Abstract

This chapter summarizes the economic understanding of insurance regulation with a focus on solvency regulation, underwriting regulation, contract regulation and competition law exemptions. Insurance industry solvency regulation is viewed as a solution to a collective action problem that might otherwise induce insurers to take excessive risk. The chapter analyzes the economics of regulations addressing adverse selection, including traditional doctrines of warranty and misrepresentation as well as recent laws relating to genetic testing. Traditional doctrines examined from an economic perspective include contra proferentum, ‘extra-contractual damages’ and resolution of conflict of interest problems arising out of typical liability insurance policies. Competition law exemptions are seen as perhaps an overly broad vehicle to foster the economies of scale in production of information needed to prevent adverse selection and insolvency.

*JEL classification:* K20, K29

*Keywords:* Insurance Regulation, Solvency Regulation, Underwriting Regulation, Adverse Selection, Genetic Testing, Warranty, Misrepresentation, Competition Law, Contra Proferentum, Contract Remedies

A. Introduction and Scope Note

In almost all jurisdictions, government heavily regulates the business of insurance. To varying degrees in different states and nations, it uses its legislative and judicial branches to regulate solvency, underwriting practices and contract structure. It likewise often distinguishes the business of insurance from other businesses regarding the extent to which separate enterprises must behave competitively rather than cooperatively. While economics-oriented scholarship has not yet spanned the breadth of these varying forms of insurance regulation or plumbed the intricacies of every doctrine thereunder, an increasing mass of literature has begun to review these regulatory practices. This entry seeks to develop the main contours of the existing body of research.

It should be noted that several critical topics are addressed only briefly in this chapter. The focus is on insurance regulation through legislative action and
judicial rulings as it has developed in the United States. The chapter does not
discuss purely legal analyses of insurance regulation or analyses that draw on
other disciplines such as political philosophy, nor does it discuss purely
economic analyses of the insurance industry, including the extensive empirical
work in the field. Rather, the focus is on analyses that have attempted to bring
the tools of economic analysis to bear on the laws regulating insurance.
Economic analysis of laws requiring purchase of insurance is addressed in a
separate chapter.

B. Solvency Regulation

1. Traditional Economic Understanding

Economic theory has often understood efforts by government to monitor and
regulate the solvency of insurers as a solution to a collective action problem not
unlike that arising with banks and other financial intermediaries (Winter,
199b). Insureds in this view are analogous to fixed-debt investors putting
money into insurers who reinvest that money and return sums to the insured in
the event certain events come to pass (Hansmann, 1985). This beneficial
intermediation of investment requires an expectation by the investor/insured
that the insurer will actually be able to pay those sums when the events occur
at some point in the relatively near future (property/casualty insurance) or in
the distant future (life insurance). Given the incentive of insurer/owners with
limited liability to invest in high risk ventures, however, the likelihood of
repayment is hardly assured. This incentive exists because insurer/owners
capture all the upside of a favorable materialization of the risk while they are
able to shift part of the downside of an unfavorable materialization onto the
investor/insureds. While economic theory suggests that investor/insureds may
be able to constrain this ‘agency cost’ or ‘moral hazard’ by use of monitored
conditions in contracts (Smith and Warner, 1979; Mayers and Smith, 1981) or
through bonding (mutual ownership) (Fama, 1980) few individual
investor/policyholders can structure the complex arrangements needed to
achieve this end. Free ridership problems hinders groups of
investor/policyholders from coalescing to accomplish this aim. These problems
are particularly weighty with respect to insurance because non-payment to the
investor/insured occurs after the insured has suffered an often catastrophic loss
at times when prompt payment is particularly valuable.

It is this understanding of the insurance industry that motivates the
predominant regulatory structure: governments intervene by licensure and
public insurance mechanisms often known as ‘guaranty funds’ that provide at
least partial indemnity to selected policyholders of an insolvent, licensed
insurer (Winter, 1991b). The licensure mechanism generally imposes severe
punishment for sale of insurance without a license and conditions licensure on regulation and monitoring of insurer finances. The public insurance mechanism transfers the risk that this system will fail in the case of a particular insurer away from the individual insured to a collective of all insurers and, derivatively, their insureds, although this backup form of risk transfer, if effective, minimizes the incentive of insured/policyholders to monitor insurer solvency.

Although this mode of regulation is prevalent throughout the world, different systems of shared sovereignty differentially allocate responsibility for solvency regulation. In the United States, a complex system exists under which, for most insurers, each of the states may regulate the solvency of insurers ‘doing business’ in that state. Regulations need not be consistent among the states (Lewis and Coates, 1991). The danger of excessive regulation posed by this system, particularly with respect to sophisticated insureds, is tempered, however, by two co-existing alternative forms of insurance regulation: captive regulation and surplus lines regulation. So-called captive insurers have their access restricted to relatively sophisticated policyholders capable of banding together to monitor the insurer. With some exceptions, they are generally regulated only by one state even when policies are sold outside that state. Surplus lines insurers, instead of being regulated by all states into which the insurer is admitted to do business, likewise have their solvency seriously regulated by only one of the states. With surplus lines insurers, however, states in which these insurers do business without a license may nonetheless regulate their market conduct. Moreover, most states make their own regulated surplus lines brokers responsible for appropriate investigations and disclosures with respect to the solvency of the surplus lines insurers. These alternative modes of insurance regulation in which a home state has greater responsibility for solvency regulation of the insurer, resembles, in certain respects, that applying in the European Community (Hogan, 1995).

2. Recent Economic Analysis of Solvency Regulation

Economic scholarship regarding the laws governing insurance regulation is divided into two categories: studies of rules relating premium revenue to surplus and studies of price regulation.

