Penalty Clauses and Liquidated Damages

Steven Walt*
Penalty Clauses and Liquidated Damages

Steven Walt

Abstract

This Chapter critically surveys the recent economic literature on penalty clauses. Almost all of this work appeared between 1977 and 1996, with a few contributions appearing in the late 1990s. It differs from early law and economics scholarship on penalty regulation in both method and substance. The early work evaluated penalty regulation informally. In contrast, much of the strictly economic recent work formally models the effects of damages clauses on investment in performance, breach and trade under prescribed conditions. The early and more recent work also differ in substance, in their evaluation of penalty clauses. Earlier work generally found that efficient damages stipulations are not overcompensatory. It concluded that penalty regulation is superfluous. In contrast, more recent scholarship finds that optimal contracts can contain penalty clauses. This Chapter argues that the conditions under which optimal damages clauses are overcompensatory do not support anything resembling existing penalty doctrine.
Penalty Clauses and Liquidated Damages

Steven Walt
Percy Brown Jr. Professor, University of Virginia School of Law

1. Introduction

The common law of contracts refuses to enforce contractual stipulations of damages courts deem penalties. Although sometimes formulated differently, doctrine characterizes a stipulation a "penalty" that either unreasonably forecasts expected or actual damages arising from breach, or sets damages that are easily ascertainable by a court. Damage stipulations that either reasonably forecast expected or actual damages, or provide for damages that are difficult to ascertain judicially, are deemed "liquidated damages" and enforced. For most legal economists and many traditional legal scholars, the penalty doctrine is puzzling. Contracting parties agree to stipulate damages from breach, as they do to any other contract term, because they anticipate that the stipulation maximizes their joint surplus. A "performance terms" doctrine specifically regulating the performance terms of a contract, such as risk of loss or warranty provisions, therefore is undesirable, and contract law does not contain one. For the same reason, a penalty doctrine specifically regulating damage stipulations also seems undesirable.

This Chapter critically surveys the recent economic literature on penalty clauses. Almost all of this work appeared between 1977 and 1996, with a few contributions appearing in the late 1990s. It differs from early law and economics scholarship on penalty regulation in both method and substance. The early work evaluated penalty regulation informally. In contrast, much of the strictly economic recent work formally models the effects of damages clauses on investment in performance, breach and trade under prescribed conditions. Some of the law and economics literature relies on psychological findings about decision making to explain or justify the penalty doctrine. Unlike earlier work, researchers are more systematic in their analysis and more careful about the normative implications of their results. While most (but not all) recent work does not support the penalty doctrine, the research bases its conclusion on the doctrine's impact on specifically identified variables isolated for study. Penalty clauses may have different efficiency effects on variables not studied. The early and more recent work also differ in substance, in their evaluation of penalty clauses. Earlier work generally found that efficient damages stipulations are not overcompensatory (Goetz & Scott, 1977; Schwartz, 1990). In contrast, more recent scholarship finds that optimal contracts can contain penalty clauses (Edlin, 1996, 1998).

The more systematic recent work is also nuanced in its normative implications. It finds that penalty clauses can produce inefficient performance or efficient breach under some conditions, and efficient or inefficient investment or trade also under other conditions. Penalty clauses can have anticompetitive effects on the size of product markets in which they are used. They also can produce inefficient breach at the same time they induce efficient investment. Thus, whether penalty clauses maximize social welfare depends on the robustness of the specified parameters as well as an assessment of the comparative impact of the clauses on the variables identified. This conclusion resembles the conclusion usually drawn from work on the efficiency of traditional damages measures: none of these measures are unambiguously efficient because
each measure impacts different variables affecting contract performance differently (Posner, 2003; Craswell, 1988). In critically surveying the recent literature, this Chapter argues that it does not support anything remotely resembling existing penalty regulation. Edlin and Schwartz (2003) survey much of this work, for different purposes.

2. The Legal Regulation of Penalty Clauses

Damage provisions a court finds unenforceable are designated “penalties;” those it finds enforceable are designated “liquidated damages.” The traditional common law rule voids stipulations of damages that bear no reasonable relation to the damages expected to be suffered from breach (Ibbeston, 1999). Courts also traditionally invalidated damage provisions where damages anticipated from breach were not difficult to prove (Farnsworth, 2004). Both applicable statute and the trend in case law are less restrictive. Under section 2-718(1) of the Uniform Commercial Code, damages may be stipulated in an amount reasonable in light of anticipated or actual harm caused by the breach (Uniform Commercial Code, 2008). Section 356 of the Restatement (Second) of Contracts describes the same rule extracted from case law (Restatement, 1981). Thus, damages stipulations that are ex ante unreasonable but reasonable in relation to actual damages are enforceable. For the same reason, stipulations are enforceable even when anticipated damages are easily provable, as long as they are reasonable in relation to expected or actual damages. Although provisions deemed penalties are unenforceable, the remainder of the agreement is enforceable, and resort to default remedies for breach remains available. By contrast, civil law systems generally enforce damage provisions, whatever their amount. The contract law of some of these systems allows courts to reduce or increase the amount provided if it is manifestly excessive or inadequate (Hatzis, 2003; Mattei, 1995; Council of Europe, 1990).

Stipulated damages are either exclusive or optional remedies for the contracting parties. They are exclusive if the contract expressly or implicitly so provides; otherwise, the breach victim has recourse to default remedies, as under section 2-719(2) of the Uniform Commercial Code (Uniform Commercial Code, 2008). Most courts restrict the remedies available when a stipulate damage provision is optional. They make available the right to specific performance, where appropriate, but not damages. This restriction is puzzling. To vet a damage provision, a court must determine actual damages or the reasonableness of the damages stipulated in relation to actual damages. In both cases the court therefore needs to gauge the extent of the victim's loss from breach. Preventing the victim from measuring its loss by damages is odd when the court has measured loss, particularly when the loss may be less than the amount stipulated in the damage provision.

The rules regulating damage provisions are limited in two respects. First, they forbid only damage stipulations that are excessive in relation to expected or actual damages. Courts frequently enforce provisions that stipulate damages in amounts less than the expected or actual damages from breach (Scott & Triantis, 2004). Thus, penalty regulation does not apply to “underliquidated” damages. Contract law allows enforcement of contract provisions that exclude recovery of consequential damages resulting from breach. These damage limitation clauses effectively underliquidate damages. At most the regulation therefore protects against overcompensation of the breach victim, not her undercompensation. Second, penalty regulation does not apply to some contract clauses that function as damage provisions. An important
example is “take or pay” clauses. A “take or pay” clause requires the buyer to accept the entire quantity contracted for at the contract price or pay for a stipulated minimum quantity at the unit price. Such clauses are useful to sellers because they avoid the need for sellers to prove market damages or the size of their variable unit costs. Courts sometimes enforce “take or pay” clauses by characterizing them as part of the buyer's performance obligations, not a stipulated remedy for breach (Gillette & Walt, 2008). The clauses nonetheless function as stipulated remedies, apparently immune from penalty regulation. A full justification of existing penalty regulation must account for these two limitations. None of the accounts in the literature justify both of them.

