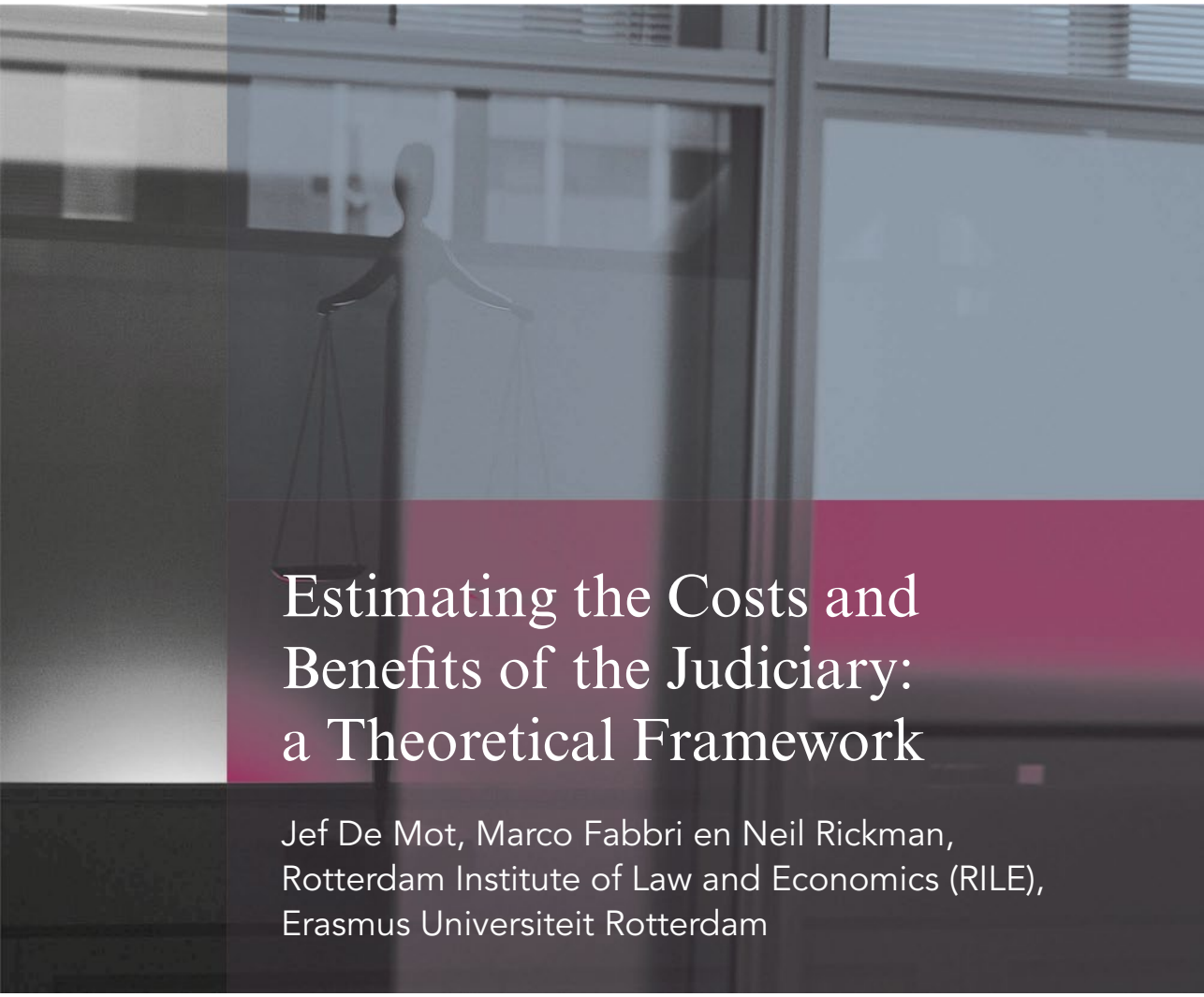
de Rechtspraak

Raad voor de rechtspraak

Research Memoranda
Nummer 5 / 2018
Jaargang 13



Estimating the Costs and Benefits of the Judiciary: a Theoretical Framework

Jef De Mot, Marco Fabbri en Neil Rickman,
Rotterdam Institute of Law and Economics (RILE),
Erasmus Universiteit Rotterdam

Colofon

Raad voor de rechtspraak
Postbus 90613
2509 LP Den Haag

Deze publicatie verschijnt in het kader van het wetenschappelijk onderzoeksprogramma van de Raad voor de rechtspraak. Uitgave daarvan betekent niet dat de inhoud het standpunt van de Raad voor de rechtspraak weergeeft.

Onderzoek in opdracht van de Raad voor de rechtspraak.

Jef De Mot, Marco Fabbri en Neil Rickman, Rotterdam Institute of Law and Economics (RILE), Erasmus Universiteit Rotterdam

Begeleidingscommissie
Guidance committee:

dr. F. (Frans) van Dijk (director Council for the Judiciary)

dr. F.P. (Frank) van Tulder (researcher Council for the Judiciary)

De integrale tekst van dit rapport is gratis te downloaden van:
www.rechtspraak.nl/Organisatie/Raad-Voor-De-Rechtspraak
Rubriek: wetenschappelijk onderzoek

Uitgever
Sdu Uitgevers bv, Den Haag

Vormgeving
Corps, Den Haag

Opmaak binnenwerk
LINE UP boek en media bv, Groningen

In opdracht van de Raad voor de rechtspraak

Estimating the Costs and Benefits of the Judiciary: a Theoretical Framework

Contents

Introduction

i. Project goals

Although the volume of scientific studies of civil justice systems is expanding, very few models exist which examine the various stages of civil litigation simultaneously. Most theories focus on one or a few of the stages. While the results of these theories are important, partial models do not allow much to be said about the total social costs and benefits of civil justice. In each stage – the *ex ante* stage, the filing stage, the settlement stage, the trial stage, the appellate stage and the enforcement stage – there are various elements which influence social costs and benefits. Moreover, these elements have a profound impact on each other through a complex web of decision making.

This project aims to provide, based on a thorough knowledge of the literature, a solid welfare economic framework for the future empirical work of the Netherlands Council for the Judiciary. That work aims to quantify the social benefits and costs of the judiciary in general and, more specifically, to measure the impact of concrete measures such as the reduction in the duration of court cases, a more uniform application of the law or lowering the costs of bringing a case before the court. The current measurements need to be improved in terms of completeness, theoretical underpinning and reliability. The framework should consist as far as possible of variables that have been measured or estimated empirically or can be measured or estimated in further empirical research, and will be linked explicitly with the approaches followed in earlier research of the Council for the Judiciary (Van Dijk 2014; Van Tulder 2014, 2016). The ultimate goal is to provide additional insights so that such models can be refined. In concrete terms, we will:

1. Provide a welfare economics framework based on a state-of-the-art overview of the literature related to the costs and benefits of each stage of a dispute and their influence on each other. Such a framework is indispensible when creating an aggregate model.
2. Point out deficits in the theoretical literature.
3. Identify which type (or types) of empirical research would be most promising to undertake in the future.
4. Examine which key variables are difficult to measure empirically and discuss methods which could be used (and perhaps have been advanced elsewhere) to shed light on the magnitude of these variables.

ii. Methodology

When discussing each of the stages, the following principles will be taken into account:

1. Models should start from the micro-economic decisions of the litigants at each stage, taking into account several policy variables which influence these decisions.

Our methodology is as follows. First we identify the various stages of a dispute and describe the types of costs and benefits of that stage; in other words, the factors which are relevant for welfare analysis and empirical research. Then we analyze in detail the benefits and costs of each of the factors specified in stages I to VI of the litigation process.

We will build upon the following framework/principles:

1. We analyze the costs and benefits as *comprehensively* as possible. This is a crucial step for future empirical research, because all too often studies focus on one or a few aspects of these factors (the most obvious ones), while ignoring other aspects which may however also influence the social calculus, sometimes in an unexpected or counterintuitive way (see e.g. the "freeway principle" introduced by Danzon and Lillard (1983)).[1] For example: a less than perfect recovery rate obviously has disadvantages, but what are the benefits of a recovery rate of less than 100 percent at the enforcement stage? Models on litigation expenditures can

1 To give a specific example of how reforms can backfire, Fenn, Grembi and Rickman (2016) show that measures taken in England and Wales to shield claimants from litigation costs when legal aid was removed actually increased these costs by allowing claimants to stop monitoring their lawyers' expenditure decisions.

partly shed light on this, (Katz, 1988). Another example involves increased litigation costs: they can obviously have a negative influence on social costs, but they also have benefits (e.g. increased settlement, less filings (Shavell, 1999)).

2. We examine the costs and benefits of these factors by explicitly taking into account their *interplay*. To do this, for each stage that we examine, we look at the consequences of this stage for earlier stages (through "backward induction"). For example: a less than perfect enforcement rate (stage VI leads to a lower amount at stake in stage IV (and stage V if an appeal is filed), which may lead to smaller litigation investments by the litigants at that stage. This will then trickle down to the earlier stages. Smaller litigation expenditures at stage 4 affect the previous stages (III, II and I), for example the willingness to settle, to file, to take care, to enter into a contract etc.