2.1 Rules Relating Premium Revenue to Surplus

Either through the so-called Kenney Rule or its more modern equivalent via the ‘IRIS’ system of insurance solvency monitoring used in much of the United States, most jurisdictions attempt to enhance solvency by restricting the ratio between the volume of premiums an insurer can write in some given period and
its stock of capital surplus. The notion is that restrictions on the rate at which a financial intermediary such as an insurer can engage in new borrowing decreases the likelihood of insolvency and preserves the ability of monitoring agencies to provide early warning of problems (Mayerson, 1969). Several scholars have challenged this implementation of solvency regulation by arguing that it creates or exacerbates positive feedback cycles in insurance pricing and availability (Winter, 1988, 1991b; Gron, 1994a, 1994b). These instabilities hamper the economy.

The Winter model assumes that this method of regulation creates a topology known as a 'fold catastrophe'. Winter argues that the regulatory constraint establishes a peculiar short-run supply curve for insurance which, when coupled with a relatively inelastic demand curve for insurance, can establish a dynamic relationship between surplus and insurance price. This dynamic process, while often equilibrating insurance markets in a relatively smooth fashion, can also lead to almost instantaneous increases in insurance price and a correlative decrease in insurance supplied - a phenomenon last observed in the late 1970s and early 1980s.

2.2 Price Regulation

Other scholars have considered the economic effects of government price regulation. This regulation may occur either through the direct setting of a price for various forms of insurance or by constraining the prices an insurer may charge to some range. By requiring that rates be 'adequate', insurers are protected from 'trembling hand' instabilities (Selten, 1975) in the system whereby the perhaps irrational decision of one insurer to underprice insurance can drive out more rational insurers and lead to the collapse of an insurance system (Kimball, 1992). While history suggests that fear of insurance underpricing is not irrational, the ability of pricing regulation to combat insolvency is questionable. If insurers compete, suppression of price competition merely bolsters non-price competition in areas such as marketing (Stigler, 1968). It lowers consumer welfare because consumers no longer get the product they desire most. The loss to the United States economy from this non-price competition was estimated in 1980 at $1 billion (Frech and Sambrone, 1980). Moreover, to the extent the insurers do not compete, price regulation based on assuring an adequate return may deter efficient methods of production by insurers.

Price regulation is also troublesome to the extent that permitted rates do not encompass those which would clear the market. If permitted rates fall below the market clearing rate, insurance shortages may exist that harm insureds and insurers alike (Winter, 1991a). While the existence of legal 'black markets' such as the surplus lines market and the captive market may reduce the impact of the distortions and hardships created thereby, the transaction costs associated with purchase of these sorts of policies, coupled with their limited availability,
renders these alternatives at best only a partial cure. If permitted rates stay atop the market clearing rate, a deadweight loss occurs resulting from a reduction in profitable transfers of risk. Insurers, under such a system, obtain economic rents. Costs incurred by insurer in obtaining or preserving such a hospitable regulatory state may also constitute economic waste.

C. Underwriting Regulation

3. Adverse Selection with Costly Sorting

Adverse selection, the tendency of persons with private and accurate beliefs about their high risk to transfer risk more extensively than persons with low risk or without private and accurate beliefs, is the central economic problem of insurance contracting. Unchecked adverse selection can destroy an insurance market and the benefit conferred thereby (Akerlof, 1970). The law has thus developed various mechanisms to support ‘underwriting’, the process of curbing the ‘asymmetric information’ between insured and insurer that causes adverse selection. Doctrines such as representation and warranty have been central in this effort. These mechanisms have long come in occasional tension with concerns regarding insurer power and the misallocation of risk that results when these doctrines are interpreted by imperfect courts. These tensions have led only to adjustments to the central legal doctrines. In recent years, however, it has become increasingly common for laws in support of private adverse selection control to collide with other social goals such as wealth redistribution or antidiscrimination entitlements. Accordingly, the law of underwriting has, in recent years, undergone significant change.

As shown by Borenstein (1989), Rea (1992) and others, when individuals with different expected losses are ‘pooled’ and charged the identical premium, the result is a loss that depends on the elasticity of demand for insurance. In general, however, the loss results from (1) the consumption of insurance by high-risk insureds beyond the point where the marginal value they attach to risk transfer equals the marginal cost of their risk transfer and (2) the consumption of insurance by low-risk insureds in an amount less than that where the marginal value they attach to risk transfer equals the marginal cost of their risk transfer. Although, as noted by Rea (1992) and Chamberlain (1985), this efficiency loss from pooling should perhaps be offset by the gain that pooling - which itself amounts to insurance of a particular ‘commodity’ price - creates for those who, at some earlier position (Harsanyi, 1967) behind a ‘veil of ignorance’ (Rawls, 1971), do not know their risk class and are unable to adjust measured indicia of their expected loss.
Separation of otherwise pooled classes is generally considered to be efficient when the cost of the separating program is less than the gains resulting from enhanced allocative efficiency through pricing that more closely reflects the marginal costs of risk transfer posed by each potential insured. Separation thus becomes more sensible as the (1) cost of the developing and implementing the separation algorithm decreases; (2) the homogeneity of expected losses posed by the separated risk classes increases; and (3) the elasticity of demand for insurance increases. Reduction of the cost of a separation program in turn depends on (a) technological advances in testing, (b) reduction of variable costs by administration of binary tests only to those likely to be in the numerically smaller of the two separated classes, and (c) legal rules supporting separation.

As also shown by Schmalensee (1984), Borenstein (1989) and Rea (1992), however, low-risk individuals and insurers may develop a separation program even when these efficiency conditions are not met. The low-risk insureds, this argument contends, fail to take into account the interests of the high-risk insureds in not implementing a testing program that, while it restores allocative efficiency, ends the cross-subsidization they enjoyed under pooling. Coasean bargaining by the high-risk insureds to pay off the low-risk insureds not to be tested is even more difficult than usual here, because, almost by definition, the high-risk insureds do not know the identity of the low-risk insureds. The unregulated market equilibrium in favor of separation is also troubling, as suggested by Abraham (1986), because it favors separation techniques that examine readily ascertainable characteristics such as race or gender that may have no causal relationship to expected loss but that bear a statistical correlation with the difficult-to-observe factors that indeed determine loss.