3. Arguments Against Penalty Regulation

Damage stipulations are costly to write. If the promisee's valuation of performance were observable to both parties and verifiable to a court, the promisor's breach decision unaffected by contract provisions and the parties' investment in performance of the contract impossible, damage provisions would not be written. Because a court could accurately determine the promisee's loss from the promisor's breach, such provisions would not be needed to fix damages accurately. Promisors could offer the particular contract each promisee demanded based on the different value the promisee placed on the promisor's performance without the need for damages provisions to convey this information to the promisor. Damage provisions could not affect the efficiency of the promisor's decision to breach or perform, by assumption. Nor could these provisions induce investment in contract performance and therefore the size of the parties' joint surplus, because such investment is impossible. None of these assumptions are realistic. Early research studied the effect of a damage stipulation on the breach decision when valuations are verifiable, the parties symmetrically informed and specific investment in the contract infeasible (Clarkson et al. 1984; Muris 1982; Posner 1977). Under these conditions penalty clauses induce inefficient breach decisions. More recent formal and informal models relax one or more of these strong assumptions. They show that, under specified conditions, either damage provisions are not penalties or that penalty provisions can induce efficient performance. The former result makes penalty regulation superfluous; both results make it undesirable.

3.1 Unverifiable Valuations

As with any contract term, contracting parties have an incentive to select an efficient measure of damages from breach. This is because an efficient damage measure maximized the joint surplus. If courts accurately measured damages from breach, the parties would not stipulate damages. The stipulation would yield no benefits and is costly to write. The law governing damage stipulations assumes that courts accurately measure damages. However, the assumption is unrealistic. Parties sometimes attach idiosyncratic values to performance. They sometimes also prefer not to disclose private information to a court that otherwise could establish the value they attach to performance (Ben-Shahar & Bernstein, 2000). Even identifying the relevant markets by which to measure market price can be difficult. For these reasons, courts make errors in determining the loss from breach. Judicial errors in measurement of loss increase the contract price. This is because contract price reflects damages payable in the event of breach, and judicial error increases the cost of breach for the breaching party. If the parties on average can accurately estimate damages, they can reduce the contract price by stipulating damages. A damage stipulation reduces contract price when the negotiation costs of the stipulation are less than the
sum of proof and error costs from reliance on a court to measure damages.

Goetz and Scott (1977) present an early informal model incorporating judicial error in measuring damages. The value of performance is observable to the parties but unverifiable to a court. Renegotiation is infeasible and a damage stipulation is implicitly assumed to have no effect on the choice of contracting partner or investment in performance of the contract. The parties decide to breach or perform the contract. With these assumptions, Goetz and Scott show that only stipulated damages induce an efficient breach decision. The promisor will perform when the cost of its performance is less than the promisee's loss in value from breach; it will breach when its performance cost is greater than the promisee's loss in value from breach. Courts make errors in determining loss from breach because the promisee's value is unverifiable to it. Thus, without a damage stipulation, the promisor will either inefficiently perform or inefficiently breach. Because the value of performance is observable to the parties, on average their estimation of loss from breach is accurate. A damage stipulation therefore induces the promisor to make an efficient breach decision: to perform when the cost of doing so is less than the promisee's loss in value from breach and to breach when the converse holds.

Stipulated damages are not penalties in Goetz and Scott's model: they measure the actual loss to the promisee from breach. Thus, penalty regulation is superfluous. It also is undesirable because the judicial unverifiability of valuations makes courts likely to mistakenly find damage provisions to be penalties. Penalty regulation is undesirable even if optimal damage provisions sometimes are penalties, although Goetz and Scott do not draw this conclusion. If courts cannot verify valuations, they cannot accurately measure loss from breach. As a result, they cannot reliably determine when a damage provision is a penalty and when it sets compensatory damages. Courts therefore cannot regulate damage provisions effectively even if they sometimes are overcompensatory. Judicial error in valuation is enough to condemn penalty regulation.

Other elements of Goetz and Scott's presentation are irrelevant to the model of efficient breach with unverifiable information about loss. In particular, the model is introduced as part of a model of an “efficient insurer.” The two models in fact are independent, and the model of efficient breach has more general application. Damages stipulations serve as insurance when a risk-averse promisee and risk-neutral or risk-preferring promisor agree to shift the promisee's loss from breach to the promisor. The premium paid by the promisee includes loss that is covered by the damage stipulation but difficult to prove ex post. Critics complain that promisees usually will not insure against nonpecuniary loss (Croley & Hanson, 1995; Talley, 1994; Rae, 1984, 1982). Even if the criticism is correct, promisees will not buy insurance in excess of their expected loss. Stipulated damages instead will underliquidate the promisee's full damages, not serve as penalties. More important, the criticism goes to damage stipulations as insurance, not to Goetz and Scott's efficient breach model (Walt, 2003). The model depends only on error in judicial measurements of loss from breach, not on the parties’ attitudes toward risk. Parties stipulate damages to induce efficient breach decisions without regard to their risk preferences. The stipulation allows them to reduce the price of their contract.

3.3 Unverifiable Performance

Breach can be observable to the parties but difficult for a court to detect. Penalties can
induce efficient breach when courts cannot observe whether the parties have performed the contract. A court's determination of performance is based on evidence supplied by the litigating parties, and evidence submitted is fabricated or used for self-interested purposes (Sanchirico & Triantis, 2007; Scott & Triantis, 2006). The court's verification of breach depends on the amount and accuracy of the evidence presented to it. Thus, although verification of breach often is treated as a binary and exogenous variable, verification is more realistically considered to be a continuous and endogenous variable. Where judicial verification is imperfect, the promisor will not always be held liable for its breach. The promisor therefore might breach when its performance is efficient. To deter inefficient breaches, damages must be multiplied by the inverse of the probability of detection (Craswell, 1999, 1996; Polinsky & Shavell, 1998; Klein, 1980). Damages therefore must be greater in amount than the actual loss. For this reason, damage provisions that stipulate penalties can induce efficient breach. Where renegotiation is feasible, penalty clauses encourage bargaining when litigation is more costly. They also affect litigation cost, reducing the breach victim's cost of proving damages and effectively placing on the breacher the burden of proving that the damages provision is a penalty.