   We do not only look at the interplay of the factors which are part of different stages, but also at the interaction between factors within each stage. For example, it has been shown that judicial delay can motivate the judiciary to interpret the law differently (see e.g. De Mot, Faure and Klick (2015) on the switch from contributory to comparative negligence).
3. It is important not only to go through the various stages in reverse chronological order, but also to understand the effects that follow from feeding back the results of the early stages into the subsequent stages. This is important for future empirical research, because the various stages act as *filters* (e.g. which types of cases get filed, which types of cases get settled etc.). Empirical research needs to be able to distinguish these *selection effects* (e.g. more high-stake cases are filed) from behavioural effects (e.g. courts suddenly award more damages than before). We will discuss the "cradle-to-grave" models available in the literature (see e.g. the analysis by Hylton (2002) of how information asymmetries at the settlement stage can work their way back to earlier decisions) and how these models can be implemented empirically (see, e.g. Snyder and Hughes' (1990) estimation of the selection effects when methods for allocating legal costs were changed in Florida in the 1980s).
4. For each relevant factor (e.g. duration of settlement, outcome of trial, litigation expenditures) described above, we will break them down into *key determinants*, as far as possible. For example, existing research shows that the litigation expenditures of litigants (stage IV) depend *inter alia* on: the amount at stake, the unit costs of litigation, the merits of the case and the variability of the decision maker. Using backward induction, we will then examine how these determinants influence the decisions of the parties in earlier stages.

Having a clear view on all key determinants of the relevant factors is a crucial step for future empirical research. It can shed light on the *necessities and priorities in data collection*. In the area of civil justice, the value of detailed data collection is enormous. Long time series are necessary to discover trends in litigation, while cross sections and panel data are valuable for analyzing how different types of cases or litigants are affected by (say) a legal rule. There is a vast interest in the litigation process among scholars but obtaining solid and useful data is often extremely difficult. Data should be gathered at the source, thus by the judicial system itself, but a precondition is that the institutions know which data to gather.

5. For each theory on the interplay between several factors and each theory on the key determinants of the factors, we will investigate whether empirical studies exist which corroborate these theories. This is important for the type of future empirical research the Council has in mind, since contrasting theories yielding divergent predictions have been proposed with respect to some of the factors affecting the judicial process.

2. Whenever possible, the model should work with closed-form solutions

Many models of civil procedure remain relatively simple, so that closed-form solutions (i.e. solutions which can be derived from the model in a formal way) can be obtained using formal equilibrium analysis. The great advantage of this is that subsequently a relatively straightforward comparative statics analysis can be conducted which will deliver clear intuitions. When developing comprehensive models however, complications may arise making closed-form solutions unfeasible. For instance, this might be the case when it is uncertain which distribution one should use for certain parameters. For example, the more uncertain the law is or the more complex a contract is, the more likely it will be that the parties' opinions will diverge. The distribution from which the parties' expectations are drawn may be empirically uncertain, so that many distributions are possible. Simulation may shed light on the different consequences of differing distributions. Simulation may also be necessary when the model becomes too complex. This can for example happen when modelling the expenditure decisions of the parties under complex legal rules, because the success function is too complicated (De Mot and Miceli, 2015;[2] Van Tulder, 2016). We will indicate when it is unlikely that such closed-form solutions could be obtained.

2 They compare an all-or-nothing rule and a rule of proportionate damages.

3. Models should incorporate endogenous litigation expenditures

Although most models in the economic analysis of litigation assume that the probability of success for litigants is exogenously determined, more sophisticated accounts of litigation understand that litigation expenditures are variable and influenced by endogenous factors.[3] In an assessment of Van Tulder's model, Prof. Dari-Mattiacci suggested that endogenous litigation costs should be taken into account. We agree with this and moreover, we consider it absolutely necessary for several reasons. First, theoretical models suggest that litigation expenditures vary with the amount at stake, the unit costs of litigation, and the inherent merits of the case (Katz, 1988; Farmer and Pecorino, 1999; Hirshleifer and Osborne, 2001). In other words, assuming uniform litigation costs for disputes with different strengths, different amounts at stake or different unit costs may lead to inaccurate predictions. Second, empirical research confirms that these theoretical insights are valuable, for example by showing that higher stakes induce additional litigation expenditures (Kakalik *et al*. 1998; Willging *et al*. 1998).[4] Third, the classic model with exogenous litigation costs does not make it possible to distinguish clearly between meritorious cases with a low probability of winning on the one hand and weak cases on the other hand. We elaborate on this point below.

4. Models should take into account the quality of claims

Strong claims may sometimes have a low probability of success at trial, for example when access to evidence is troublesome for the party with a meritorious case.[5] Most models of litigation do not allow one to distinguish these claims from weak claims. Formally, both types of claims may have a low value of "p". For example, a probabil-

3 Note that the issue of endogenous litigation expenditure was initially recognized in the economic literature on litigation (Landes, 1971; Gould, 1973; Posner, 1973), but subsequently received relatively little attention until fairly recently (Farmer and Pecorino, 1999; Hirshleifer and Osborne 2001; Parisi, 2002; De Mot, 2013; De Mot and Miceli, 2015)

4 Kakalik *et al*. show that higher stake disputes "are associated with significantly higher total lawyer work hours on discovery and significantly longer time to disposition" (Kakalik *et al*. 1998, p. 639). Similarly, Willging *et al*. (1998) find that the duration and costs of litigation are strongly correlated with the monetary stakes in litigation. Additionally, empirical research describes situations where additional expenditures are shown to increase a litigant's success rate (Ashenfelter and Bloom, 1990; Ashenfelter and Dahl, 2003; Seron *et al*., 2001). For instance, Ashenfelter and Bloom (1990) observe higher final arbitration awards when unions retain representation and Seron *et al*. (2001) document superior outcomes for *pro bono*-represented tenants in New York City's Housing Courts.

5 Strong claims are claims in which the inherent merits of the claim are large. They are claims the plaintiff should win.

ity of winning (p) of 20 percent may indicate the presence of a weak claim, but also of a strong claim with problematic evidence issues. It is however crucial that such a distinction can be made, because of the different impact on various elements influencing the social costs and benefits (e.g. the decision to take care and the duration of settlement negotiations).[6] Models with endogenous litigation expenditures (see no. 3) are well suited to make such a distinction, because in these models the probability of a plaintiff victory depends *inter alia* on the inherent merits of the case (F). Note that models with exogenous trial effort can also make the distinction between weak and strong claims (e.g. weak claim: the defendant took due care; strong claim: the defendant took no care; probabilities of plaintiff victory can be derived by introducing fixed probabilities of type 1 and type 2 errors). However, such models have the disadvantage that they do not recognize that trial effort varies with the amount at stake and the merits of the case.

5. Models should incorporate uncertainty

The issue of legal uncertainty pervades each stage of the litigation process. Uncertain laws can not only undermine efficient *ex ante* behaviour, but also make it more difficult for the parties to form expectations on how the court will decide.[7] Thus more uncertainty will make it more likely that the opinions of the parties will diverge, potentially leading to more trials relative to settlement (at least in the short run). Also, when litigants are risk-averse, the uncertainty of the outcome leads to additional costs, and this may in turn significantly affect the volume of suit, the duration of settlement negotiations, and the outcome of such negotiations. Uncertainty and attitudes towards risk may furthermore influence the investments of the parties in a lawsuit, and thus the transaction costs of the legal system.

6. Heterogeneity should be recognized

Theories or specific findings that make sense for one type of dispute or one type of litigant may be less relevant for other types of disputes or litigants. In the context of what we have described above (at 1), this will often come down to major differences related to the key determinants. For example, some types of cases may be more difficult to decide and may thus be more sensitive to judicial error. In terms of the key determinants, the variability of the decision maker will increase, so that larger random

6 In his analysis of the model by Van Tulder (2014), Dari-Mattiacci also stresses the importance of not limiting models to situations in which those who bring suit have really been wronged. One should allow for the possibility of parties using the legal system strategically.

7 Note that complex contracts can have the same effect.

influences on the judicial outcome will become more common relative to smaller random influences.

We also recognize the importance of heterogeneity in the context of the two major theories explaining the incidence of trials: relative optimism (divergent expectations) and asymmetric information. Empirical research seems to indicate that both theories are relevant, with relative optimism affecting a whole range of disputes, and asymmetric information being significant for some specific types of disputes. In our discussion of all the stages of a dispute, we discuss insights from both canons. For modelling issues however, we focus on divergent expectations.

7. Micro and macro level

The theories we will describe will shed both light on the decision making of firms and individuals as well as on the social costs. Van Tulder (2014)[8] makes the distinction between the micro level (the former) and the macro level (the latter). The decisions at the micro level directly influence the elements of the macro level. For example, in the theories we will discuss, the following elements clearly belong to the micro level: the decision of a potential injurer/contractual party on whether or not to take care (and how much)/whether or not to fulfil his or her contractual obligations, the decision of the plaintiff on whether or not to sue, the decision of the parties on whether or not to settle, the decision of the parties on how much to spend at trial, the decision of the loser at trial on whether or not to file an appeal, etc. Turning to the macro level, the following social costs can be identified:

- the costs of the courts (both publicly and privately financed);
- the costs of legal assistance to the litigants (both publicly and privately financed);
- other costs to litigants (e.g. emotional costs; travelling expenses etc.);
- the costs of taking care in a non-contractual setting;
- damages related to physical harm or to property, both in a contractual and non-contractual setting;
- if the parties did enter into a contract, the non-realization of the surplus that a contract would have realized if the parties had performed their contractual duties;
- the costs associated with deterring efficient activities in a non-contractual setting;
- the unrealized surplus of contractual transactions which did not take place.