In theory, the unregulated market equilibrium described above operates by establishing different pricing terms for identical contracts and achieves separation through verification mechanisms that reliability and validly classify individuals by risk. A second mechanism described first by Rothschild and Stiglitz (1976) and explicated by various scholars (Rasmusen, 1994) achieves separation by having insurers offer a menu of contracts with varying shortfalls in the indemnity to be paid - and commensurately lower prices. This menu of contracts induces only low-risk insureds (who know themselves to be low risk) to purchase the most incomplete contracts. The mechanism saves the cost of verification but results in inefficiently low levels of risk transfer to the low-risk insureds.

4. Classification of Legal Responses to the Problem of Adverse Selection

Legal responses to the tensions between adverse selection control and other goals, such as non-discrimination, differ in the way that they control the
problem of ‘error’ and in the zones of legality they establish according to classification cost and classification benefit. One prevailing response is ‘libertarian’, to give the insurer and insured complete freedom to use whatever classification devices they choose. Perhaps hoping that the market will punish inaccurate classifiers, the government itself makes no effort to control error. The government likewise fails to evaluate whether the classification method employed is inefficient, either classifying too coarsely or too finely. In some ways, this response, which at least prevents the possibility of government-created error in classification, has been the prevailing legal response in the United States and elsewhere for many years.

Laws that focus primarily on error control represent a second response mode to the tension between adverse selection control and other goals. Laws, for example, that prohibit classification for automobile insurance based on marital status of the driver or that prohibit different health insurance charges for disabled individuals absent ‘sound actuarial principles’ for doing so exemplify this sort of law. Such laws act as a safeguard against the cooperative behavior generally permitted in the insurance industry by prohibiting ‘irrational’ collective discrimination by insurers against disfavored groups. They fail, however, either to require classification in this way or, apparently, to prohibit overly coarse classification in which separated groups have apparently different expected losses but in which that difference is actually caused by an underlying variable that greater classification expense would have revealed. Such laws likewise generally fail to criticize insurance classifications that result in reasonably accurate separation of applicants into groups with equivalent expected losses but that do so too finely through testing methods that are so costly as to make the whole classification enterprise inefficient.

A third response mode makes classification impermissible, except when it is extremely fine, even where the cost of fine separation exceeds the economic benefit. Laws generally prohibiting use of gender as a basis for classifying individuals for life insurance or annuity policies exemplify this mode of response (Arizona Governing Committee v. Norris, 1983). In the western world, gender correlates with life expectancy and it is unclear whether many other underlying variables that determine expected life expectancy may be ascertained at any reasonable cost. Gender is nonetheless sometimes prohibited as a classification device. Laws of this sort protect both against ‘irrational’ collective discrimination but also against the ‘accident’ of certain variables that correlate with expected losses being cheaper to observe that others that actually determine expected loss. They also may serve as a safeguard against the perpetuation of irrational collective prejudices whose existence is difficult to prove.

A fourth response (‘non-actuarialism’) is to prohibit classification even if it is efficient and even if it is accurate. Laws barring use of race for classifying
Insurance applicants for life insurance and annuities exemplify such laws. Such laws redistribute wealth according to political and social preferences apart from notions of economic efficiency.

Interestingly, laws that implement the ‘efficiency concept’ described above - do not appear yet to exist, although, as noted below, restrictions on use of genetic information may fall into this category of regulation. This absence of ‘efficient actuarialism’ from the law as it actually exists may in part be due to the extraordinarily high legal and administrative costs that would be incurred in determining accurately whether any particular classification system met the efficiency standard.

5. Traditional Regulation via Legal Doctrines of Misrepresentation and Warranty

As part of a libertarian impulse, the law has long facilitated separation and the economic benefits it often entails by permitting the insurer to investigate the risk characteristics of a potential insured even where this investigation probes generally private information about the insured. Moreover, subject to recent restraints on use of cheap observables such as race or gender, the law has not intervened in insurer’s selection of characteristics used to determine expected loss, even where this selection - often done in concert - causes substantial loss to identifiable social groups.

Pre-contractual investigations of all potential insureds can, in many instances, be quite costly. The doctrines of warranty and representation represent potentially cheaper vehicles to create symmetric information and thereby reduce adverse selection. Under both doctrines, the insured is given an opportunity prior to contract formation to make statements that, if true, demonstrate a lower risk and thereby lower premiums in the event a contract is formed. The insured accepts some penalty in the event that some subset of otherwise insurable events come to pass and the statements made prior to contract formation turn out to have been false to some pre-specified extent as determined by some post-event investigation.

The economic concept behind both warranty and representation is to restructure the ‘game’ being played by the insured and insurer into one where the insured’s dominant strategy is to reveal truthfully all private information about his risk and therefore to render the information symmetrical. The success of this concept becomes more likely as (a) the amount of the potential penalty increases, (b) the scope of the subset of events triggering the potential penalty increases, (c) the degree of falsity required to trigger the penalty decreases; and (d) the likelihood that falsity will be detected increases. The joint benefit to both insurer and insured increases as the expected cost of this post-event
determination decreases. Measurements of this expected cost should reflect both the cost of the investigation itself and the risk cost imposed on the insured as the result of a determination that is either incorrect or that occurs absent truly asymmetric belief structures.

The traditional doctrine of warranty is a structure that results in strong inducements to tell the truth but creates substantial costs. Under a polar version of the traditional doctrine, the penalty (defeasance of an otherwise existing insurance obligation) is exacted whenever any claim for insurance benefits arises and any statement made by the insured is shown to have been false to any degree. Thus, with warranty, a claim for insurance is defeated by the making of even minimally false statements even when the complete truth would not have lowered the premium for a policy providing indemnity only for the event that actually occurred. Because warranty bars claims even when the insured cannot be shown to have known that the statement was false, the doctrine protects the insurer (and, in a competitive market, insureds) from statements that the insured in fact knew to be false in conditions where the insurer cannot prove that knowledge. The downsides of the warranty mechanism include the frequent investigations into the truthfulness of pre-contractual statements and the undesirable transfer of risk to the insured in situations where belief structures were indeed symmetric but the insured was careless in his pre-contractual statements.