The role of penalties in deterring breach often is limited. This is because damage stipulations do not function as damage multipliers in many contracts. Some contracts contain only precise material terms that describe aspects of performance that are easily observable by the parties and courts. Courts will not fail to detect breach in these contracts. Damage multipliers therefore serve no role in inducing efficient performance of contracts with only precise material terms. However, commercial contracts frequently contain a mix of precise and vague terms (Scott & Triantis 2005, 2006). The vague terms include best efforts, good faith or material adverse change of conditions clauses. Although such terms allow judicial error in determining the promisor's performance obligations, they are unlikely to induce inefficient performance, for two reasons. First, unless the judicial error is biased in favor of the promisor, the mean error rate is zero and the promisor's incentives to breach therefore are unaffected. Bias is unlikely because a court's findings about performance are based on evidence presented by both contracting parties. Second, nonremedial contract provisions and non-contractual devices can deter breach. Damage multipliers are not needed. This is because the promisor incurs nonrecoverable litigation costs in defending against an allegation of breach. Burdens of proof and presumptions adverse to it increase its litigation costs. They can be set so that the sum of the promisor's litigation costs and liquidated damages exceeds its gains from breach (Choi & Triantis, 2008). By increasing litigation costs, contractual and non-contractual devices can serve as substitutes for damage multipliers to deter breach. Stipulated damages with damage multipliers assure efficient performance only in contracts with vague terms lacking such substitutes. Relatively few contracts are of this sort. Thus, asymmetries in information, investment incentives, or unverifiable valuations are more likely to explain the presence of the use of stipulated damages clauses in the vast range of other contracts.

3.4 Asymmetric Information About Valuations

Contracting parties might not observe the promisor's cost of performance or the value of performance to the promisee. In this case information about cost and value is private to the promisor and promisee, respectively. Under limited conditions, parties will agree to a set of contracts that produce efficient trades: the promisor profits and its promisees select the contracts
each prefers to other contracts offered. Stole (1992) and Schwartz (1990) show that promisee-buyers select contracts containing different amounts of stipulated damages depending on the value each attaches to performance. In their models a seller with monopoly power makes a set of offers to sell a good to buyers with different valuations for the good. The menu of offers is rich enough to accommodate any valuation revealed by a buyer. Offers differ in price and the amount of the liquidated damages clause accompanying them. The promisor-seller knows the distribution of valuations buyers place on the good but not each buyer's particular valuation. Investment in performance is excluded and renegotiation is infeasible.

Stole and Schwartz model the buyers' choice of contract as a screening game. Buyers reveal their valuations to their seller by selecting a contract with a specific price and liquidated damages clause. The seller, who has market power, can therefore price discriminate and charge each buyer a price equal to its revealed valuation. Knowing this, a buyer must tradeoff the desire for compensation if the seller breaches against the desire not to pay a price equal to its revealed valuation (Edlin & Schwartz, 2003). Buyers will not select contracts with damage stipulations above their actual valuations because contract price will be above the value of performance to them. However, the buyers will select damage stipulations equal to or below their actual valuations. Buyers with the highest valuations select contracts with fully compensatory damage stipulations: they desire to protect their high valuation more than their desire to avoid paying a higher contract price that results from their revelation of valuation. Buyer with lower valuations select contracts with undercompensatory stipulations: they desire lower priced contracts more than they desire to protect their low valuations. Each type of buyer gets the type of contract optimal for it.

However, the outcome in Stole and Schwartz's models might not be socially optimal. Underliquidated damage stipulations result in efficient trades because buyers are not offered a blended price. Each buyer instead is offered a price-damages proposal associated with its revealed valuation. No buyer therefore exits the market because price is above this valuation. But contracts with underliquidated damages induce inefficient breach because the promisor does not have to pay damages equal to the buyer's value. Instead, the promisor will breach when its performance cost is greater than the amount of the damages stipulation, which is less than the buyer's actual value. Thus, there is a tradeoff between inefficient breach and efficient trade produced by underliquidated damages. These stipulations maximize social welfare only if the magnitude of inefficient breach (a cost) is less than the magnitude of efficient trade (a benefit). This inequality in turn depends on the distribution of different valuations among the population of buyers. For instance, if most buyers have high valuations, they will select contracts with compensatory damage stipulations. There will be therefore be few inefficient breaches accompanying efficient trade. On the other hand, if most buyers have low valuations, contracts with underliquidated damage stipulations will dominate. Sellers therefore frequently will breach inefficiently.

The role of underliquidated damages as a screening device is consistent with an aspect of existing penalty regulation: courts regularly enforce underliquidated damages. However, screening does not justify judicial scrutiny of any damage stipulation. Efficient trade instead requires the enforcement of all damage stipulations as written, even penalty clauses. Buyers reveal their different valuations by selecting offers with different damage stipulations. The fact
that buyers will not accept price-damage proposals with damages set above actual value to them is irrelevant. Offers with different stipulated damages enable trade when the promisor-seller cannot observe information about the promisee-buyer's valuation. The screening mechanism works whether damages in the offers are under- or overliquidated. Thus, screening does not support existing penalty regulation, which scrutinizes overliquidated damages.

### 3.5 Unverifiable Investment: Selfish Investment

Investment in a contract's performance can increase the joint surplus by reducing the cost or value of performance, or both. However, default remedies encourage the breach-against party to make inefficiently high investments. Expectation damages gives the breach victim an award equal to its value from performance. If the contract is performed, it will get its value from performance. If the contract is breached, expectation damages give it the returns on the investment had the contract been performed. In this case too the breach victim receives the value performance would have given it. Thus, the breach victim receives the return on its investment whether or not the contract is performed. In deciding to invest, it therefore does not take into account the likelihood that breach would be efficient. For this reason, the breach victim will overinvest in the contract's performance (Shavell, 1980; Rogerson, 1984; Sloof et al., 2006). Courts could reduce the incentive to overinvestment by reducing recoverable damages by the amount of overinvestment (Goetz & Scott, 1980; Cooter, 1985). However, this requires courts to observe the breach victim's investment. The requirement is too strong.

When remedies are unavailable, contracting parties also might underinvest. Some investments increase the joint surplus but have less value if deployed elsewhere. Non-redeployable investments enable the noninvesting party to appropriate through efficient renegotiation some of the returns from the investments (Williamson, 1985). Anticipating this, a party will not make investments that maximize these returns. Thus, even when the contract is performed, performance produces less value than it would have produced with investment. It would be a coincidence if the incentives to underinvest and overinvest generally cancel each other.

Research in contract design shows that penalty clauses can play a role in producing efficient investment incentives. The research exploits the insight that damage measures that decouple recoverable damages and investment can induce efficient investment. An expectation measure, for instance, calculates loss from breach net of the breach victim's cost of investment. It does not decouple damages and investment: the breach victim recovers its loss from breach less its investment (and reliance generally). Expectation damages therefore encourage the breach victim to overinvest in its performance of the contract. In contrast, damages measures that fix damages without regard to investment separate recoverable damages and investment. Because fixed damages measures make damages invariant to investment, investment can affect the breach victim's net return from breach. Thus, such fixed damages measures force the breach victim to take into account the effect of its investment on its net returns if the contract is breached. As a result, "decoupled" damage measures can be written to encourage efficient investment incentives.