8 See Van Tulder, F (2014), In de schaduw van de rechter: Individuele en maatschappelijke kosten en baten van de juridische infrastructuur, p. 9.

iii. Summary of the conceptual framework and relationship with previous studies on the costs and benefits of the Netherlands Judiciary

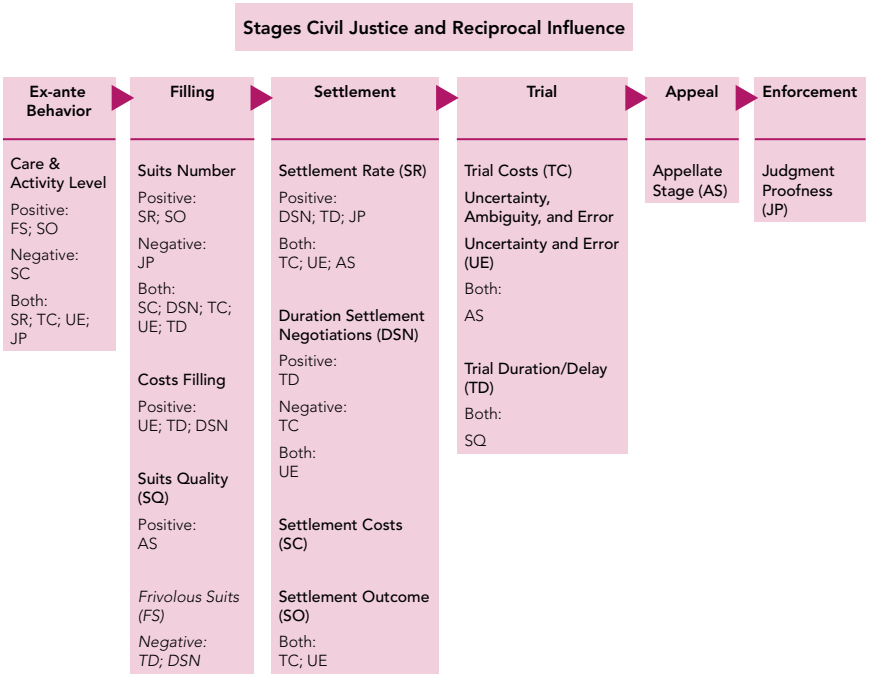
Following the parliamentary debate concerning the 1998 reforms to the Netherlands Judiciary, a study took place to identify bottlenecks in the judicial systems and to calculate associated costs and benefits deriving from potential improvements (Ministerie van Justitie, 1998). Fifteen years later, a recent publication of Van Dijk (2014) analyzes the actions undertaken to remedy the shortcomings listed by the 1998 study and empirically estimates costs and benefits associated with the reform process. Van Dijk's analysis recognizes the importance and validity of the main conclusions of the 1998 study. However, it also highlights a series of limitations of the theoretical framework adopted in it. In particular, the author observes that the micro-economic approach followed by the 1998 study is based on a simplified theoretical framework. For instance, the study considers only direct effects of the elements of the system (Van Dijk, 2014, pp. 3). Therefore, an improved theoretical framework could better explain observed elements of the judicial system and predict more precisely the effects of any envisaged reforms.

A first attempt to propose a comprehensive conceptual framework for engaging in micro-founded studies of the Netherlands Judiciary was made in the research memorandum of the Council for the Judiciary "In de schaduw van de rechter" by Frank van Tulder (2014). The author develops a model capturing several stages of the judicial process. The model is then calibrated using real data and employed to simulate the functioning of the judicial system. The model has been recognized as "a line of research that should be encouraged and supported, as it has enormous potential for policy improvement" (Koopmans and Gerritsen 2014, p. 23). At the same time, the model was considered to be not yet ready to serve policymaking purposes, but more suitable as a basis for further research and for the development of larger data sets and the next generation of models.

The present project builds on the contributions by Van Dijk (2014) and Van Tulder (2014, 2016), complementing and extending the theoretical framework proposed by the authors. In particular, our analysis further elaborates on the following elements necessary to produce a more comprehensive costs and benefits estimation of the judiciary:

1. We discuss the consequences related to uncertainty and inconsistency deriving from the non-uniform application of the law. We show how these elements affect each stage of the judicial process, specifying the links between stages. We also explicitly highlight the recursive effects of these elements, explaining how consequences in one stage trickle down to previous stages.
2. We elaborate on the case disposal time. We emphasize how delay, in general, has a negative effect on social welfare. We specify how expected delay during the trial influences agents' behaviour in previous stage. We also examine some consequences connected to trial delay that can induce welfare-increasing behaviours in previous stages and, therefore, reduce the total associated costs.

The theoretical framework characterizing our analysis includes six stages: *ex ante* behaviour, filing, settlement, trial, appeal, and enforcement. Figure 1 below summarizes the elements determining costs and benefits of the judicial process and specifies how these elements link each individual stage with previous stages.



List of Abbreviations	Explanation
AS = Appellate Stage DSN = Duration Settlement Negotiations FS = Frivolous Suits JP = Judgment Proofness SC = Settlement Costs SO = Settlement Outcome SQ = Suits Quality SR = Settlement Rate TC = Trial Costs TD = Trial Duration/Delay UE = Uncertainty & Error	Each column represents a stage of the trial. Quadrants report aspects of the trial that that arise at the corresponding stage and that have an influence of on other aspect of the trial reported in another quadrant (mainly previous stages). Each aspect of the trial is coded by reference to its initials. For instance, Settlement Costs is coded as SC. The influence of one aspects of the trial on another is signaled by indicating the code corresponding to one aspect of the trial below another aspect of the trial. For instance, SC under Suits number indicates that Settlement Costs affect the number of suits brought to trial. The type of influence is indicated as either Positive; Negative or Both. Both means that the influence can be either positive or negative. For instance, the fact that an increase in Settlement Outcome (SO) increases the number of suits brought to trial is indicated by writing SO in the "Positive" row under Suits.

Figure 1: summary of the conceptual framework

1

Stage I.
Ex ante behaviour

1. Description

In the *ex ante* stage, we consider the harm that one person does to another, and the decisions persons make which affect the likelihood and extent of the harm. Harm can take many forms, such as breach of contract, tortious injury, trespass upon property, etc.

1.1. Tort law

In the economic analysis of tort law, harm prevention may take two forms. First, potential risks can be decreased by *investments in care*. For example, drivers may reduce the probability of accidents by driving more slowly and factories can reduce pollution by using filters. Second, the risk of harm can further be lowered by altering one's *activity level*. Less driving, for example, just like careful driving, diminishes the probability of car accidents. In an ideal world, tort law rules would induce parties to behave optimally with respect to levels of both care and activity.

1.2. Similarities and differences between tort law and other fields of law from an economic perspective

Cooter (1985) showed the power of economic models to provide a unifying framework for different fields of law. He used the classic accident model (the "model of precaution") to reveal an underlying unity in torts, contracts, and property law. Cooter argues that most legal conflicts, independent of the subject matter, can be conceptualized as problems of care in the sense that they involve a dispute between two parties, each or at least one of them possessing the ability to take some action to avoid

or mitigate the resulting harm. Cooter showed that the basic accident model used to assess the efficiency of various liability rules in tort law can also be used to study certain rules in contract law and property law. For example, in contract law, the breach of a contractual promise can be seen as an accident, with the "injurer" being the breaching party (the promisor), and the "victim" the party to whom the promise was made (the promisee). Note that the fact that the breach can be a deliberate act does not alter the mathematical logic of the model. The promisor's "care" in this setting is the decision on whether or not to breach the contract. The promisee's "care" is how much to invest in reliance on performance. One can easily show that awarding victims full damages for breach (the "expectation damage" measure) is equivalent to strict liability in torts. It leads to efficient incentives for promisors to avoid breach, but inadequate incentives for promisees to mitigate damages.[1] Cooter also applies the model of precaution to property law, more specifically to government takings of private property under eminent domain. Here, the injurer's care corresponds to the government's taking decision, and the victim's care represents the landowner's investment in improvements to the land (this influences the value of the property in the event of a taking). So even though the actual rules that have come up in these different areas of the law differ in their outward form, Cooter's analysis shows that they often exemplify the same fundamental mathematical structure. The law has converged on analogous solutions but with different forms to a set of common economic problems.

While the conceptual framework of tort law can be applied to many contract situations in which a contractual party needs to take some care measure to reduce the probability or extent of harm,[2] some other types of contractual relationships fit less well in such a framework and we need to emphasize some other elements which influence social costs.

a. With respect to *debt contracts*, at first sight one may think that not paying back a loan involves merely redistributional issues: the party who does not pay back the loan wins what the other party loses. However, the possibility that a loan will not be paid back creates social costs:
   - monitoring costs: in order to assess the probability of a default, the lender will incur costs *ex ante*, and sometimes also *ex post*.

1 Note that a rule that limits damages to the amount that promisors could have *reasonably foreseen* will, just like negligence, create efficient incentives for both parties. Such a rule essentially creates a threshold for the promisor by limiting his/her liability to the level of damages that would result from efficient promisee reliance. This creates the same efficient bilateral incentives as a negligence rule.

2 E.g. medical malpractice cases, product liability cases.

- a higher interest rate: when the rate of default increases, lenders will increase their interest rates. This can influence the social calculus since some socially valuable projects will not be undertaken. Suppose for example that a person, A, contemplates engaging in a project which will bring an added value of 5 percent of the invested sum. If there was no default on loans, the interest rate would be 4 percent, and A would choose to make the investment. If there is a certain fraction of default, interest rates may increase above 5 percent, in which case A will not be willing to make the investment. The higher interest rate can have secondary effects:
    1. Adverse selection: those who agree to pay a higher interest rate may be those who were not planning to repay anyway.
    2. Individuals who were planning to repay may now face greater difficulties in repaying, further increasing the default ratio.
- Transaction costs related to collateral: the creditor may want to secure his loan by collateral. This can create additional contractual transaction costs and in some cases transaction costs to inform third parties.
- Costs of non-use: for some types of collateral, the debtor will not be able to use it until he repays the loan.

b. *Production contracts (for goods or services).* The possibility of contractual breach can influence social costs and benefits in several ways:
    - Monitoring costs: the parties may spend resources in order to reduce the probability that the other party will breach his or her promise.
    - The cost of lost investments: when value is only added when both parties make an investment, but only one party has actually made such an investment while the other party has committed a breach, then (part of) the investment may be wasted. This will be especially socially harmful in the case of relation-specific investments.
    - Fewer value-adding transactions: as the probability of inefficient breach becomes larger, the willingness of some parties to engage in value-creating transactions will diminish, since the expected benefits of entering into a contract decrease.