The doctrine of representation represents another method of controlling adverse selection. Although many variants of the doctrine exist, it may generally be idealized as a rule under which the insured is penalized (defeasance of an otherwise existing insurance obligation) whenever the insured knowingly made a false statement, a claim for insurance benefits arise, and the premium for a policy providing indemnity for the event that in fact occurred would have been materially higher had the insured been truthful. Since fewer events trigger an inquiry into the truthfulness of pre-contractual statements and since exaction of the penalty depends on the insured’s being shown to have knowingly lied, the expected penalty for false statements is less than that under traditional warranty. The incentive to be truthful and the correlative reduction in adverse selection may be tempered as well. On the other hand, representation rather than warranty lowers expected investigation costs and substantially reduces the chance of imposing a penalty on the insured in situations where beliefs were truly symmetric.

During the early nineteenth century, many American and English courts saw considerable virtue in a legal regime where warranties were unfettered in their ability to control adverse selection (Horwitz, 1992). In recent years, however, legal authorities, apparently believing representation either to be a superior vehicle for controlling adverse selection or in an effort at protecting insureds from the costs of warranty, have generally imposed a preference for
representation, at least in situations where the legal status of pre-contractual statements has been ambiguous. Some legislatures have thus enacted statutes that deem statements to be representations absent clear statements to the contrary. Others have enacted ‘contribute to cause’ statutes that permit penalization of the insured only where the premium for a policy providing indemnity for the event that in fact occurred would have been materially higher had the insured been truthful. Few jurisdictions have considered a contribute to cause standard coupled with a significant penalty (beyond mere defeasance of otherwise existing contractual obligations) for discovery of a material falsehood, although such a mechanism with low policing costs and high penalties would fit well with economic theory regarding control of illegal behavior (Polinsky and Shavell, 1979).

Also reducing the efficacy of the pre-contractual statement mechanism, though arguably decreasing its undesirable side effects, has been the doctrine of incontestability. Incontestability, which acts as a kind of statute of limitations against insurer defenses of misrepresentation of breach of warranty, generally addresses situations where there is a high risk of an erroneous determination that a pre-contractual statement was false: the insured is dead and thus unable to testify regarding knowledge of falsity or falsity itself against a life insurer seeking to avoid an indemnity obligation. Thus, at least for the subset of risk factors that the insurer could reasonably have determined by an investigation either before the contract or during some period of time (1 or 2 years) after formation of the contract, the law refuses to let post-event investigations result in a penalty to the insured (Simpson v. Phoenix Mutual Life Insurance Company, 1969).

Laws reducing the penalty for false statements likewise potentially weaken the use of post-event investigations as an underwriting tools for controlling adverse selection. These laws reduce, however, the costs associated with erroneous determinations of falsity and determinations of falsity in the face of symmetric beliefs. Legislative distrust of the market’s determination of optimal methods of adverse selection control may explain statutes limiting the penalty for false declarations of age on a life insurance policy to a reduction in the indemnity equal to the difference between the face amount of the policy and the amount of insurance that could have been purchased for the same premium had the insured been truthful about age. Similar judicial distrust of the market may explain occasional rulings limiting the penalty for false representations to a reduction in indemnity equal to that which the insured would have been entitled to had it told the truth in the application process.
6. Genetic Testing

Legal support for both investigatory and post-event methods of separation has recently collided with the growing interest in acquisition of information regarding the genetic make-up of individuals. The result has been the emergence of new laws or ‘voluntary’ agreements on the part of insurers restricting use of genetic information. The Health Insurance Portability and Accountability Act of the United States, laws recently passed in several American states, and laws enacted in Austria and Belgium, all restrict the ability of insurers either to investigate genetic information or to rely on traditional doctrines such as representation with respect to genetic information. Indeed, the Belgian law would appear to prohibit even voluntary disclosure of favorable genetic information. Insurers in Australia, France, the Netherlands, Norway and Switzerland have restricted use of genetic information through industry trade associations.

Most scholars who have examined ‘genetic information privileging laws’ from an economic perspective have been critical of these efforts. They fear it will lead to serious problems of adverse selection in health and life insurance (Christianson, 1996; Pokorski, 1995, 1997). Those with privileged information of a genetic defect that increases the probability of early death but without other symptomatic manifestation might, for example, purchase large amounts of life insurance, knowing it to be a tremendous investment under those circumstances. The investment could be liquidated prior to death through viatical settlements and accelerated death benefit provisions now becoming more prevalent within insurance policies. Similarly, those with knowledge of a genetic predisposition to debilitating illnesses such as Alzheimer’s disease might stock up on long-term disability insurance.

Others, however, have suggested that permitting insurance underwriting to deter genetic testing may lead to delayed diagnosis and treatment of disease. The harm caused thereby outweighs dislocations to the insurance market caused by privileging genetic information (McGoodwin, 1996). The economic component of this argument rests, however, on the value to the insured (and perhaps related parties) regarding foreknowledge of genetic defect. Moreover, this argument for privileging genetic information has greater weight with respect to health insurance and possibly long-term care insurance than it does for life insurance. With respect to the latter, it is hard to develop a traditional economic argument supporting genetic privileging laws, except perhaps one that sees the prohibition as the solution to a collective action problem under which individuals with ‘good genes’ would benefit from testing by costly methods even when the harm done to those with bad genes, coupled with the cost of testing, outweighed the total benefit created by the classification scheme.
Still others (Rothstein, 1993, 1997; Stone, 1996; Jacobi, 1997) have argued that genetic testing poses a perilous dilemma for a system that relies heavily on private risk-based insurance for basic health insurance. If genetic tests are not privileged, few will take them and those who do may find themselves denied financial access to good health care. If genetic tests are not privileged, the private insurance market is seriously endangered. An intriguing escape from the dilemma is suggested by one scholar, however, who proposes to bar individuals from undergoing genetic tests absent prior purchase of insurance to indemnify the insured for the higher health or life insurance premiums that would result from a finding of defect (Tabarrok, 1994). Such a system might itself be subject to some adverse selection, however. People rationally suspecting a genetic defect based either on family history or testing not covered by the law (such as testing in a foreign nation), would purchase this bundle of genetic testing and ‘insurance premium insurance’ with greater frequency than those without suspicion.