A penalty clause is a type of fixed damages measure that decouples damages and investment. The simplest type of contracts which give efficient investment incentives are those in
which only one of contracting parties can invest in performance. Investment is unverifiable and "selfish": it benefits only the investing party, either by reducing its performance costs or increasing the value it receives from performance. Edlin (1996, 1998) describes the design of contracts of this type which produce efficient investment. The key is to structure the contract so that only the investing party will breach and it receives all the returns from its investment. Three conditions assure this. (1) The contract price is set below the investing party's marginal cost of performance. This induces the non-investing party to enter the contract and not breach. At the same time, the below-marginal cost price gives the investing party an incentive to breach. (2) The level of performance demanded by the contract is set so high that efficient renegotiation is unlikely. This assures that the noninvesting party never has the opportunity to appropriate through renegotiation any of the returns from investment. (3) The non-investing party makes a nonrecoverable payment to the investing party large enough to make the contract is profitable for the investor. The payment is needed to induce the investor to enter the contract.

The first and third conditions are particularly important. Condition (1) assures that only the investing party is likely to breach. Expectation damages gives the noninvestor the value performance of the contract would have given it. Thus, any residual value produced by investment goes to the breacher. The breacher therefore will calibrate the costs and benefits of its investment to make efficient investment decisions. Strictly, condition (3) does not require that the large nonrefundable payment to the investing party be transferred up-front. It is sufficient that the noninvestor be obligated to pay the investor this amount. More important, the noninvestor must not be able to recover its up-front payment or cancel its payment obligation if it breaches. Otherwise, the investing party's incentive to overinvest is reinstated (Edlin, 1996; Edlin, 1998; Edlin & Schwartz, 2003). Investment increases the investor's profit from performance and therefore also its expectation damages from the noninvestor's breach. The large damage award reduces the amount of the up-front payment the investor must refund or the payment obligation the investor must reduce. This gives the investor an incentive to invest even when investment is inefficient. The investor's incentive to invest efficiently remains when the investor's payment is nonrefundable. Because the large payment to the investor is unrelated to the investor's damages from breach, it is a penalty.

3.6 Unverifiable Investment: Bilateral and Cooperative Investment

The contract design literature identifies other contracts in which penalty clauses promote efficient investment. Two types of contracts are noteworthy. One type involves contracts in which both contracting parties can make investments that benefit only the investing party. These investments are “bilateral.” Edlin and Reichelstein (1996) show that a contract cannot be written that gives efficient investment incentives to both parties under an expectation damages remedy. They also demonstrate that specific performance will produce efficient investment in contracts with bilateral investment. Edlin and Schwartz (2003) show that penalty clauses have the same effect on bilateral investment. The intuition behind Edlin and Reichelstein's result can be described in general terms. Efficient investment discounts the investment by the probability of its returns. Expectation damages gives the breach victim its profit, which takes into account returns from investments in performance the victim has made. The remedy therefore encourages overinvestment by the victim, as noted above. Where the contract is efficiently renegotiated, part of the return from investment is appropriated by the non-investing party. This encourages the
investor to underinvest in performance. The investor's incentive to invest is efficient when its incentive to overinvest is balanced by its incentive to underinvest.

Edlin and Reichelstein notice that contracts subject to expectation damages cannot balance the incentive to overinvest against the incentive to underinvest for both contracting parties. This is because expectation damages do not give the breacher the full returns on its investment. Instead, the breacher keeps only returns, if any, above the amount of the breach victim's loss from breach. Because the noninvestor appropriates part of the returns from investment when renegotiation is efficient, the breacher's incentive to underinvest remains. Thus, the breacher's incentive to overinvest will not offset its incentive to underinvest. The breacher's investment incentives therefore will be inefficient. Setting the performance obligations of the contract high gives the breacher efficient investment incentives. Because renegotiation of this contract is remote, the nonbreacher is unlikely to appropriate part of the returns from the breacher's investment through renegotiation. The breacher's incentive to underinvest therefore is weak. Thus, the breacher's weak incentive to overinvest will dominate its weaker incentive to underinvest. However, a high performance standard increases the breach victim's incentive to overinvest because expectation damages give it the returns on its investment. The breach victim's incentive to overinvest therefore will dominate its diminished incentive to underinvest. By contrast, contracting involving selfish investment, described in Section 3.5, do not have this result. This is because the level of performance set by the contract is high so that only the investing party will breach. In these contracts the likelihood of an efficient renegotiation of the contract is remote and the incentive to underinvest therefore weak.

Contract penalties restore efficient investment incentives to both parties. To see this, recognize that breach will not occur when performance under the contract is inefficient. The parties instead will renegotiate to obtain efficient performance. Because both parties can gain when performance is efficient, their renegotiation shares the surplus from efficient performance. The inefficient contract either sets performance above or below the efficient level. A penalty enables the nonperforming party to obtain a renegotiated share even when contract performance is set inefficiently high. This is because it serves as credible threat to the performing party: the penalty will be enforced if the performing party does not share the efficiency surplus from performance at the lower, efficient level. The performing party's loss from performing under the contract or paying the penalty is greater than its loss from performing at the lower level. Thus, the performing party will renegotiate to share the returns from performing at the lower, efficient level. The nonperforming party's share of the returns encourages it to overinvest in the contract. At the same time, the nonperforming party's incentive to underinvest remains when contract performance is set inefficiently low. This is because the performing party can appropriate by renegotiation a share of surplus from modifying the performance level upward to an efficient level. The nonperforming party's incentive to overinvest therefore balances its incentive to underinvest. The previous reasoning applies to both contracting parties. Thus, penalties encourage both parties to make efficient investments.

This result is limited to bilateral investment. Che and Chung (1999) demonstrate that expectation damages and stipulated damages both give inefficient investment incentives when investment benefits only the non-investing party. This sort of investment is "cooperative"; it benefits the non-investor while not diminishing the investor's cost of contractual performance.
Cooperative investments produce positive externalities. As a result, when renegotiation is infeasible, a contracting party will not make them. This is because the party receives the same return whether or not it invests. If the counterparty breaches, expectation damages do not increase the investor's award, because cooperative investment does not reduce its costs of performance. If the counterparty performs, investment again does not increase the investor's returns, for the same reason. A damages stipulation can encourage cooperative investment by setting damages above the cost of performance. But the stipulation produces inefficient trade when the damages set exceed the non-investor's valuation of performance. Renegotiation eliminates trading inefficiencies and encourages cooperative investment. However, it makes an initial contract superfluous, because the parties can bargain for a share of the surplus from such investment. A damages stipulation clause therefore is not needed to encourage cooperative investment (Che & Hausch, 1999).

3.7 Unverifiable Investment and Contract Design

The second type of contract that encourages efficient investment is one in which a penalty clause induces the disclosure of accurate information about investment. As before, investment, costs and valuation are unverifiable. The contracting parties can observe them; the court cannot. Realistically, cost and valuation depend on investment in performance. Writing a contract with terms specifying these variables is useless because a court cannot ascertain whether performance complies with the contract. Ex post bargaining after investment has been made and costs and valuations realized can be costly. Models of "mechanism design" show that the parties can devise a procedure or mechanism which elicits from them accurate private information about investment, cost and valuation (Moore & Repullo, 1988; Moore, 1992; Paltry, 2001; Bolton & Dewatripont, 2005). Under the narrow conditions of the models, a court enforces the contract based on the information elicited. Some of the contracts in these models use penalties to induce accurate disclosure of private information.

