defendant might manipulate the type of information he reveals to the court. Recognizing the plaintiff's difficulty in proving the magnitude of the benefit she conferred upon the defendant, courts might allow a plaintiff who cannot prove the magnitude of the benefit to at least recover her costs, whenever these costs appear reasonable. Namely, in asymmetric information environments, courts might be willing to award recovery based on the full *ex-post* benefit enjoyed by the defendant whenever reliable information about the actual benefit is provided, but award only some *ex-ante* measure (either cost or expected benefit)

otherwise. This adjudication regime quickly transforms into a *lesser-of* regime if the defendant can selectively disclose information. The defendant would hide information about his actual benefit whenever this benefit is high, thereby limiting the plaintiff to the more moderate cost-based measure of recovery. And conversely, the defendant would reveal information about his actual benefit whenever this benefit is low, to limit the magnitude of the plaintiff's recovery to the (low) actual benefit. To the extent that discovery procedures enable the defendant to manipulate information in this way, the plaintiff is effectively governed by a *lesser-of* regime.

The limited ability of courts to verify actual benefits might also translate into a *greater-of* regime, when it is the plaintiff who can manipulate the information provided to the court. One way the plaintiff can control the informational-basis of the recovery is by affecting the *timing* of the suit. A plaintiff who, on the basis of private information, knows that the *ex-post* benefit will be low, can time her suit prior to the verifiable realization of the benefit, expecting the court, which cannot verify the actual benefit, to employ instead a cost-based recovery measure. Conversely, a plaintiff who knows that the *ex-post* benefit will be high can await the verifiable realization of the benefit and recover the full *ex-post* value.

3. *Information about the Cost of the Investment*

Another type of information problem arises when courts cannot verify the cost of the investment. Because of the stochastic nature of the benefit, any given observable *ex-post* realization of benefit can be associated with any number of different costs of investment. This problem might be particularly acute in the context of pre-accident precautions taken by an insured party. As many of the precaution measures that the insured can take are both non-verifiable in court and non-observable to the insurer, it is less puzzling why the parties to the insurance arrangement do not contract over them and why the courts cannot apply the pure cost-

based recovery rule to them. Courts must look to the verifiable benefit to ascertain, not only whether the precaution was desirable, but also whether the alleged precaution was ever taken. Thus, when the realization of the benefit is high, the inference that some unobservable precaution had been taken is more plausible than when the benefit is low. When a ship sinks, courts are less likely to believe that the insured ship-owner took the necessary, yet subsequently futile, precautions. A cost-based recovery regime might, in the presence of this Bayesian inference strategy, transform into a *lesser-of* regime.

However, the *lesser-of* regime applied in practice to precautions taken by insured parties cannot be fully explained as a by-product of this information problem. If the non-verifiability of the precaution investments were the reason for the transformation of a cost-based rule into a *lesser-of* rule, one would expect that in cases where precautions are observable and verifiable a pure cost-based regime would survive. This, however, is not the case. While many *pre-accident* precautions are indeed non-verifiable, most *post-accident* harm-reducing mitigation actions taken by the insured—which are another category of precautions—are more easily observable and verifiable, and yet are subject to the same *lesser-of* rule. For example, the actions taken by a ship's crew to avoid maritime hazards prior to an accident (e.g., safer routes, maintenance of machinery) might be non-verifiable, whereas actions taken by the crew to reduce the harm after an accident (e.g., raise a sinking vessel) are more readily verifiable. Case law, however, can hardly be partitioned according to this verifiability property of precautions; the adherence to the *lesser-of* rule and the denial of recovery for failed precautions are more robust than this conjecture would imply.

D. The Potential "Incentive-Irrelevance" of Hybrid Regimes

In our discussion of legal doctrines, we often took for granted that a beneficiary is liable

for a given investment, and focused instead on how the courts measure damages. As we have demonstrated, courts employing the hybrid regimes set damages too high or too low, and thereby skew incentives to make the investment in the first instance. But, one might speculate, courts could avoid this problem by deciding, first, whether the investor's behavior was efficient, and then, based on that determination, apply one or the other hybrid regime. Thus, if the court were to find the investment was not cost-justified, it could limit recovery by employing the lesser-of approach. If the court were to find that the investment was efficient, it could apply the greater-of approach instead. Used this way, the hybrid regimes would not distort investment incentives. There would be no incentive to make socially undesirable investments, and no lack of incentive to make socially desirable ones.