1.3. Determinants of disputes

In the discussion of the various stages, we will examine the influence of each stage (the filing stage, the settlement stage, the trial stage, the appellate stage and the enforcement stage) on earlier stages, including the *ex ante* stage. The determinants of these stages will obviously influence the *ex ante* stage. For example, everything else held constant, larger settlement amounts may lead to an injurer taking more care or a party to a contract committing fewer breaches.

2. Consequences for other stages

In the discussion of the other stages (the filing stage, the settlement stage, the trial stage, the appellate stage and the enforcement stage), we will discuss the influence of each stage on the previous stages. Of course, given that the *ex ante* stage is the first stage, such an analysis cannot be provided here. What is important here is to list the variables which are endogenously determined by the *ex ante* decisions of the parties and which influence the behaviour of the parties in the subsequent stages:

1. *The inherent quality of the case*: The way parties behave *ex ante* will determine the inherent quality of the case. For example, when a party to a contract intentionally and without any justification decides not to perform any of his contractual duties, the inherent quality of the case will be strong. Likewise, when a potential injurer took due care (or more) to avoid an accident, the inherent quality of the case will be weak. We would stress that a low or high inherent quality does not automatically lead to a very low or high probability of a plaintiff victory. As we will discuss thoroughly in the discussion of the trial stage further on, the probability of a plaintiff victory may not only depend on the inherent quality of the case, but also on the expenditures of the parties to persuade the court.
   The influence of the inherent quality of a case is very important. As we will see further on, its influence on many elements of the social cost function of litigation has been theoretically and/or empirically established. The inherent quality influences the expenditures of the parties, the probability of a plaintiff victory, the duration of settlement, the settlement outcome, the probability of a trial[3] etc.
2. *The unit costs of effort*: *Ex ante* behaviour can influence the amount of evidence that will be available in court. For example, it may be easier to gather evidence regarding behaviour which falls far below the demands of the legal standard or

3 The parties may e.g. find it more difficult to estimate the probability of victory for closer cases compared with very weak or very strong cases, and so their estimates are more likely to diverge for closer cases.

rule than for behaviour which is almost but not entirely compliant with the legal standard or rule. Also, *ex ante*, parties can make investments in order to make proof *ex post* more difficult or easier. From an economic perspective, all these elements will influence the unit costs of litigation of the parties. If effort is expressed as the number of hours that a party and his or her lawyer spend on a case, then the unit costs can be seen as the cost of one hour of legal services. For example, when the behaviour of a defendant can be considered as gross negligence, a plaintiff will more often find it easier to find evidence in his favour than when the behaviour of the defendant can only be considered as slight negligence. In other words, the unit costs of the plaintiff will be lower in the first instance. As we will see, these unit costs can have great implications for the ultimate effort levels of the parties, the settlement frequency, the accuracy of the trial outcome etc.

3. *The amount at stake*: Sometimes, the amount at stake is exogenously determined, in other words not within reach of the parties *ex ante*. Frequently however, the *ex ante* behaviour of the parties will not only determine whether a contract is breached or an accident occurs, but will also influence the extent of the harm in the case of breach or an accident. For example, driving faster will not only lead to more accidents, but also to more severe accidents on average. A party to a contract can decide to not fulfil his or her contractual obligations at all or to fulfil them to some degree, etc.
   Like with the previous variables, the influence of the amount at stake in the future stages of the dispute is great. Its influence on many elements of the social cost function of litigation has been theoretically and/or empirically established. The amount at stake influences the expenditures of the parties, the probability of a plaintiff victory, the duration of settlement, the settlement outcome, the probability of a trial etc.

3. Modelling issues

3.1. Existing cradle-to-grave models

There are very few models in the existing literature which incorporate the *ex ante* stage and several later stages (most importantly the filing stage and the settlement/trial stage). Here we discuss two such "cradle-to-grave models": one that focuses on divergent expectations, and one that deals with asymmetric information.

3.1.1. A divergent expectations model

Van Wijck and Van Velthoven (2000) build a model which incorporates a potential injurer's decision to engage in a harmful activity, the plaintiff's incentive to sue, and the parties' incentives to settle or go to trial, and analyze whether a shift from the U.S. rule of cost allocation (each party bears his or her own legal costs) to the U.K. rule of cost allocation (the loser at trial pays all legal costs) results in an improvement in efficiency.

The model is structured as follows. An individual, A, can engage in an activity that causes harm, S, to individual B. If he or she engages in the activity, he/she will derive a personal gain GA. The net gain of the activity not only depends on this personal gain, but also on the costs that may be involved in a settlement or a trial. The applicable legal rule is strict liability. If it comes to a trial and B wins the trial, he will receive damages equal to S. The trial costs of A and B are equal to C_A and C_B respectively. Under the U.S. rule of cost allocation, A and B each pay their own trial costs. Under the U.K. rule, the loser at trial pays all trial costs ($C_A + C_B$). A thinks his probability of success at trial equals P_A, while B thinks his probability of success equals P_B.

The authors show the importance of adding the *ex ante* stage. Adding this stage can significantly change the results of studies which do not take this stage into account. Most importantly, the standard model of litigation has shown that moving from the U.S. rule to the U.K. rule of cost allocation will result in an increasing number of trials. Intuitively, the U.K. rule magnifies the effect of relative optimism[4] (the parties may now also be optimistic as to who will end up paying the trial costs). As Van Wijck and Van Velthoven argue, the rule of cost allocation also affects whether a potential injurer will engage in a certain activity, and this also affects the number of trials. The conclusion from the standard model is no longer generally valid when the *ex ante* stage is taken into account. A shift from the U.S. rule to the U.K. rule may lead to an increase in the number of trials at the expense of the number of settlements, but at the same time there may be an increase in the number of discouraged activities at the expense of a reduction in the number of trials. Taking these two effects together, a reduction in the number of trials is possible. From an efficiency point of view, the U.S. rule delivers better results for activities that provide a personal gain to the injurer which is larger than the harm inflicted on the victim (thus activities that should not be discouraged). For other types of activities however, one cannot draw any unambiguous conclusions. A shift from the U.S. to the U.K. rule may or may not lead to an improvement in efficiency.

4 Which means that the plaintiff estimates his chances of winning as being higher than the defendant's estimate of the plaintiff's chances of winning.

3.1.2. An asymmetric information model

Hylton (2002) provides a cradle-to-grave model of liability incorporating the decision to comply with the legal standard, the decision to file suit, and the decision to settle.

The model is structured as follows. The litigants are risk-neutral. Potential injurers can take costly precautionary measures (x, which has distribution G, and x is unobservable to potential victims; injurers however know their cost of care) to reduce the probability that victims will suffer losses (v, with distribution function H, and this loss is observed by all parties once realized). Potential injurers choose between taking no care at all and taking care (care is thus not a continuous variable). Obviously, the probability of loss when no care is taken (p) is larger than the probability of loss when care is taken (q). Litigation is costly. The plaintiff's cost of litigating equals Cp, the cost of the defendant Cd. The court sometimes makes mistakes in deciding liability. There is a probability of Q1 that the court will erroneously fail to hold a defendant liable (type 1 error), and a probability of Q2 that the court will erroneously hold a defendant liable (type 2 error). Both injurers and victims know these probabilities. Although the courts are not fully accurate, they are assumed to be sufficiently accurate in the sense that the probability that a defendant who did not take care will be held liable is larger than the probability that a defendant who took care will be held liable ($1- Q1 > Q2$). An injurer violates the legal standard if he fails to take care when the cost of care is smaller than the increase in the expected loss of the victim ($x < (p-q)E(v)$). Hylton stresses that this test cannot only be applied to tort law, but that it can serve as a model of a general balancing test similar to those used in most areas of law.

From these elements, Hylton derives the plaintiff's and the defendant's estimate of the probability of a verdict for the defendant. Given the existence of asymmetric information, the plaintiff does not know whether the defendant complied with the legal standard. The plaintiff's estimate thus equals: $Pp = W(1-Q1) + (1-W)Q2$, where W is the probability, given an injury, that the injurer did not comply with the legal standard. This non-compliance probability is endogenous and depends on the expected liability of a guilty defendant and the expected liability of an innocent defendant. The defendant of course knows whether or not he complied with the legal standard. When the defendant did not comply, then $Pd = 1-Q1$, when the defendant complied $Pd = Q2$.

The rest of the model follows the classic structure of many asymmetric information models. First, the plaintiff decides whether or not to file suit. Then the informed

defendant makes a settlement offer (this is a "signalling model"), which the plaintiff either accepts (settlement) or does not. If the plaintiff rejects the offer, the plaintiff either goes to trial or drops the suit.

Incorporating several stages shows important relationships between those stages. For example, Hylton shows that litigation to judgment occurs only in equilibria in which some but not all actors comply with the legal standard. Second, the probability of a settlement is determined not only by litigation expenses and litigation stakes, but also by the rate of compliance. Third, the author shows that plaintiff win rates will generally be smaller than 50 percent if defendants have an informational advantage. Fourth, the relationship between changes in the plaintiff win rate and changes in the probability of judicial error are not as simple as basic intuition may suggest. Suppose there is an increase in the probability that innocent defendants are found guilty. Intuitively, one would expect an increase in the plaintiff win rate. However, an increase in the number of such type 2 errors may actually lead to a reduction in the plaintiff win rate. If informed guilty defendants settle more often in response to the change in errors, the sample of disputes litigated to judgment will contain a higher proportion of innocent defendants, which could result in a lower win rate for plaintiffs. Fifth, the model shows that even with some level of inaccuracy in assessing liability, the legal system may still perform quite well as a deterrent mechanism. Finally, using simulation techniques, Hylton compares four rules of cost allocation, including the U.K. and the U.S. rule of cost allocation, on overall welfare grounds (welfare is defined as the negative of the sum of injury costs, avoidance costs and litigation costs). He finds the U.K. rule is the superior rule. This is mainly due to its effect on compliance. More specifically, the U.K. rule creates the highest level of compliance because it maximizes the spread between the expected levels of liability of guilty and innocent defendants.