A very simple model involves two risk neutral parties: a seller and buyer. The parties contract for the seller to produce and deliver a good to the buyer. Only the seller can invest in performance, which reduces the seller's production cost. The contract specifies the following mechanism to fix the contract price of the good, adapted from Schmitz (2001): The buyer and seller both report the seller's cost and the buyer's valuation of the good to a court. If their reports match, the good is traded at a price equal to the buyer's reported valuation, unless the reported costs exceed the reported valuation. If the reports differ, no trade occurs and both the buyer and seller each pay a large penalty to the court. The parties' agreement prohibits renegotiation of the contract. Under these conditions, making a truthful report is a weakly dominant strategy for each party. If the party's report is truthful and the other party's report also truthful, the good is traded as long as the buyer's reported valuation exceeds the seller's reported cost. The party is better off than if it lied: lying results in no trade and assessment of a large penalty against the party. If a party's report is truthful and the other party's report lies, no trade results and a penalty is assessed. In this case the party is no worse off in the circumstances than if its own report were truthful. Because the same reasoning applies to both parties, both have an incentive to truthfully report costs and valuation to the court. The penalty gives the parties part of their incentive to submit truthful reports.
Two of the model's key assumptions are unrealistic. The model assumes that the parties' commitment not to renegotiate is irrevocable, as do most mechanism designs (Tirole, 1999; Maskin & Tirole, 1999; Bolton & Dewatripont, 2005). If the parties' choices produce the no-trade outcome, they have an incentive to renegotiate rather than pay a penalty to the court. The threat of the no-trade outcome is not credible because renegotiating to trade makes both parties better off. Recognizing this, each party might choose to submit an untruthful report of valuation and costs. To deter selection of this off-equilibrium strategy, the mechanism in the contract must be enforced (and known by the parties to be enforced). Courts generally will not enforce a contract clause prohibiting renegotiation. Even with judicial enforcement, the parties remain free to renegotiate on their own to avoid paying a penalty (Brooks, 2002). To prevent renegotiation, the mechanism must be implemented so that there is no gap in time between its use and enforcement of its outcome (Maskin & Moore, 1999). Maskin and Tirole (1999) suggest in passing that the mechanism might be implemented before an arbitrator, who enforces the clause. But arbitration still does not guarantee that the no-trade penalty will be enforced in the simple mechanism above. The parties must pay the penalty to the arbitrator, and they have an incentive to renegotiate on their own to avoid doing so. Their arbitration agreement might require establishment of standby letters of credit to assure payment, but parties worried about arbitral self-dealing will reject the arrangement. Whether a commitment not to renegotiate is irrevocable obviously is an empirical question. However, it appears that renegotiation remains possible in many cases.

The other key assumption is that the no-trade penalty will be implemented. This too might be unrealistic. Existing law does not allow courts to collect penalties as dictated by the parties' agreement. A court's authority to impose penalties instead is given by statute expressly or by jurisdictional grant. Courts usually are not authorized to do so simply because parties submit conflicting reports to them. True, the parties' contract perhaps could select a jurisdiction's law which will enforce its penalty provision. Combined with a choice of forum clause, the choice of law might guarantee that a court or arbitrator will order a no-trade or other penalty. However, other jurisdictions might refuse to enforce the judicial order or arbitral award, deeming it an unenforceable penalty. This limits enforcement when the cooperation of courts in other jurisdictions is required. Again, a standby letter of credit or other payment device could assure payment of the penalty to the court or arbitrator. But concern about judicial or arbitral self-dealing might dissuade the parties from assuring payment of the penalty. Although penalties sometimes can be implemented through artful drafting and payment design, they frequently will not be enforceable.

4. Implications for Existing Penalty Regulation

Existing penalty regulation assumes that penalties are undesirable and that parties sometimes put penalty clauses in their contracts. The research described above denies this assumption. It finds that the damage stipulations either are not penalties or that penalties can provide efficient incentives to breach, investment or trade. The former finding makes penalty regulation superfluous; the latter finding shows that penalties sometimes are desirable. None of the research supports penalty regulation. It cannot because existing doctrine does not take into account the different variables the research identifies as affecting efficient investment, breach and trade. In vetting damage stipulations, penalty regulation assumes that courts accurately estimate
the parties' valuations from performance. By contrast, models with symmetric but unverifiable information assume that courts make measurement errors. Screening models with asymmetric information require courts to enforce damage stipulations as written. Mechanism designs also require enforcement of penalties in order to induce disclosure of private information about valuations to the court. Penalty doctrine voids damage stipulations that are penalties. This undermines the role of damage stipulations in promoting efficient trade. Existing doctrine also does not take into account the effect of penalties on incentives to invest in contract performance. It application therefore is indifferent to whether investment is one-sided, bilateral or beneficial to both parties. For all of these reasons, a penalty doctrine that incorporated the variables identified by the research above would look very different from existing regulation of damage stipulations.

5. Arguments for Penalty Regulation

Several arguments in the more recent literature support some form of penalty regulation. Penalty clauses can create externalities by deter entry into a product market. They also can lead inefficient signaling in contracts with asymmetric information. And contracting parties might incorporate penalty clauses in their contracts because they systematically misjudge the likelihood of breach. These arguments are either unconvincing or, if convincing, apply only in limited settings. None support penalty regulation in its present form.

5.1 Deterring Efficient Entry

A contract with a penalty clause might benefit both contracting parties. However, the clause makes it less likely that one party (the "buyer") will breach by buying from a lower-priced seller. This reduces the size of a potential entrant-seller's market and therefore the likelihood of entry by more efficient sellers. Consequently, other buyers pay a higher price for the relevant product. Penalty clauses do not deter entry completely. Entrants who are sufficiently efficient to offer a price below that of the incumbent seller by the amount of the penalty still will enter. The penalty clause merely transfers some of the entrant's surplus to the buyer and incumbent seller. However, penalty clauses deter entrants who are only moderately more efficient than the incumbent. By deterring entry, penalty clauses create a negative externality for other buyers.