There are, however, several problems with this conjecture. To begin, it does not describe what actually occurs in practice. Courts rarely follow this two-step approach of checking whether an investment was efficient before choosing which of the two hybrid regimes to apply. Investors may have to satisfy certain requirements to establish liability (such as proving a mistake was "reasonable"), but demonstrating the efficiency of an investment is not one of them. For example, in contract damages, there is no such hurdle the plaintiff would need to clear. In fact, in many cases in which breached-against parties recovered the greater restitution damages, expectation damages would have left them with a loss, namely, these are losing contracts that involve seemingly inefficient investments.

Moreover, to correctly encourage efficient investment and discourage inefficient ones, courts must decide which hybrid regime to apply on a case by case basis. In cases that are found to involve efficient investments, courts must apply a *greater-of* approach, whereas in cases that are found to involve inefficient investments, courts must apply a *lesser-of* approach. Courts,

however, do not partition the use of the regimes in such a case by case basis, but instead apply a unified rule across an entire category of investment. Accordingly, it would be impossible for courts to tailor the recovery regime according to the efficiency of the investment. To illustrate this argument, consider again the rule governing mistaken improvements. The same *lesser-of* rule is applied to all mistaken improvements of property, regardless of their efficiency attributes. Suppose a court following the two-step approach determined, first, that an investment was efficient and that the improver should be compensated. Under the two-step approach, the true owner would be held liable, even when no benefit actually materializes from the investment. The problem is, the mistaken improver's recovery in that case would be 0, inadequately compensating the improver for his costs. And when a benefit does materialize, the improver's recovery would be limited to his cost. On average, the mistaken improver would expect to recover less than the benefit generated or the cost incurred, even for efficient, desirable investments. Under the existing doctrine, the court cannot escape the *lesser-of* result, even in situations in which it is clear that the investment merits more complete compensation. Thus, courts that apply a lesser-of regime to all investments of a particular type (here, mistaken improvements), cannot prevent the hybrid regime's distortion by using the two-step approach.

Further, the hindsight bias we discussed in Part C above might make it difficult for courts to execute a two-step approach accurately. The hindsight bias might skew courts' determinations of whether an investment was, from an *ex-ante* point of view, efficient. A court may be more likely to determine (erroneously) that some cost expenditure was unjustified, if the expenditure did not actually generate a benefit. As we discussed, this bias appears to explain why courts have adopted hybrid regimes in a number of legal settings discussed in Part II. By causing courts to declare some cost-justified investments to be "inefficient" at the first step, the bias would tend to

discourage efficient investments.

Finally, the use of the hybrid regimes adds an element of arbitrariness to the compensation award. Even if a two-step approach were to be executed properly, such that the hybrid recovery regimes would not distort the investment decision, the measure of recovery is hard to rationalize as a good compensatory scheme. Raising or lowering the recovery in a manner that depends only on the variance of the distribution of benefits can hardly be said to accomplish the stated purposes of the hybrid regimes, such as encouraging parties to take due care. It is difficult to see why the law should permit an investor who created a stochastic benefit of either 0 or 200 (equally likely) to collect a larger recovery than an investor who created a less volatile benefit, or even a certain benefit of 100. Recovery simply should not depend on such irrelevant factors.

CONCLUSION

This article has identified a distortion in the structure of legal rules that deal with chance. Although the type of uncertainty examined here—uncertainty over the value of the investment—is (usually) resolved by the time the law steps in to determine the recovery, the confusion between the *ex-ante* (cost-based) and the *ex-post* (benefit-based) measures of value leads to hybrid recovery practices with their associated distortions. By identifying the generality of the problem—potentially arising any time the net external benefit of an investment is probabilistic—the analysis can be applied to any situation that exhibits this structure. The article explored some half-dozen applications of the hybrid approaches, all from seemingly unrelated areas of law, but all sharing the same analytical structure. The list is, of course, far from exhaustive. Accordingly, the usefulness of the analysis would prove greater to the extent that the trans-substantive tool offered here is found applicable within other areas of the law as well.

One possible application of the analysis is extending it to the case of probabilistic costs, rather than benefits. Some investments, say, in improving property have fairly certain benefits but involve random costs (e.g., excavating land). Here, too, in the absence of a contract the investing party may be subject to a hybrid recovery regime. For example, the law of special assessment, which taxes property owners for improvements made by the city near their property, allows the city to recover at most its costs, not to exceed the enhancement value of the affected properties.[43] According to the analysis in this paper, this *lesser-of* regime diminishes the city's incentives to make socially valuable improvements.