3.2. Strengths and weaknesses of existing cradle-to-grave models

These two cradle-to-grave models show how important it can be to study all the stages of a trial simultaneously. However, both models discussed have several gaps:

1. Although both models incorporate the *ex ante* stage, they only do so partially. In the model by Van Wijck and Van Velthoven (2000), it is assumed that the activity can be performed with one specific level of care only. There is no choice between care levels. The injurer can only decide whether or not to engage in a certain

activity. In reality, both activity level and care level will influence the social calculus. In the model by Hylton (2002), the opposite occurs. Here, the injurer can choose between taking care or taking no care, but his activity level is fixed. Note that the assumption of discrete care (either care or no care at all) is also restrictive. Models with a continuous function for the probability of accidents may be more realistic for several types of situations.

2. The model by Hylton (2002) makes it possible to distinguish between strong and weak claims (claims in which the injurer did not take care and claims in which the injurer took care). Through the introduction of type 1 and type 2 errors, one can then compute the probabilities of victory in court for both types of claims. The model by Van Wijck and Van Velthoven does not however make it possible to distinguish between strong and weak claims. The probabilities of victory are given. Note that there is not necessarily a one-to-one relationship between the strength of a claim and the probability of victory. In some situations for example, a plaintiff can have a strong claim but at the same time a low probability of victory (e.g. due to evidence problems).
3. Both the Van Wijck and Van Velthoven model and the Hylton model do not take into account the following elements:
   a. endogenous litigation expenditures;
   b. the duration of trial and settlement;
   c. the appeal stage;
   d. the enforcement stage.

4. Empirical issues

4.1. Does the law have a deterrent effect?

We need to – even if only briefly – touch upon the issue of whether the law is actually capable of deterring certain behaviour. The importance of this question is that if the answer is negative, then the *ex ante* stage will solely be influenced by factors such as economic growth, and not by the (elements of) other stages.

With respect to torts, there is a major debate in the literature as to whether tort law deters risky behaviour. Some scholars find it plausible that individuals and firms respond to the risk of liability with increased care or by altering their activity level. Others find deterrence theories unrealistic because they assume that prospective

tortfeasors understand risk and law (which may be especially troublesome for many individuals). There have been many empirical studies that have tried to find out whether the law has a deterrent effect. It would be impossible to give a complete overview of all these studies. We limit ourselves to a selection, some finding no effect, others finding a significant effect.

Some studies of particular industries in the United States find no or little evidence that tort law significantly affects product safety. For example, Priest (1988) found no association between insurance premiums and injury rates, nor between times of increased tort liability and changes in death rates and injuries. However, this study is not conclusive, because the author did not distinguish other factors that could be relevant to the findings. Also, the accident rates Priest considered were *total* accident rates rather than the rates due exclusively to accidents caused by product defects. Dewees, Duff, and Trebilcock (1996) review literature on the effect of product liability on product safety. They conclude that product liability does not lead to a decrease in product-related accidents. These authors however stress some limitations of the data. With respect to the effect of product liability on the safety of general aviation aircraft, there was a sharp increase in liability from the 1970s to the 1980s: the liability and defence expenditures of manufacturers of these aircraft grew approximately ten-fold in that period. However, some authors established that the rate of fatal accidents did not exhibit a decline that could be associated with the sharp increase in liability (Craig 1991; Martin 1991). The accident rate had been falling for several years, and in the years during and following the increase in liability, the accident rate did not decline more sharply. On the contrary, it actually fell less steeply. There are however several reasons why one may expect that product liability may not (strongly) affect the safety of general aviation aircraft: (1) there are strong market forces at play: purchasers of aircraft have a strong incentive to gather information about the safety records of the planes they will be flying; (2) regulation in this area is already extensive. With respect to motor vehicles, Graham (1991) used regression analysis to find out whether product liability reduced motor vehicle fatalities in the period from 1950 to 1988. He established that higher liability was associated with a higher accident rate, but the effect was not statistically significant. The author also undertook five case studies of specific safety problems in motor vehicles (fuel tanks, gears, the roll bar, air bags and seatbelts). His conclusions remained the same. When safety problems came up, manufacturers responded first and foremost because of a concern about their reputations with consumers and because of pressure from regulators.

Other studies do find an effect. Landes (1982) examined the incentive effects of no-fault automobile insurance in the U.S. She used panel data from the 50 states and the District of Columbia in the time period 1967-1975. She found that no-fault states experienced up to 10 percent more fatalities per capita than tort states. This article received much criticism because of several alleged shortcomings (e.g. the model does not fit the U.S. no-fault laws because these laws are far from pure no-fault). Devlin (1992) examines the consequences of (pure) no-fault in Quebec and concludes that no-fault resulted in a significant increase in the fatality rate. However, the article does not account for some other legal changes (regulations) that were made at the same time. Also, these "earlier" papers do not tackle the problem of endogeneity. Indeed, the no-fault variable could be endogenously determined. It is possible that states adopt no fault partly based on the state's observed fatality rate. If that is the case, then the estimates of the response of fatalities to no fault would be biased. Cummins *et al*. (2001) took into account this endogeneity issue. Their results indeed suggest that the no-fault variable should be regarded as endogenous. They also find that no-fault significantly increases the fatality rate (12.8 to 13.8 percent). However, also this study is not immune from criticism. To tackle the endogeneity issue, the authors use the instrumental variables technique. More precisely, they use the predicted probability that a state has no-fault as an instrumental variable. This predicted probability comes from an equation the authors construct based on five exogenous variables. Two of these variables may be problematic: the population density and the percentage of the population which lives in urban areas. The problem is that these variables may be correlated with the fatality rate, making the instrumental variables technique problematic. Sloan, Reilly and Schenzler (1995) examined the effects of no-fault insurance on the incidence of drunk driving and binge drinking. They found that the propensity of drivers to binge drink is slightly higher in no-fault states. Focusing on a different area of the law, Horwitz and Mead (2009) examine the relationship between volunteer tort immunity and volunteering. In the United States, almost every state provided some level of volunteer immunity in the 1980s and 1990s.[5] Congress followed in 1997, with the introduction of the Volunteer Protection Act. This Act provided immunity for volunteers in states without immunity laws. The authors compare volunteer rates across states with different liability regimes in order to identify the effect of a reduction in tort exposure on volunteering. They use data from the Independent Survey's Giving and Volunteering surveys. They find systematic differences in volunteer rates between states with and without immunity. The evidence suggests that individuals reduce their activity level by foregoing volunteering in the face of

5 Lawmakers were worried that tort liability harmed nonprofit organizations.

exposure to liability. This study is interesting for three reasons. First, it shows how liability exposure can influence individuals rather than corporations, which are more likely to internalize the costs of their behaviour than individuals. Second, there are no laws (at least in the U.S.) forcing volunteers to buy liability insurance. Third, the study attempts to estimate the costs of foregone services: $4.4 billion per year. This is based on: (1) the finding that there is 7.5 percent more volunteering in states with some immunity than in states with no immunity, (2) that volunteers spend 50 hours per year on average on volunteering, (3) the assumption that the value of a volunteer hour equals $17.80, and (4) that about 65 million people volunteered during the relevant period. Note that the results of this kind of study are not sufficient to shed light on the social utility of tort law. The study for example does not identify the potential benefits of tort law (the costs of the accidents avoided, the additional care induced by the exposure to liability).

4.2. Critical note on the generalizability of existing studies

Existing studies on automobile accidents and other types of accidents (e.g. medical malpractice) do not lead easily to generalizable conclusions. Automobile accidents are muddied by criminal law and insurance mandates. Medical malpractice is confounded for example by mandatory insurance laws, professional duties and organizational relationships. All these complications influence deterrence, and these complications may be absent in different contexts.

4.3. Estimating the probability of accident functions

In the analysis of stage II we will examine the condition under which lawsuits are socially valuable. One element that is of crucial importance in this endeavour is the function for the probability of an accident, p(x), with x being the level of care. At a minimum, one needs to know the probability of harm with a current level of suit and the probability of an accident when the injurer takes no care.[6] And even if such information is available and it is established that the current level of suit is socially valuable,[7] the question arises of whether more suit is even more socially valuable. To be able to determine this, one would (*inter alia*) need information on the probability of an accident at the new equilibrium level of care. In other words, to be able to determine the

6 If however reputation effects are at play, or if the injurer may suffer harm himself or herself in the event of an accident, then the injurer will take some care even if there is no legal obligation to do so. In that case, one needs to know the probability of an accident at that level of care rather than at zero care.

7 Taking into account many other elements, such as the costs of filing and trial.

optimal amount of suit, one needs to estimate p(x) on the full range of x. As far as we were able to check, this has not been done in the literature for any type of accident.[8] This makes sense: for many types of accidents, it is already difficult to determine the accident rate for a level of prevention which is actually chosen by potential injurers,[9] let alone for levels of prevention which are not actually chosen by injurers.