D. Contract Regulation

7. Introduction

Two significant themes pervade regulation of insurance conflicts: the first is the extent to which government intervention is desirable to protect the insured from the presumed power of the insurer; a second is the extent to which government intervention is desirable to protect non-parties to the transaction from the coalition created by insurer and insured.

8. Insured Protection

8.1 Contra Proferentum

Construing ambiguous contracts against the drafter, usually the insurer, has long been part of American insurance law. One economic theory for understanding this doctrine of ‘contra proferentum’ sees it as beneficially internalizing the social costs fostered by ambiguity. Imposing a ‘penalty’ for ambiguous drafting in the form of expanding the enforceable promises made by the drafter or contracting the conditions protecting the drafter theoretically induces the drafter to avoid ambiguity. An overlapping economic theory sees contra proferentum as correcting information problems in the insurance market: interpreting ambiguities in the contract in favor of the insured provides the insured with the optimal expected coverage; the insurer’s foreknowledge of
the doctrine insures that the insured pays the correct premium for the expanded coverage (Abraham, 1996).

While the theory behind contra proferentum, so stated, may not be particularly controversial, its application in sculpting existing insurance law is. To begin with, if the non-drafting party can understand the terms of the contract without incurring significant cost, as is the case when the insured is sophisticated or a ‘repeat player’ in the transactional form, the assignment of liability to the drafting party appears arbitrary. Indeed, under such circumstances the labeling of one party as ‘the drafter’ appears arbitrary as well. Many courts nonetheless hold the doctrine still to apply in this circumstance (Ostrager and Newman, 1995). This arbitrariness might be tempered, to be sure, if the parties were costlessly able to reallocate the private burdens of ambiguity through the insurance contract itself, but impasse costs, bargaining costs, and the risk that a court would not recognize the reallocation of risk would hinder such a cure (Rappaport, 1995).

Second, while the burdens of ambiguity fall substantially on society in the form of higher dispute resolution costs, the benefits of the doctrine accrue largely to the non-drafting party (at least in the case of ambiguity unrecognized by the drafter). The doctrine thus creates a disincentive for non-drafting parties to become educated and, indeed, creates a possible incentive for non-drafting parties to prefer contracts that have unrecognized ambiguities. Moreover, some have argued that the doctrine of contra proferentum discourages innovative terms in contract drafting because new terms are subject to the doctrine whereas old terms, already having been interpreted by the courts, are better understood (Goetz and Scott, 1985). The social costs of ambiguity - more judicial disputes - are thus not reduced as much as might initially be thought by the existing judicial doctrine of contra proferentum. Examination of contract terms by government officials before litigation arises, though creating its own difficulties, may thus have some place in an economic theory of insurance regulation.

Scholars have also profitably used the tools of economics to classify the variants of contra proferentum found in American courts. Professor Abraham, for example, drawing on traditional law and economics scholarship, describes the optimal degree of clarity as one that would minimize the sum of maldrafting costs (uncertainty, unwarranted reliance) plus maldrafting avoidance costs (lawyer time and lengthier policies, Rappaport, 1995; Abraham, 1996). Negligence-like regimes are those in which contra proferentum is invoked only where the degree of clarity created by the drafter falls short of the optimal level. Strict liability regimes are those in which contra proferentum is invoked whenever an ambiguity is found. Drawing on recent scholarship regarding contract default rules (Ayres and Gertner, 1989), Abraham further classifies legal regimes regarding contra proferentum as penalty or majoritarian-oriented. ‘Majoritarian’ approaches alter the terms of a contract deemed ambiguous (by one of the above measures) to that which a
A significant development in American insurance law has been the willingness of courts and legislatures to grant insureds injured by breach of an insurance contract damages in excess of that which would have been paid by the insurer had it honored its contractual obligations. These ‘extra-contractual damages’ have been controversial since their inception, have been the subject of some retrenchment in recent years (Macintosh, 1994), and have the subject of considerable economic analysis.

There are two sorts of arguments made for augmenting damages for breach of an insurance contract. One, not the subject of examination in this entry, rests on the general inadequacy of traditional contract damages and applies those grounds in the insurance arena (Sebert, 1986). A second type of argument justifying extracontractual damages is that otherwise the winning strategy for insurers is frequent ‘opportunistic breach’ with respect to meritorious claims. This strategy is particularly effective against those who show no signs of sophistication and with respect to smaller claims (Abraham, 1986).

Economic arguments in favor of extracontractual damages generally rest on forms of information failure. Unsophisticated insureds, often ignorant of their actual rights under the policy, particularly as expanded by doctrines such as contra proferentum, fear that if they sue and lose they will be out significant time and money. Attorneys are unlikely to correct any misinformation because insureds have imperfect access to the legal marketplace even for advice (Curtis, 1986). Moreover, even insureds accessing the legal market are likely to stall there because attorneys will have difficulty cheaply determining the merits of the insured’s case, particularly where the insurer has failed to create a detectable pattern signaling wrongful denials. And even an attorney determining that the case has merit will be unlikely to bring it because of the small rewards relative to the cost of the case. The reward is small because the insurer pays only what it owed in the first place, plus any positive difference between court-system interest rates and the money earned by holding on to the money, plus some attorneys fees, possibly including those of the insured (Jerry, 1986).

According to this theory, the ‘equilibrium’ under a legal regime denying extracontractual damages is for the insurer to breach frequently and the insured frequently to do nothing about it. The consequence is lesser entry into insurance...
transactions and less risk transfer than desirable. Extracontractual damages thus creates a Pareto superior contract as a default rule (Perlstein, 1992).