The analysis in this paper can be read "narrowly", as a remark on the benefit principle within the law of quasi-contract. Under this principle, the liable party has to pay only when an actual benefit is conferred upon him. The analysis in this paper provides an argument for expanding the definition of benefit to include, not only actual benefits, but also potential yet unrealized benefits. Receiving a *chance* for enrichment is valuable to the beneficiary in similar fashion that receiving a lottery ticket is beneficial. Recovery, though, has to be consistent across realizations. If the beneficiary pays only for the *ex-ante* value of the chance when the chance materializes *ex-post* into a substantial benefit, he should also pay for the value of the chance when a benefit does *not* materialize *ex-post.*

Lastly, this paper can be read more broadly, as a comment on the appropriate interface between cost-based and benefit-based liability in private law. Costly actions that are identical from an *ex-ante,* cost-based perspective, can appear dissimilar *ex-post*, once the stochastic benefit from them materializes. This appearance can lead—and as we showed, it has often led—courts to apply an inconsistent treatment of the right to recovery, bouncing in an arbitrary fashion

[43] *E.g.*, McNally v. Teaneck, 379 A.2d 446 (N.J. 1977).

between cost-based and benefit-based liability. While the paper does not take a position concerning the choice between the two pure methods of measuring liability, it highlights the distortion that an inconsistent choice creates.

References

Annotation. 1957 "Measure or Basis of Attorney's Recovery on Express Contract Fixing Noncontingent Fees, Where He is Discharged Without Cause or Fault on His Part," 54 A.L.R.2d 604.

_____. 1970 "Recoverability, Under Property Insurance or Insurance Against Liability for Property Damage, of Insured's Expenses to Prevent or Mitigate Damages," 33 A.L.R. 3d 1262.

_____. 1998. "Limitation to Quantum Meruit Recovery, Where Attorney Employed Under Contingent-Fee Contract is Discharged Without Cause," 56 A.L.R.5th 1.

Atiyah, Patrick S. 1979. *The Rise and Fall of Freedom of Contract* (Oxford University Press)

Brinkman, Lester. 1998. "Will Legal Ethics Go Up in Smoke?," WALL ST. J. A18 (June 16, 1998).

Conte, Alba. 1993. *Attorney Fee Awards* (2nd Ed., Shepards Pub.)

Corbin, Arthur L. 1964. *Corbin on Contracts Vol. 12* (Interim Ed., New Jersy: LexisNexis Pub.)

Dickinson, Kelvin H. 1985. "Mistaken Improvers of Real Estate," 64 N.C. L. Rev. 37.

Dukeminier, Jesse, and James Krier, *Property* (4th Ed., New York: Aspen).

Farnsworth, E. Alan. 1987. "Precontractual Liability and Preliminary Agreements: Fair Dealings and Failed Negotiations," 87 COLUM. L. REV. 217.

_____. 1999. *Contracts* (3rd Ed.,)

Hazard, Geoffrey C. and W. William Hodes. 2000. *The Law of Lawyering* (3rd Ed., New York: Aspen Pub.)

Jackson, Thomas. 1978. "Anticipatory Repudiation and the Temporal Element of Contract Law: An Economic Inquiry into Contract Damages in Cases of Prospective Nonperformance,"

31 STAN. L. REV. 69.

Kull, Andrew. 1994. "Restitution as a Remedy for Breach of Contract," 67 S. CAL L. REV. 1465.

Landes, William, and Richard A. Posner. 1978. "Salvors, Good Samaritans, and Other Rescuers: An Economic Study of Law and Altruism," 7 J. LEG. STUD. 83.

Levmore, Saul. 1985. "Explaining Restitution," 71 VA. L. REV. 65.

_____. 1994. "Obligation or Restitution for Best Efforts," 67 S. CAL. L. REV. 1411.

Note. 1957. "Right of Cotenant to Contribution from other Cotenants for Unauthorized Repairs and Improvements Made to the Common Property," 32 NOTRE DAME L. R. 493.

Polinsky, A. Mitchell, and Steven Shavell. 1994. "Should Liability Be Based on the Harm to the Victim or the Gain to the Injurer?," 10 J. L. ECON. & ORG. 427 (1994).

_____. 1998. "Punitive Damages: An Economic Analysis," 111 HARV. L. REV. 869.

Posner, Richard A. 1998. *Economic Analysis of Law* (5th Ed., New York: Aspen Pub.)

Rachlinsky, Jeoffrey J. 1998. "A Positive Psychological Theory of Judging in Hindsight," 65 U. CHI. L. REV. 571.

Schoenbaum, Thomas J. 2001. *Admiralty and Maritime Law* (3rd Ed., West).

Stoebuck, William B. and Dale A. Whitman. 2000. *The Law of Property* (3rd ed.).

Williston, Samuel. 1960. *A Treatise on the Law of Contracts* (3rd Ed.).

Wittman, Donald. 1995. "Should Compensation be Based on Costs or Benefits?," 5 Int'l Rev. L & Econ. 173.