Sometimes (but rarely), one can find an example of a specific function for the probability of an accident in the literature. For instance, Miceli (2008) uses the function $p(x) = 0.005e^{-\theta x}$ to be able to determine whether suit is socially valuable in a hypothetical case. Note that θ can be seen as a parameter reflecting the productivity of care. This function clearly satisfies some properties which are thought to be generally true for most types of accidents: (1) more care reduces the probability of an accident ($p'(x) = -0.005\theta e^{-\theta x} < 0$), and (2) this reduction becomes increasingly smaller as prevention increases ($p'' = 0.005\theta^2 e^{-\theta x} > 0$). Note that the function $p(x) = 0.005e^{-\theta x}$ is an example of the more general function $p(x) = \alpha e^{-\theta x}$, where α represents the probability of an accident when the injurer takes no care (x=0). In other words, even if $p(x) = \alpha e^{-\theta x}$ would be the correct functional form for the probability of an accident for some types of activities, the social desirability of suit would depend on both α and θ, and many combinations of these parameters are possible. The figures below demonstrate the profound influence these parameters have. In (a) α=0.005 and θ=1,in (b) α=0.005 and θ=2, in (c) α=0.02 and θ=1, and in (d) α=0.02 and θ=2. The x-axis always represents the level of precaution, and the y-axis the probability of an accident.

8 We have double checked this with two renowned empirical scholars, Prof. Jonathan Klick and Prof. Yun-Chien Chang.

9 For example, it may be difficult to document some type of accidents.

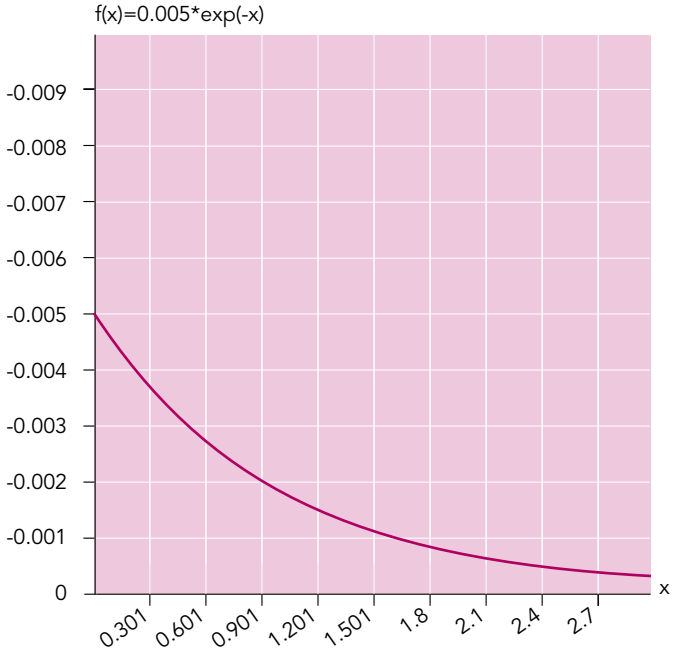


FIGURE a

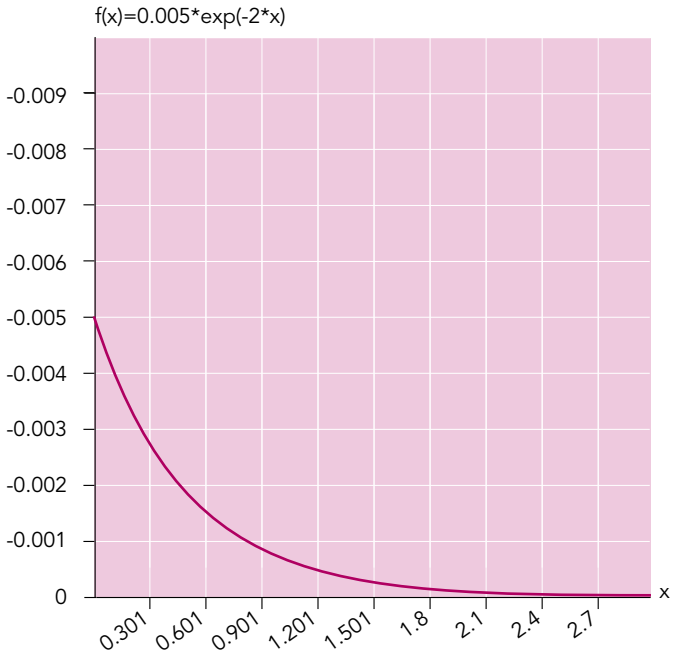


FIGURE b

FIGURE c

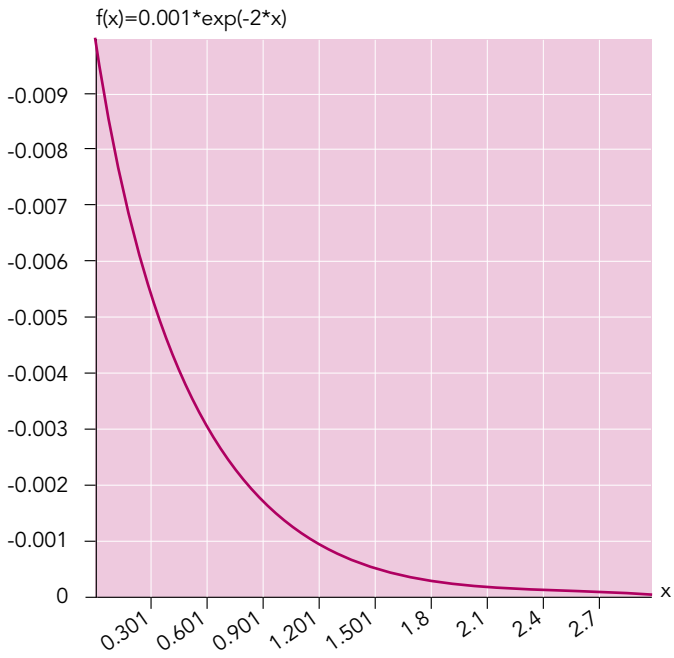


FIGURE d

4.4. Examining the level of influence of legal changes on *ex ante* behaviour

When trying to establish the extent to which a legal change has altered the *ex ante* behaviour of contractual or extra-contractual parties, one obvious difficulty is that other (legal) changes may have occurred simultaneously, and that these other changes need to be controlled for. A more subtle problem is the bilateral character of *ex ante* behaviour. Outcomes (whether an accident occurs, whether a contract is performed) often depend not just on how one party behaved but on how both parties behaved. As the new legal rule changes the incentives of one party, in light of this, the incentives of the other party may change as well, and sometimes in the opposite direction (e.g. a legal change causes the first party to take more care, and because of this, the other party takes less care). An ideal experiment would randomly assign to each party potentially involved in a contract breach or an accident its own rule. For example, we would observe a party (e.g. consumer) who is compensated for his harm even though the other party is not held liable, and also a party who is not held liable even though the other party was compensated.

Sometimes, a quasi experiment with characteristics similar to the ideal experiment is possible. This is the case when the choice regarding the *ex ante* behaviour of one of the parties is made well before the legal change, while the other party still needs to make decisions after the legal change. For example, Helland and Tabarrok (2012) examine whether the U.S. General Aviation Revitalization Act of 1994, which exempted manufacturers of small aircraft from product liability claims when their aircraft reached 18 years of age, had an impact on the behaviours and safety investments by pilots and owners. The authors exploit the fact that the manufacturers of aircraft can do little to change their investment in safety once an aeroplane is in the hands of the buyer. The authors find that for aircraft no longer covered by liability, the probability of an accident decreases. Pilots and owners of aircraft increase their investments in safety and decrease the use of aircraft. So in the quasi-experiment, investments in safety by manufacturers are held constant even as consumers lose the possibility to sue.

Another problem that arises is that often data is lacking regarding the *ex ante* behaviour of the parties (e.g. level of precautions; behaviour leading to contractual breach). In some cases, data is available but only for a selection of cases: e.g. those cases in which an accident happened, or those contract cases which ended up in court. The

problem is that these cases are unlikely to be a random sample of all cases. One can for example expect that the level of care was higher in those cases that did not lead to an accident or to a breach of contract.

Sometimes, this problem can be avoided. For instance, some types of accidents are less influenced by safety investments than others. For these types of accidents, the information on safety investments in the accident file may be more likely to be a random sample of safety investments in the whole population. For example, in the study by Helland and Tabarrok (2012), the empirical model is estimated using not only all accidents in the sample, but also for a subset of accidents in which it was determined that weather was the sole causal factor in the accident. The distribution in accidents for which the weather was thought to be the sole causal factor approximates more closely the distribution in the population of general aviation aircraft. That distribution can more accurately estimate the influence of the legal change on changes in safety investments.

2

Stage II. The filing stage

1.1. Description

In this stage, we focus on three elements which directly or indirectly influence the private and social costs and benefits: (a) the number of suits, (b) the costs of filing suit and (c) the quality of the suits which are filed (e.g. meritorious suits/frivolous suits).

1.1.1. The number of suits

The number of suits equals the number of situations in which the person who allegedly suffered harm either officially files the required documents with the court or informally asserts his claim by privately contacting the other party (Cooter and Rubinfeld, 1999). It includes:

1. The disputes that are settled before or during trial;
2. The disputes that end with a court verdict;
3. The disputes that end because the plaintiff decides not to pursue the case anymore after he has asserted his claim.[1]

The fact that most legal disputes are settled (or dropped) without ever coming to the attention of legal authorities means that generally there are no good data available on the total number of suits. Only the officially filed suits are well documented.

1 Note that we did not include the disputes that are ended because the plaintiff decides not to assert a claim, even though he may have incurred legal costs before making this decision. For example, an injured person may, after having contacted his lawyer, decide not to take any further steps. The alleged injurer may even never know about the harm he may have caused.

1.1.2. The costs of filing suit

In civil disputes, a litigant may face different types of costs:[2]

a. the fees charged by one's own lawyer;
b. the fees for the services of the bailiff (who may be needed to serve a writ of summons on the other party or who will recover the outstanding claims awarded);
c. court fees (these fees constitute a contribution to the costs of the proceedings, and must be paid in advance to the court that will be hearing the case);
d. expenses incurred by witnesses and experts who are called to testify (e.g. travel and accommodation expenses);
e. legal costs of the winning party (of course this is only relevant if some type of fee shifting exists);
f. psychological costs;
g. reputational costs.