Against this theory rests a basic question: if extracontractual damages are so superior why do they not exist in a competitive insurance market absent legal intervention? If insureds operating in a competitive insurance market really felt it were necessary and desirable to have extracontractual damages, some sort of mechanism would be built into the contract, such as a formula damage clause for some subset of failures by the insurer timely to pay a meritorious claim. Yet such a system was seldom if ever seen. In part, economic critics of extracontractual damages say, this absence may exist because the harmful reputation an insurer would acquire through consistent bad faith denials of a claim would lead to market punishment at least as harsh as any the legal system could provide. And even if the market failed, public regulators and licensure mechanisms could succeed. It may not exist because insurers and insureds know that the legal system would be unable to distinguish ‘bad faith’ breach from simple mistake. This inability would lead to precisely what has occurred: perversions of desired coverages, induced breach (Richmond, 1994; Schmidt, 1994), and needlessly higher premiums for insureds. And it may be absent in part because the legal market functions better than proponents of extracontractual damages assert (Sykes, 1994a). Awards of legal fees to prevailing parties substantially correct the attorney incentive problems without resort to radical reformulation of contract law. The ability of some insureds to ‘sell’ their rights against insurers in a competitive market to others whose need for funds is less immediate may sometimes hinder the ability insurers would otherwise possess to force cheap settlements out of insureds. Finally, some economic scholars have argued that (ex ante) insureds subject to suffering a catastrophic loss of an irreparable commodity want less than full compensation for their financial loss rather than the extra compensation provided by much current law on extra-contractual damages (Cook and Graham, 1977; Fenn, 1987).

This attack on extracontractual damages is not without problems. Deterrence of opportunistic breach through reputational injury rests on the dubious ability of the market accurately to distinguish breach from lawful claims denial. Public regulators are often underfunded, incompetent or captured by insurers. The competitive market that supposedly exists in fact consists of insurers who may lawfully collude. An award of simple legal fees to prevailing insureds may not entice attorneys uncertain of their clients’ prospects. The supposed cure of the insurer’s ability to exploit the weakness of an insured through assignment of claims relies on markets that seldom exist precisely because such transactions are often difficult and, in some instances, illegal (State Farm Fire and Casualty Co. v. Gandy, 1996).
8.3 Conflicts of Interest in Liability Insurance Policies

Protection of the insured is likewise illustrated by judge-made law in the United States governing conflicts of interest created by the contract between a liability insurer and its insured over private settlement of legal actions brought against the insured. The conflict, which resembles that arising in the corporate context between various forms of claimants against corporate assets, arises from the extraordinarily complex decomposition of responsibility liability insurance policies almost always create.

The responsibilities created by typical liability insurance policies in conjunction with legal doctrine may be crudely represented by positions along three not entirely orthogonal axes. One axis distinguishes responsibility for ‘discharge’ payments such as settlements and judgments made to reduce the liability of the insured from ‘defense’ payments made on behalf of the insured in defending against claims of liability. Certain discharge payments such as those that ‘exhaust’ certain limits of the policy may eliminate future defense obligations of the insurer. This feature of conventional liability insurance contracts effectively gives the insurer a ‘put’ on future defense obligations provided it can completely discharge its insured’s liability to one potential victim of the insured. Thus, a conflict occasionally arises between insurers, who want to exercise this put (at a price negotiated with a victim of the insured) and an insured, who does not want to take on defense responsibilities even when the insurer’s failure to exercise the put increases at least one insured’s expected payments in discharging liabilities to various victims (Texas Farmers Ins. Co. v. Soriano, 1994). The law is not yet well developed on whether the insurer will be permitted unfettered discretion in exercising its apparent contractual authority in this area.

One axis that creates frequent conflicts distinguishes responsibility for ‘discharge’ payments made to reduce the liability of the insured based on the amount of the payment in question. Often, the insured takes the first level of responsibility pursuant to deductible or ‘self-insured retention’ provisions in its policy. A primary insurer will often take the next level of responsibility up to some policy limit, but with payments thereunder reducing its responsibility for certain other judgments against the insured and, potentially, litigation expenses in other cases. An excess insurer may take the next level of responsibility up to some policy limit, with the insured having responsibility for judgments beyond the final layer of excess. All of these responsibilities can be shifted in whole or in part to reinsurers. To the extent that litigation strategies, such as reactions to settlement offers, can shape the probability distribution of future discharge and defense obligations, which distribution in turn determines the expected loss from the case suffered by each party, the decomposition along this axis creates conflicts of interest among the various parties regarding litigation strategy (Chandler, 1993; Sykes, 1994a, 1994b; Silver and Syverud, 1995; Pryor, 1997).
In the absence of clear language to the contrary in the insurance contract, courts typically react to this conflict by simplifying the decomposition in a way somewhat analogous to a judicially-imposed corporate restructuring. In the face of an opportunity to settle for some ‘reasonable amount’ with the plaintiff, the insurer that fails to tender the requisite amount or who acts in bad faith with respect to settlement of the case, is held to ‘own’ all layers above the one it nominally owns by contract. The faulty layer is thus liable directly to the victim for judgments piercing that layer or, through subrogation and other mechanisms, to higher layers who discharged the responsibilities that would have been theirs absent the fault of a lower layer (Stuhr, 1992). And, in most American jurisdictions, this liability exists even where the higher layer would not actually have the resources to pay the judgment.