Aghion and Bolton (1987) and Chung (1992) describe models in which penalty clauses deter efficient entry. In Aghion and Bolton's model a monopolist seller enters into a contract with a buyer that contains a penalty clause. Both the incumbent seller and buyer expect sellers with lower marginal costs to enter but cannot identify them in advance. The penalty clause enables the incumbent and buyer to collude to exercise monopolistic power against more efficient entrants. Without the penalty clause, entrants could offer the buyer a lower price and the buyer would breach its contract with the incumbent. The buyer would pay damages to the incumbent equal to its lost profits, and the entrant retains the economic surplus from its lower costs. With a penalty clause, an entrant must offer the buyer a price below the incumbent's price by an amount equal to the amount of the stipulated penalty. This is the amount the buyer must pay the incumbent in damages if it breaches by buying from the entrant. The penalty clause therefore enables the incumbent and the buyer to extract a portion of the economic surplus from the entrant: the incumbent captures a portion of it in the penalty, and the buyer in the lower price it pays the entrant. The difficulty is that it allows the extraction of surplus from the entrant only when entry occurs. Potential entrants whose marginal costs are not lower than the incumbent's price by the amount of the penalty will not enter. In this case the penalty clause inefficiently deters entry.
Segal and Whinston (2000) extend Aghion and Bolton's result to settings in which production involves economies of scale and incumbents can make discriminatory offers. In these settings incumbents can use penalty clauses to exploit externalities among buyers. The presence of economies of scale means here that entry is not profitable for a rival if a minimum number of buyers have entered into contracts containing penalty clauses with an incumbent monopolist. Discriminatory offers allows the incumbent to make different offers to buyers to achieve the minimum scale needed to deter entry. In this way the incumbent can lower its monopoly price to entice the minimum number of buyers to agree to contracts with penalty clauses. Agreeing to a contract with a penalty clause is rational for each buyer because they gives her a lower price than otherwise. Having obtained contracts with these buyers, the incumbent can make offers with higher monopoly prices to remaining buyers. Thus, buyers who agree to contracts with penalty clauses increase prices for remaining buyers. They therefore impose a negative externality on remaining buyers. As in Aghion and Bolton's model, penalty clauses thereby enable the incumbent to inefficiently deter entry by more efficient rivals.

Spier and Whinston (1995) study the effect of investment on deterring entry when renegotiation is possible. They assume that the monopolist incumbent seller can invest to increase the profit from its contract with the buyer and that renegotiation is costless. When a buyer receives a lower priced offer from a more efficient entrant, it will renegotiate with the incumbent to eliminate the penalty clause. Their renegotiation allocates the buyer's surplus from the entrant's offer between the buyer and the incumbent, according to their bargaining power. Renegotiation therefore undoes the deterrent effect of a penalty clause on entry. However, investment deters entry even when renegotiation occurs. This is because the incumbent's investment in performance reduces its costs and therefore increases its profit from performance. An entrant therefore must offer a lower price to the buyer sufficient to cover the incumbent's expectation damages resulting from the buyer's breach. Because investment increases the incumbent's profit from performance, it increases the incumbent's expectation damages. Thus, the incumbent has an incentive to overinvest in the contract. If the buyer breaches when a more efficient entrant enters, investment increases the incumbent's expectation damages. This deters entry by moderately more efficient rivals. If entrance does not occur, its investment again increases its profit from performance.

The incumbent's overinvestment in Spier and Whinston's model results from its market power. In order to extract some of an entrant's surplus from entry, the incumbent and its buyer set high stipulated damages. This gives the incumbent a monopolistic advantage over a more efficient entrant: entry is profitable for a rival only when its marginal costs are below those of the incumbent by the amount of the stipulated damages. High stipulated damages in turn encourage the incumbent to overinvest in the contract's performance because it recoups its investment whether or not the buyer breaches. The incumbent's incentive to overinvest is not present when its contract lacks high stipulated damages. In this case the entrant pays only the incumbent's expectation damages from breach and keeps the economic surplus from its lower-priced offer.

The possibility that penalties can create barriers to efficient entry does not justify penalty regulation. This is because the rationale's central assumption of market power is unlikely to hold in the broad range of settings in which penalty regulation applies. In markets where buyers can identify competing sellers, they can determine with certainty whether entry will occur. Buyers
need not agree to contracts with penalty clauses in advance of entry. Rival sellers in these markets therefore can induce buyers to reject contracts from sellers with penalty clauses by offering them lower priced contracts. Incumbents and rivals compete for the buyer's business, driving down price to a competitive level. The buyer gets a lower price contract and the entrant keeps the surplus from its lower priced contract. Where sellers lack market power, the rationale that penalty clauses have anticompetitive effects on price does not justify penalty regulation. The penalty doctrine nonetheless applies even to contracts in competitive markets.

More generally, the penalty doctrine applies broadly, without regard to the market structure in which contracting occurs. It holds for contracts to which both monopolist and competitive sellers are parties. As noted in Section 4., it also applies without regard to whether investment in contractual performance is feasible. Penalty regulation applies too even when renegotiation is infeasible. At best, market power might justify a presumption against the enforcement of stipulated damages clauses. For example, a monopolist incumbent justifiably might bear the burden of proof that a damages stipulation in its contract is not a penalty. Or the incumbent might be required to show that its penalty clause does not deter efficient entry. Such proposed allocations of proof apply only to monopolists. They do not resemble penalty regulation, which applies generally in all market settings. Thus, the conditions described in Aghion and Bolton's model, and its extensions, are not robust enough to justify existing penalty regulation.

5.2 Inefficient Signaling

Penalty clauses can inefficiently signal information about the quality of a contracting party’s performance. Under specific conditions banning them induces parties to provide an efficient amount of information. Aghion and Hermalin (1990) describe a model of financial investment under asymmetric information in which legal restrictions on contracts can increase welfare. The model assumes a number of entrepreneurs need financing for their projects. Projects are of two sorts: those with a high probability of success (“good” projects) and those with a low probability of success (“bad” projects). Entrepreneurs know whether their projects are good or bad; investors do not. The outcomes of projects are verifiable, so that contracting parties can contract on them. However, the quality of a project is known only to entrepreneurs; it is not contractible. Lastly, the model assumes that entrepreneurs are risk-averse and investors risk-neutral or risk-averse.

Entrepreneurs get financing from investors in return for a promise to repay a larger amount if the project succeeds or if it fails. Because investors likely will not be repaid in full if the project fails, they prefer to invest in good rather than bad projects. Entrepreneurs therefore want to signal in their investment contracts that their projects are good. A promise to repay a large amount conveys this information. At the same time, the promise imposes a risk of significant loss on the entrepreneur whose project fails, which is a cost to it.

To indicate that it has a good project, a good entrepreneur will offer an investment contract with a large repayment promised. A bad entrepreneur may or may not mimic the same signal. If it sends the same signal, investors will find the signal uninformative and conclude that the entrepreneur is offering the project of average quality. Investors will demand an investment contract with a repayment promise for an average project. If the bad entrepreneur sends a
different signal or none at all, investors will be able to identify the project as "bad." Investors therefore could distinguish good from bad projects and invest accordingly. The pooling equilibrium in which both types of entrepreneurs signal and the separating equilibrium in which only good entrepreneurs signal can be inefficient. Good entrepreneurs in both equilibria assume risks of significant repayment obligations. Because entrepreneurs are assumed to be risk-averse and investors can be risk-neutral, this risk might more efficiently be allocated to investors. A restriction on signaling can make both good and bad entrepreneurs better off than in the equilibria identified. With the restriction in force, investors will consider all projects to be of average quality. Good entrepreneurs avoid the cost of taking on excessive risk of significant repayment, and bad entrepreneurs have their bad projects considered average.