Only some of these costs are incurred (if at all)[3] at the filing stage, namely some of the fees for the services of the bailiff and court fees; the influence of other costs is most heavily felt in later stages (e.g. lawyer fees at the trial stage).

1.1.3. The quality of filed suits

There is absolutely no consensus in the literature on how to define the quality of a lawsuit. This can be easily seen by focusing on the literature dealing with frivolous lawsuits. As Stone and Miceli write: "Different groups perceive frivolous lawsuits in different ways, and hence, there is some ambiguity with respect to defining, and correspondingly evaluating, frivolous suits" (Stone and Miceli, 2013, p. 304). The United States Supreme Court has defined a frivolous suit as follows: "a complaint, containing as it does both factual allegations and legal conclusions, is frivolous where it lacks an arguable basis either in law or in fact".[4] Black's Law Dictionary defines a frivolous suit as "lacking a legal basis or legal merit" (Black's Law Dictionary, 2009). Although one may conclude on the basis of these descriptions that certain causes of action are clearly frivolous, for some cases, the dividing line between a frivolous and a legitimate lawsuit may be difficult to identify *ex ante*. Logically, the term "frivolous" is defined differently in different jurisdictions. In some jurisdictions, subjective bad faith is implied, while in others a more objective standard is used (Keeling, 1994).

2 Note that these costs will be part of the social costs as well, but the social costs include a broader category of items (e.g. benefits of creating a precedent, which are seldom incorporated by the actual litigants).

3 Since a case can also be filed without bringing it to court.

4 *Neitzke v. Williams*, 490 U.S. 319, 325 (1989).

It is important to recognize that a frivolous suit is not the same as a negative expected value (NEV) suit. A very meritorious suit can have negative expected value. For example, suppose the law and the facts are all very much in favour of the plaintiff, but that providing evidence is quite costly for the plaintiff (and the U.S. rule or a limited version of the U.K. rule of cost allocation applies),[5], and that the harm suffered was quite low (e.g. probability of victory is 90%, harm = 1,000, cost = 1,500). Then the suit may have negative expected value ($0.9x1{,}000<1{,}500$), without being frivolous.

Note that there may exist several motivations for initiating lawsuits without merit. An undeserving plaintiff may simply try to extract a positive settlement offer. But frivolous suits can also be initiated to harass another person or entity, or be filed to further an ulterior motive (e.g. sham litigation in antitrust, in which improper use is made of the courts and other government adjudicative processes against rivals to achieve anti-competitive ends).

One element which may make the filing of frivolous suits more likely is the existence of incomplete information on *ex ante* behaviour (and thus on contractual and non-contractual liability). Contract disputes are more likely than many other kinds of litigation to exhibit (almost) complete information about liability. This may be especially the case when the contractual relationship involves periodic opportunities to monitor performance. This would mean that frivolous suit issues in the contract setting are unlikely to be as serious as in other litigation contexts. Moreover, such problems are likely to be even less important for long-term contract relationships, since relation-specific investments give parties incentives to act reasonably so as to maintain the relationship.

1.2. Theories on the decision to file suit/determinants

1.2.1. The basic model

When the plaintiff decides whether or not to file suit, he takes into account what may happen after filing: the parties may settle or they may go to trial.[6] A rational plaintiff

5 In the Netherlands, the loser at trial needs to reimburse a lump-sum part of the costs of the winner.

6 This section focuses on the relative optimism model. When the plaintiff considers filing a suit, he does not necessarily know whether he is more optimistic than the defendant. He only knows that this possibility exists. Here we consider a model in which the plaintiff finds out how the defendant estimates the plaintiff's chances of winning after a round of negotiations. Before the negotiations, the plaintiff only knows the distribution of defendant types (for example: he knows that half of the defendants will estimate his chances at 30 percent and the other half at 70 percent). From this distribution, he can make an estimation of the probability of trial.

will choose to bring suit when the weighed expected net benefits of settlement and trial exceed the costs of filing suit. The plaintiff will thus have to form estimations of the probability of trial, the probability of settlement, the expected pay-off in the case of a trial and the expected pay-off in the case of a settlement.

Formally, at the moment of the plaintiff's decision whether or not to file suit, his or her expected value equals:

$$EV = p_{tp}(p_{vp}J - C_{tp}) + (1 - p_{tp})(S - C_{sp}) - C_F$$

with $S = p_{vp}J - C_{tp} + \alpha((p_{vd}J + C_{td}) - (p_{vp}J - C_{tp}))$

with:
Ptp = the plaintiff's subjective probability that a complaint will eventually lead to a trial;[7] it follows that 1-Ptp is the plaintiff's subjective probability that a complaint will eventually lead to a settlement;
Pvp = the plaintiff's subjective probability of a victory at trial;
Pvd = the defendant's subjective probability of a plaintiff victory at trial;
J= amount at stake;
Ctp = the plaintiff's cost of a trial;
Ctd = the defendant's cost of a trial;
S= the settlement amount conditional on a settlement being reached;
Csp = the plaintiff's cost of settlement;
CF = the plaintiff's cost of filing;
α= the plaintiff's relative bargaining power ($0 \leq \alpha \leq 1$).

We can now examine the basic determinants of the decision to sue:[8]

- *The plaintiff's subjective probability of prevailing at trial*: There will be more suits when the plaintiff's estimation of his probability of success increases. First, the plaintiff's expected value from trial increases. Second, settlement amounts will increase as well.

7 The basic model does not allow for the plaintiff to drop his claim after some expenditures have been incurred. However, further on we discuss more advanced models (e.g. on sequential litigation), in which a claim can be dropped at some intermediate point.

8 We will discuss the influence of each determinant while holding all other determinants constant. For example, when we discuss the influence of the plaintiff's subjective estimation of his probability of prevailing at trial, we hold the probability of settlement constant. In reality, the probability of settlement will be influenced by the plaintiff's subjective estimation of his probability of prevailing at trial.

- *The defendant's estimation of the probability that the plaintiff will prevail:* There will be more suits when the defendant's estimation of the plaintiff's probability of success increases. The plaintiff's expected value from trial remains the same, but settlement amounts will increase because defendants have more to lose when they fear that there is a higher probability of being defeated in court.
- *The amount at stake:* There will be more suits when the amount at stake increases. First, the plaintiff's expected value of trial increases. Second, settlement amounts increase as well.
- *The trial costs of the plaintiff:* There will be fewer suits when the trial costs of the plaintiff increase. First, his expected value from trial decreases. Second, settlement amounts will decrease as well.
- *The trial costs of the defendant:* When the trial costs of the defendant increases, so does the settlement amount (unless the plaintiff has zero bargaining power)
- *The probability of settlement:* There will be more suits when the probability of settlement increases. The reason is that settlement amounts are always higher than the expected value of trial: plaintiffs obtain a fraction of the settlement surplus when they successfully reach a settlement agreement.[9]
- *The settlement amount:* logically, the incentive to sue increases when settlement amounts increase.
- *The settlement costs of the plaintiff:* There will be fewer suits when the settlement costs of the plaintiff increase, since the expected pay-off from settlement will decrease.
- *The costs of filing suit:* Logically, there will be fewer suits when the costs of filing increase.
- *The bargaining power of the plaintiff*: Settlement amounts increase when the plaintiff's bargaining power increases.

1.2.2. Extension 1: cost divisibility

In the previous section, it was assumed that once a case is filed, the parties get one opportunity to settle and that during trial, litigation costs are incurred all at once in a lump-sum fashion. Consequently, if and only if $p_{vp}J - C_{tp} > 0$ will the plaintiff have a credible threat of going to trial. In the opposite case, the defendant knows that the plaintiff will drop the case, and the defendant will consequently not be willing to offer any positive settlement amount. However, more subtle models lead to situations in which some plaintiffs with negative expected value suits are able to extract a positive settlement offer from a defendant, even when the defendant knows the plaintiff's suit has such a negative value. Bebchuk (1996, 1998) shows that the divisibility of the litigation process can provide such a plaintiff with a credible threat. In his model, litiga-

9 Unless the plaintiff has zero bargaining power.

tion costs are not incurred all at once in a lump-sum fashion but are spread over time, and bargaining is possible on various occasions. One can see the implications of divisibility with a simple numerical example. Consider a case in which the amount at stake equals 100 and the litigation costs of both the plaintiff and the defendant are 150 (so total trial expenses are 300). When there is only one litigation stage in which all of the parties' litigation expenses must be incurred, the plaintiff will not have a credible threat to proceed to trial. If the defendant refuses to settle, the plaintiff will drop his or her case. Anticipating this, the defendant will refuse to settle. Suppose now however that the litigation costs are incurred in two equal-cost stages (75-75 for each party) and that the parties can settle before the first stage (before the parties have spent any fraction of the costs) and also between the two stages (after each party has spent 75). Because of the sequential nature, we need to analyze this case by backward induction. Starting with the settlement negotiations between the two stages, conditional on the parties reach this round, the plaintiff will have a credible threat to proceed to trial: the cost of the first stage of litigation is sunk and going to trial would provide an expected judgment of 100 and involve an additional expense of only 75. So if the parties reach the bargaining round between the two stages, the plaintiff will have a credible threat and will be able to extract a positive settlement amount. Given an expected judgment of 100 and costs of 75 for each party in the second stage, the parties will settle somewhere in the range between 25 and 175. Suppose they have equal bargaining power. Then they will settle for an amount of 100. If we now go back in time to the bargaining round before the first litigation stage, it is easy to see that the plaintiff will have a credible threat to proceed to the next stage. Going through the first stage will cost the plaintiff 75 but will lead to receiving an amount of 100. Consequently, the plaintiff will be able to extract a settlement offer in the bargaining round preceding the first stage of litigation.