Economic theory might well view this restructuring as one vehicle (among others) for correcting inefficient behavior through internalizing of the costs created when parties in control of litigation strategy are otherwise able to foist the costs of certain strategies on to others. The restructuring often induces insurers with settlement responsibility to make decisions that maximize the joint wealth of insurer and insured, thereby leading to optimal consumption of liability insurance. The problem, as noted by two scholars (Sykes, 1994a; Logue, 1994), occurs when the higher layer (usually the insured) would not have the financial resources to pay the judgment. Under these circumstances, the conventional legal rule induces settlements that fail to maximize joint wealth, which in turn leads to underconsumption of liability insurance, which in turn leads to undercompensation of victims. A minority legal doctrine has recently re-emerged that limits the liability of the faulty layer to no more than the higher layers would actually have been able to pay. While this ‘Michigan rule’ may cure the distortions in the settlement process caused by the traditional legal decomposition of responsibility, it may create other problems such as undercompensation of victims and underdeterrence of wrongful conduct. Whatever the merits of the various rules, however, there is always the theoretical possibility of Coasean bargaining to restore efficiency (Sykes, 1994b) but it is unlikely to be particularly useful here given (a) the high costs of bargaining in this extraordinarily complex field and (b) a likely judicial reluctance to accept private party modification of legal rules.

Interestingly, consonant with economic theory, liability insurance contracts tend not to grant the insurer authority to settle cases without trial absent the consent of the insured when the insured’s reputation and future income depend on exoneration from the allegations made in the lawsuit (Mayers and Smith, 1981). Courts, concerned about the major change in expected discharge responsibilities created by such a provision, generally refuse, however, to imply it into contracts otherwise silent on the matter (Dear v. Scottsdale Insurance Company, 1997; McKinley, Moody and Barlow, 1997). Insurers and insureds
protect against exploitation of this consent clause by frequently providing that a settlement held up by the insured’s non-consent re-layers discharge responsibilities such that the insured takes on responsibility for discharge payments in excess of the amount of the rejected settlement (Rauch et al., 1986).

A third axis in which conflicts of interest arises surrounds the liability insurance contract’s decomposition of responsibility for various types of events. The insurer, for example, generally has discharge obligations arising out of liability from negligence but generally does not have discharge obligations arising out of more serious types of fault. This conflict leads to an intricate ballet of litigation strategies and counterstrategies described well by one recent scholar (Pryor, 1997). In essence, however, the injured party may ‘underlitigate’ the dispute in an effort both to induce the insurer to incur heightened expected defense and discharge obligations, which increase may be partly appropriated by the injured party through offering to withdraw the triggering underlying dispute against the insured in exchange for a payment mostly from the insurer.

9. Public Protection

9.1 Anti-Diminution
In the past, some liability or indemnity insurers sought to reduce the amount they paid victims of their insureds based on the insolvency of their insureds. The insurers’ theory was that their own liability derived from actual payment or the possibility of payment by the insured, which possibility was reduced by the insured’s insolvency. ‘Anti-diminution laws’, or ‘bankruptcy provisions’ now frequently prohibit this practice (Jerry, 1987) and are a classic example of insurance regulation acting to the detriment of the insured and in favor of the public at large. Assuming with Huberman that an insured’s utility of wealth is constant below some ‘insolvency’ level, but that the indemnity owed as a function of insured loss does not have to be non-decreasing, an insured would prefer a policy whose indemnity obligation disappeared above some threshold level of loss to a policy that had some constant upper limit regardless of the amount of the loss. Indemnity obligations on the part of the insurer that fail to increase the utility of the insured in the event of some serious adverse increase needlessly decrease wealth in all other states by increasing the actuarially justified premium owed by the insured for the insurance. And, indeed, during the early years of indemnity/liability insurance, such ‘diminution clauses’ existed in insurance policies.

The problem with such ‘diminution clauses’ is that they result in inadequate compensation to victims of the insured and exacerbate the moral hazard problem caused by the existence of insurance (Shavell, 1986). Many American states now require that insurance policies not diminish on account of the
insolvency of the insured.

9.2 Treatment of Proceeds in Insolvency
The conflict between the interests of the insured and those of potential stranger victims of the insured shows itself again in the legal treatment of liability insurance proceeds upon the insolvency of the insured. Treating these proceeds as property of the estate and thus available to all creditors of the insured, including contract creditors, has the potential generally to enrich the insured by permitting it to borrow at lower interest rates, which inures to the benefit of the insured in the event a subsequent insolvency does not occur. Treating the proceeds as dedicated to the victims of conduct covered by the insurance policy, as the law now often does, means that uncovered claimants such as contract creditors or victims of tortious conduct excluded from the insurance policy’s indemnity obligations must look solely to the insured’s non-insurance assets in the event of an insolvency. This treatment thus has the potential to increase the interest rate the insured must pay. This result likewise may occur where the proceeds of the policy are nominally part of the insured’s estate but where the law permits victims of conduct covered by the insurance policy to ‘cut through’ and bargain with the insurer for a direct payment in exchange for a reduction of the claim against the insured.

9.3 Extreme Conduct Limitations
Government limitations on the ability of insureds to protect themselves against liability either for their intentional misconduct to others or for conduct deemed appalling or sadistic enough to be worthy of ‘punitive’ or ‘exemplary’ damages further illustrates intervention in the market to protect potential victims of the insured’s and insurer’s cooperation (Chandler, 1996). Generally, these limitations restrict coverage for appalling conduct. Other times, however, notably in the area of automobile insurance, government requires coverage for appalling conduct so as not to defeat a ‘compulsory insurance’ regime (State Farm Fire and Casualty Co. v. Tringali, 1982).

Many commentators applaud limitations on the insurability of punitive damages or damages for intentional conduct on grounds that such restrictions prevent undesirable moral hazard (Ellis, 1982). These efforts generally fail to confront the questionable need for such governmentally imposed constraints, however, given the ability of insurers who fear moral hazard and who will pay the price for failure to control it on their own through exclusions and conditions in the insurance contract. One explanation is that the refusal to enforce contracts permitting indemnification for severely deviant conduct lets insurers and insureds, apparently confused at the time of contracting, ‘come to their senses’. Such a legally created ‘reprieve’ from a faulty contract may be particularly important given the possibility that indemnity for punitive damages
will increase the frequency of wrongful behavior by insureds and the frequent
inability of the law, even with punitive damages, to restore their victims to the
status quo ex ante. This argument explanation draws support from simulations
conducted by one scholar suggesting that few rational contracting parties would
want to provide indemnity for conduct grossly below an efficient level of care,
even with court systems prone to error (Chandler, 1996). Perhaps the simplest
explanation made with respect to the non-insurability of punitive damages is
that deterrence models are irrelevant to punitive damages and that punishment
is destroyed where punishment is indemnified (Ellis, 1982).