Aghion and Hermalin's model of asymmetric information might justify voiding penalty clauses in contracts. Assume that contracting parties are of two types: good and "bad." Good types are likely to perform according to the contract; bad types are unlikely to do so. While the performing parties know their own type, their counterparties do not. Counterparties prefer good types to bad types because they are unlikely to be compensated fully in the event of nonperformance and good types are more likely to perform. To signal that it is a good type, a good type can promise to pay a penalty if it fails to perform. Bad types might mimic this signal or not, as in Aghion and Hermalin's model. The resulting equilibria are inefficient if the promised penalty allocates to the good or bad risk-averse promisor a risk that is more efficiently borne by risk-neutral counterparties. If penalty clauses in contracts are void, these inefficient equilibria will not result.

Aghion and Hermalin's model does not justify existing penalty doctrine, for at least two reasons. First, as Aghion and Hermalin acknowledge, their model only shows that legal restraints on contracts can increase welfare. It describes a mere possibility. Whether restrictions in fact increase welfare depends on the robustness of the conditions underlying the model (Hermalin et al., 2007). However, the model's underlying conditions are fragile. The superiority of a "no signaling" equilibrium depends on the particular shape of risk-averse good and bad entrepreneurs respective indifference curves. A good entrepreneur with a project that will almost certainly succeed (a "very good" entrepreneur) might prefer to signal the project's above-average quality. It values the larger reduction in the investment contract price associated with a very good project more than avoiding the small risk of its project failing. At the same time, a bad entrepreneur with a project that is almost certain to fail (a "very bad" entrepreneur) might prefer to signal that its project is below average. The very bad entrepreneur is more concerned with avoiding the repayment obligations associated with average projects than obtaining the advantageous investment contract price attached to average projects. A "no signaling" equilibrium makes both types of entrepreneur worse off: the very good type, because it cannot distinguish itself from bad entrepreneurs, and the very bad type because investors will infer that its project is of average quality. More generally, the continuum in quality of actual contracts is unlikely to exhibit the specific quality of projects assumed by Aghion and Hermalin's model. The same is true for the potentially wide range of contracts with penalty clauses. This is consistent with Ayers' finding that corporate charters and by-laws do not exhibit excessive signaling (Ayers, 1991).

Second, existing penalty doctrine applies more broadly than models of asymmetric information allow. These models find that legal restrictions on contracts can improve on equilibria produced with parties with private information signal excessively. Implementing these
restrictions requires courts to identify inefficient signaling equilibria (Posner, 2003). Courts lack the information needed to do so. (Adler reaches a similar conclusion with respect to judicially created penalty default rules for damage limitations (Adler, 1999).) More important, models of asymmetric information do not support penalty regulation when contracting parties are symmetrically informed. When parties have no private information about performance, their contracts are efficient (Hermalin & Katz, 1993). Legal restrictions on penalties therefore cannot improve on these contracts. However, penalty regulation is not sensitive to the information parties possess. Courts void penalty clauses in contracts without regard to whether parties are symmetrically or asymmetrically informed. For this reason, models of asymmetric information do not support existing penalty regulation.

5.3 Cognitive Error

Some scholars maintain that stipulated damages clauses often are the product of systematic errors made by the contracting parties in estimating damages from breach (Eisenberg, 1995, 1998; Marrow, 2001). They conclude that these cognitive errors justify penalty regulation. Their conclusion relies on laboratory studies documenting that experimental subjects systematically make choices and probabilistic judgments that violate standard axioms of rational choice theory (Kahneman & Tversky, 2000; Govitch et al., 2002; Camerer, 1995). Subjects make different choices in response to differently formulated but equivalent descriptions of options. They exhibit different preferences among options depending on the procedure for eliciting preferences. Subjects also show more aversion to losses relative to an initial level of assets than they are attracted to gains from the same asset level. In some experimental contexts they underestimate or overestimate the likelihood of events. These and other findings suggest that people exhibit a range of cognitive errors in making decisions (Jolls et al., 2000; Kahneman & Tversky, 1996). Contracting parties who make cognitive errors could draft stipulated damages clauses that inaccurately forecast the damages from breach. Accordingly, penalty regulation might be a justified legal response to the tendency to make cognitive errors in estimating damages.

Eisenberg (1995, 1998) defends penalty regulation based on the apparent ubiquity of three documented cognitive biases. One bias estimates probabilities by the ease with which a type of event can be recalled or imagined. This bias is the result of the “availability heuristic”: a simple rule of thumb that estimates likelihoods by their availability to recall or imagination. A second bias is “overconfidence”: the tendency to underestimate the likelihood that one's judgment is inaccurate. The third bias is representativeness: the tendency to base probability judgments on the extent to which evidence is representative of a category. Eisenberg postulates that all three biases often operate among contracting parties to produce systematic inaccurate damage stipulations. The bias favoring availability apparently gives prominence to the intention to perform, which is salient, not to the possibility of breach, which is remote in recall. Contracting parties subject to this bias underestimate the probability of breach. Overconfidence about performance also leads parties to underestimate the likelihood of breach. Similarly, the representativeness bias induces the parties to take their intention to perform as representative of the likelihood that they will intend to perform in the future. All three cognitive biases produce inaccurate stipulated damages provisions. In contrast, these biases apparently do not operate when parties draft the terms of contract performance. Eisenberg concludes that, consistent with
existing penalty doctrine, systematic cognitive error justifies special judicial scrutiny of stipulated damages provisions only.

Eisenberg suggests that cognitive error likely operates when there is a gross disparity between estimated and actual damages. Accordingly, he proposes that courts invalidate a stipulated damages provision where such disparity exists, unless the parties specifically intended the provision to apply to the situation in which breach occurred. The specific intent to apply stipulated damages to this situation signals that the damages stipulation is the produce of rational contract design, not cognitive error. Under the proposal the defendant bears the burden of establishing the specific intent to apply the stipulation to the situation in which breach occurred. Because parties might specifically intent penalty provisions to apply, the proposal revises existing penalty regulation without abandoning it, as Eisenberg acknowledges.

For at least four reasons, cognitive error does not justify penalty regulation, even in the revised form advocated by Eisenberg. First, the experimental evidence of cognitive bias is insufficiently robust to support such regulation. Penalty doctrine applies generally, without regard to the type of contract or the characteristics of the contracting parties. Accordingly, parties must exhibit cognitive biases in most contract settings. However, laboratory studies document bias in limited experimental environments. These environments are not sufficiently representative to allow reliable extrapolation to nonexperimental settings, including exchange (Lowenstein, 1999; Hillman, 2000; Walt, 2003). For instance, parties overconfident about their skills or optimistic about their future might underestimate their risk of breach or its consequences (Camerer & Lovallo, 1999; Brenner et al., 1996). However, individual debiasing techniques, organizational safeguards in firms and interfirm competition can mute or prevent the bias from operating (Romano, 1986; Arkes et al., 1991; Heath et al., 1997; Kadous et al., 2006). As important, individual differences in sophistication and capacity make generalizations about the prevalence of bias unsound (Mitchell, 2002). Because laboratory studies do not support a finding of systemic cognitive bias generally, a gross disparity between estimated and actual damages does not signal cognitive error.