E. Competition Law Exemptions

In the United States and in Europe, insurers have been exempt from many
‘competition’ or ‘antitrust’ laws prohibiting forms of cooperative behavior by
distinct insuring entities. In the United States, this exemption occurs via the
McCarran-Ferguson Act (15 United States Code § 1011-15) insulating insurers
from most of federal antitrust laws, and by ‘mini McCarran-Ferguson Acts’
enacted by some states insulating insurers to varying degrees from the states’
own laws barring combinations and conspiracies in restraint of trade. In
Europe, it occurs via Commission Regulation promulgated in 1992 under
Article 85(3) of the Treaty of Rome. This regulation generally permits certain
cooperative practices in the ‘insurance sector’ such as cooperation with respect
to establishment of premiums, standard policy forms and coverages that might
otherwise be prohibited by Article 85(1) of the Treaty and the regulations
thereunder.

An economic theory has emerged to justify this departure from the general
notion that competition rather than cooperation maximizes welfare: the
significant economies of scale in production of information needed to prevent
adverse selection and insolvency (Achampong, 1991; Danzon, 1992; Macey
and Miller, 1993b). Without significant data on the experience of policyholders,
it becomes difficult reliably to model the expected loss posed by individual
policyholders and therefore accurately to establish actuarially sound premiums.
Individual insurers, even large insurers operating multiple lines in many
jurisdictions, may lack adequate internal historic data to perform the necessary
analyses with confidence. Some have argued that this inability extends even to
projections of future losses on future policies. Collective analysis on this point
should thus be exempt from antitrust scrutiny (Danzon, 1992).

Advocacy of cooperation rather than competition in the business of
insurance extends to collective drafting of policy forms. Without common
forms, the data needed to make reliable predictions would be ‘noisy’ and thus
less amenable to accurate, fine-grained statistical analysis (Macey and Miller,
Consumers are thought better able to shop when policies have only a few varying parameters, such as price. Uniformity in policy provisions thus actually favors competition, this argument proceeds. Moreover, reinsurance, which serves an economically beneficial specialization function, would be hampered were collective policy forms barred and certain lines of insurance made more custom and less of a ‘commodity’ (Ayres and Siegelman, 1989).

Two other economic arguments also frequent debates about application of competition laws to the insurance industry. First, scholars such as Bork (1978) generally question whether collusive behavior between entities should ever be illegal so long as the law would permit those entities to merge. Limited cooperation between competitors might well be better than the elimination of competition through merger (Bork, 1978). Exemption from competition laws thus fosters the existence of numerous small insurers and, derivatively, competition (Danzon, 1992). Second, and somewhat relatedly, some scholars have suggested that legally supported cooperation helps prevent ‘cut-throat’ or ‘anarchic’ competition that, at least historically, appears to have jeopardized the solvency of insurers (Carlson, 1979; McDowell, 1987).

Several scholars have noted, however, the scope of the American exemption from competition laws is considerably broader than the arguable economic justification (Achampong, 1991; Macey and Miller, 1993a) Collective data gathering would almost certainly be protected by modern interpretations of existing American competition laws, even were the McCarran-Ferguson Act not in place. Moreover, a need for collective data gathering would not seem to justify collective data analysis, nor would it justify collective pricing of insurance policies, both of which are frequently permitted in the United States though practiced somewhat less frequently than in years past. Moreover, if there are economies of scale in data collection, it is not clear why for insurance but not for other industries the law would ‘prop up’ small companies operating at less than an efficient scale by permitting them to collude with each other in a way other than lawful merger or joint research permitted for all industries under federal laws enacted in the 1980s. This critique would appear to be supported by empirical research suggesting that the economies of scale in the insurance industry are limited and that many insurers could survive at the minimum efficient scale (Geehan, 1986; Joskow, 1992). It has also been suggested that the collusive rating practices facilitated by the exemption from competition laws induces sloppy and inefficient cost control measures by insurers (Angoff, 1988). Finally, collective data gathering may trap underwriting practices in some local optimum whereas the greater experimentation facilitated by barring collusion might lead to more accurate methods of assessing policyholder risk.

Interestingly, the more recent regulations of the European Community appear to echo at least some of these criticisms of the broad American exemption from competition laws. Under Regulation 3932/92, for example,
collaborative compilation of statistics on expected losses is permitted as is statistical analysis of that data. Collective determinations of actual premiums, after expense loading and other cost factors are taken into account, is prohibited, however. Obligations among competitors to adhere to a joint pricing mechanism are void. While development of common policy forms and coverages are permitted, insurers are not permitted to enter into contracts with each other than would compel use of these forms and coverages.

**F. Conclusion**

Sensible regulation of insurance requires an understanding of a broad span of economics including finance theory, price theory, game theory, dynamics, international trade and econometrics as well as ancillary subjects such as statistical and information theory. It also requires an understanding of jurisprudential issues such as the appropriate specificity of rules, the appropriate centralization of legal authority across networked cultures, the appropriate allocation of law-making authority between legislature and judiciary, and the costs of various forms of dispute resolution. While the broad contours of insurance regulation have been profitably and extensively examined using tools from each of these disciplines separately, and while this entry evidences some considerable progress in a more interdisciplinary approach, many of the intricacies of legal doctrine that determine the actual course of insurance transactions remain unconquered by the powerful combination of law and economics.

**Bibliography on Insurance Regulation (5700)**


Jerry, Robert H., II (1987), Understanding Insurance Law, § 95c, New York, Matthew Bender.

Other References


**Cases**