Second, there is no general theory specifying the conditions under which specific cognitive biases operate. This makes penalty regulation difficult to implement effectively. Consider “ambiguity aversion”: the preference for options with precise probabilities over options with unknown or imprecise probabilities. Studies show that experimental subjects are averse to ambiguity in limited settings (Fox & Tversky, 1995). Contracting parties averse to ambiguity will prefer liquidated damages, which can be made precise, to uncertain damages estimated ex post by a judge or jury. In drafting damages provisions, the parties might therefore carefully consider the likelihood of breach and resulting loss (Hillman, 2000). Their attention to the array of risks of breach in turn might diminish or eliminate any tendency to be overconfident about performance. In this case the parties’ bias toward overconfidence might not produce inaccurate damage estimates. Thus, the effect of cognitive bias on the accuracy of stipulated damages cannot be predicted without specifying the conditions under which bias operates (Rachlinski, 2000). The cognitive error justification of penalty regulation does not specify these conditions.

Third, the assumed ubiquity of cognitive error makes it likely that courts too are subject to cognitive bias (Rachlinski, 2000). As a result, courts are unlikely to scrutinize damages
stipulations effectively. They will overestimate the parties' ability to fix damages accurately at the time of contracting and void damages stipulations that are reasonable forecasts ex ante of actual damages. Guthrie et al. (2001, 2007) present experimental evidence of cognitive bias among judges that could produce this result. Their experimental judicial subjects exhibit hindsight bias: the tendency to use known outcomes to assess the likelihood at an earlier time of the occurrence of events (Guthrie et al., 2001, 2007). Judges with hindsight bias judge past events to have been more predictable than they actually were. Guthrie et al.'s judicial subjects also overestimate their own abilities to interpret information accurately (Guthrie et al. 2001). Courts subject to these biases will do a poor job at scrutinizing damages stipulations. In vetting stipulated damages, hindsight bias will lead courts to overestimate the ability of contracting parties at the time contracting to set damages that approximate actual damages. This bias is reinforced when judges overestimate their own abilities to accurately assess the ease of estimating damages ex ante. In combination the biases induce courts to void damages stipulations that were reasonable ex ante but disproportionate to actual damages. They lead courts to find the stipulations to be penalties that penalty doctrine, properly applied, deems enforceable liquidated damages. Because even specialist courts appear to be subject to cognitive bias (Rachlinski et al., 2006), they too might scrutinize damages stipulations poorly. For all three reasons, cognitive error does not justify penalty regulation. At most damage stipulations are properly voided when they are demonstrated to be the result of cognitive error in a particular case.

Fourth, penalty regulation reduces the incentive of contracting parties to forecast damages accurately. Parties avoid the cost of inaccurate damages stipulations produced by cognitive error, because courts effectively will rewrite damage stipulations. In drafting damages stipulations they therefore will underinvest in avoiding cognitive error or its operation. Judicial estimations of damages are superior to party-provided estimates only if courts are better positioned than parties to measure loss from breach accurately. But courts usually are comparatively poorly positioned. The assumed ubiquity of cognitive bias means that courts too exhibit bias in vetting stipulated damages. In addition, parties often have better information about the value of the contract than courts (Goetz & Scott, 1977; Ben-Shahar & Bernstein, 2000). Thus, penalty regulation implemented by courts subject to cognitive error likely results in suboptimal measurement of damages.

6. Conclusion

The literature on optimal remedies shows that the normative implications of traditional damage measures are ambiguous. Contract remedies affect the selection of contracting partner performance, investment, breach and renegotiation. None of the traditional damage measures are optimal with respect to all of these variables (Craswell, 1986, 2003). The recent literature on penalty regulation supports a similar conclusion with respect to stipulated damages. Stipulated damages affect investment in performance, breach and trade. Penalty clauses can induce efficient investment in the contract while also inducing inefficient performance. Under other conditions they produce efficient trade while also encouraging inefficient breach. In different conditions penalties encourage efficient breach when breach is difficult to detect. The recent literature shows that no damage stipulation, whether a penalty or liquidated damages, likely is optimal with respect to investment, breach and trade. Rather, the optimal damage stipulation under realistic conditions likely must trade off its incentive effect on these variables. Existing penalty regulation
unjustifiably assumes that penalties have an overall inefficient incentive effect.

Penalties may or may not produce efficient incentives even with respect to an isolated variable such as investment. The recent literature shows that the effect of penalties on the incentive to invest is sensitive to the nature of the investment, market power and contract design. Penalties induce overinvestment when only the breach victim benefits from investment. They also encourage overinvestment by a monopolist when it uses its market power to deter entry by more efficient rival suppliers. On the other hand, penalties induce efficient investment when investment benefits only the investor and only the investor can breach. Penalties in well designed contracts also can encourage efficient investment by inducing contracting parties to reveal accurate information about investment when the information is not verifiable by courts. Taken together, recent work demonstrates that penalties do not have an unambiguous effect even on investment.

The work has a normative implication for existing penalty regulation. It shows that penalty clauses can be efficient under specific conditions, depending on the nature of investment in performance or whether valuations are observable to the parties or verifiable to a court. For their part, even contributions arguing that penalties can be inefficient specify specific conditions, such as cognitive bias or market power, in which penalties result in inefficient investment, trade or breach. Some of the conditions specified require courts to obtain information often unavailable to them, such as valuations among contracting parties or the character of investment. Nonetheless, all of the work evaluates the efficiency of penalties by isolating particular variables. A fair implication of the work as a whole is that rules regulating stipulated damages, if any, should take these variables into account. Penalty doctrine does not do so. Instead, it applies generally, without regard to the variables identified in the literature. It is therefore insensitive to the character of investment, the verifiability of valuations among parties, or market structure. Penalty doctrine also applies whether or not valuations are observable to the parties or whether breach is difficult to detect. Because it does not incorporate such variables, penalty regulation is unlikely to enforce efficient penalties clauses and void inefficient penalty clauses. For this reason, penalty doctrine is remote from the scheme of penalty regulation (if any) justified by the recent literature.

Acknowledgments

I thank Paul Mahoney for comments on a previous draft and Michael Zadd for assistance in the preparation of this Chapter.

Bibliography


Council of Europe, European Union Committee on Legal Cooperation (1978), Penal Clauses in Civil Law, Strasbourg.


Lowenstein, George (1999), 'Experimental Economics From the Vantage-Point of Behavioral Economics', 109 The Economic Journal F25-F34.


Rae, Samuel A., Jr. (1982), 'Non-Pecuniary Loss and Breach of Contract', 11 Journal of Legal
Studies, 35-53.


Stole, Lars A. (1992), 'The Economics of Liquidated Damage Clauses in Contractual

http://law.bepress.com/uvalwps/olin/art57


**Other References**