The divisibility of costs can thus lead to an increase in the number of suits. Moreover, the number of negative expected value suits which will receive a positive settlement offer increases with the number of stages. Returning to our example, suppose that the amount at stake drops from 100 to 60. With a division of the litigation process into two equal-cost stages (75 and 75), the plaintiff will not have a credible threat. He will not be offered a positive settlement amount. However, with a division into three equal-cost stages (50, 50 and 50), and once again assuming equal bargaining power, the plaintiff will succeed in extracting a positive settlement amount. More generally, Bebchuk (1998) shows that a finer division of the litigation process sometimes

improves but never worsens the plaintiff's ability to extract a settlement. Note however that divisibility cannot always provide a plaintiff with a negative expected value suit with a credible threat. For some cases, no matter how finely the litigation costs are divided, the plaintiff will never have a credible threat.

1.2.3. Extension 2: cost divisibility and gradual information flow

Even if the parties can settle only once and not also at some intermediate point, a plaintiff with a negative expected value suit can still extract a settlement offer when two conditions are met: (1) the plaintiff might obtain favourable information at an intermediate point which may turn the plaintiff's suit into a positive expected value suit, and (2) if the plaintiff does not obtain such favourable information, he can drop the suit and save his residual litigation costs. The plaintiff thus has the option of whether or not to progress with litigation. In contrast to the above theory on cost divisibility, here it is not assumed that the parties' estimate of the expected judgment is expected to remain below total litigation costs throughout the litigation. Once again, a numerical example can be illustrative. Suppose that the judgment can be either 0 or 1000, both with a 50 percent chance. Consequently, the expected judgment equals 500. Suppose further that the plaintiff's litigation costs are 600. The claim clearly has NEV (500<600). Suppose now that litigation is divided into two equal-cost stages (300 and 300), that the plaintiff learns the value of the judgment between the two stages (0 or 1000), and that the parties cannot bargain between the two stages, but that the plaintiff can drop the case after the first stage. Then the plaintiff will pursue litigation only if she learns that the judgment equals 1000, leaving her with a net value of 700. Since the probability that the judgment would be 1000 is 50 percent, this means that the plaintiff's expected value from litigation, before incurring the litigation costs in the first stage, is 350, and given that this is higher than the first stage costs of 300, his or her litigation threat is credible and she could extract a settlement offer.

Grundfest and Huang (2006) combine the models by Bebchuk (1996, 1997) and Cornell (1990). In their unifying model, bargaining between stages is possible, costs are divisible and information can be revealed throughout the litigation. Their analysis points out (1) the importance of the variance of the information revealed during the course of litigation. More precisely, they show that (2) when this variance is sufficiently large, every NEV lawsuit is credible. Moreover, they show that (3) risk-neutral plaintiffs can act as though they are risk-loving because increases in variance can increase a lawsuit's settlement option value. Next, (4) litigation settlement values can behave

quite differently from option values in financial markets. Finally, the model also shows (5) how important the sequence in which litigants incur expenses can be for the plaintiff's decision on whether or not to sue.

1. In the classical model, when litigants are risk-neutral, changes in the variance of the value of the information disclosed during the litigation process do not affect the settlement amount unless these changes also affect the lawsuit's expected value. In an options model, litigants may settle for amounts that can be far higher or lower than the expected value of the claim, even if that expected value is held constant. In an options model, variance is a critical determinant of the settlement value because the larger the variance, the more potentially valuable the information waiting to be disclosed during the course of litigation and the larger the value of the plaintiff's option to proceed with or abandon the litigation in light of that information. In other words, the variance reflects the value of the ability to adjust to new information independently of the parties' attitudes to risk.
2. It makes sense for a plaintiff to pursue any NEV claim if the cost of pursuing the claim is low enough and the possibility of uncovering some sort of "smoking gun" that will lead to a large recovery is big enough.
3. In a large category of cases, more uncertainty over the outcome of the litigation causes the plaintiff's claim to become more valuable. A rational defendant will be willing to pay more to settle the case for reasons that are independent of risk-aversion but have everything to do with the value of the plaintiff's option to adjust litigation expenditures in light of the revelation of new information.
4. The value of a financial option is generally a monotonically increasing function of the variance of the underlying instrument. Grundfest and Huang (2006) show however that settlement values can be discontinuous, non-monotonic functions of the variance of the lawsuit. When variance is sufficiently low, a lawsuit can have a settlement value that initially equals its expected value. As variance increases however, the settlement value can suddenly drop in a discontinuous way. But as variance continues to increase, settlements can climb to valuations much higher than the *ex ante* expected value of the lawsuit. Some NEV suits can have dead zones: regions of variance over which the lawsuit suddenly loses and then recovers credibility. The discontinuities and dead zones can have important consequences. Small differences as to a lawsuit's variance can cause significant differences in a lawsuit's settlement value, even when the parties agree about the expected value of the suit. This can cause the parties to spend litigation costs just to settle the

dispute after some uncertainty is resolved but prior to judgment. This finding does indeed find resonance in actual litigation: the plaintiff files a suit, expenses are incurred, and the litigants reach a settlement after some uncertainty is resolved.

5. The settlement value of a lawsuit can depend on the sequence in which the parties incur costs, even if the total value of the parties' expenditures is held constant and the sequence of investment is the same for both parties. If a plaintiff can defer a larger fraction of his litigation expenditures until a sufficient degree of uncertainty is resolved, then the value of the plaintiff's claim will increase. Intuitively, the plaintiff can commit to a relatively small investment before learning most of the information about the value of a lawsuit.

1.2.4. Extension 3: the case of sequential litigation[10]

A dispute often concerns multiple issues. For example, it is not unusual that parties disagree about a substantive issue *and also* about the court's jurisdiction.[11] Likewise, the plaintiff and the defendant may disagree over the liability issue *and* over the exact amount of damages. Also, the defendant may raise a counterclaim in response to the plaintiff's claim. The law can deal with multiple issues in basically two ways. First, it can dictate that courts should decide all issues at the same time. In other words, courts are then obliged to let the parties litigate all issues before making any decision. This can be called a "unitary trial" (Landes, 1993). Second, the law can allow courts to let the parties litigate one issue after the other and decide on each issue after the parties have litigated it. The parties thus litigate "issue 1" after which the court makes a decision about "issue 1". Then, if still necessary, the parties litigate "issue 2" after which the court makes a decision about "issue 2", and so forth. This can be called a "sequential trial" (Landes, 1993). In many countries courts have broad scope for separating issues at trial. Several articles of the Dutch Code of Civil Procedure give courts the possibility to separate issues at trial. For example, article 232, paragraph 1 of the Code gives courts the power to make intermediate judgments ("*tussenvonnissen*") before rendering a final judgment. In interlocutory decisions, a type of intermediate judgment, one can often find the court's view (explicitly or implicitly) on certain issues (e.g. the liability issue) and at the same time instructions for the delivery of evidence for further issues (e.g. the amount of damages) (Stein and Rueb, 2003). Sometimes, but not always, this view is expressed in the dictum (the decision is then a "*deelvonnis*"). According to article 110, paragraph 1 of the Dutch Code of Civil Procedure,

10 Note that sequential litigation is a special case which combines cost divisibility and a gradual flow of information. What is special about sequential litigation is that after each stage the parties know the court's precise views on certain issues.

11 Over the matter or over the defendant.

motions challenging the jurisdiction of the court must be filed before presenting the defence (Bosnak and Jonk, 2003). According to article 138,paragraph 1 of the Code, the parties litigate claims and counterclaims together and the court decides on both in the same decision, unless the court is of the opinion that either the claim or the counterclaim can be disposed of earlier. Likewise, the Federal Rules of Civil Procedure give U.S. courts various possibilities for sequential decision making. Motions contesting personal jurisdiction must be filed before an answer is filed or the answer itself must contain a denial of personal jurisdiction; otherwise the defendant is deemed to have waived any issue concerning the sufficiency of personal jurisdiction (see Rule 12) (Grubbs and DeCambra, 2003). Rule 42 gives the courts various powers to separate substantive issues: "The court, in furtherance of convenience or to avoid prejudice, or when separate trials will be conducive to expedition and economy, may order a separate trial of any claim, cross-claim, counterclaim, or third-party claim, or of any separate issue or of any number of claims, cross-claims, counterclaims, third-party claims, or issues, always preserving inviolate the right of trial by jury as declared by the Seventh Amendment to the Constitution or as given by a statute of the United States."[12]

An important question is how sequential litigation influences the behaviour of the parties. Landes (1993) presented the first systematic economic analysis of sequential litigation.[13] He showed that sequential trials:

1. Lower the expected cost of litigation for both parties compared with a unitary trial. The intuition behind this result is simple. In a sequential trial, the parties avoid litigation on subsequent issues if the defendant wins the current issue. For example, if the defendant is not held liable, the trial ends and the parties do not have to incur costs to prove the exact amount of damages.
2. Increase the plaintiff's incentive to sue and so increase the number of lawsuits. This follows directly from the cost savings discussed in (1). When suing becomes cheaper, more cases will be filed.
3. Reduce the likelihood that the parties will settle out of court by narrowing the range of mutually acceptable settlements. Once again, this follows directly from the cost savings discussed in (1). Since a trial becomes less costly for the plaintiff, he will ask for a higher settlement amount. And since a trial also becomes less costly for the defendant, he will only be willing to pay a smaller settlement amount. The settlement range thus gets smaller, resulting in more trials.

12 Unitary trials are the rule and sequential trials the exception for disputes that involve liability and damages (Landes, 1993).

13 For an earlier, less systematic analysis, see Schwartz (1967). For an analysis of sequential trials with one-sided private information, see Chen, Chien, and Chu (1997).

1.2.5. Extension 4: endogenous litigation expenditures

We would refer to the discussion of the trial stage for theories related to endogenous expenditures.

1.2.6. Extension 5: introducing time

The basic model described in Cooter and Rubinfeld (1989) does not take into account the element of time. This can be amended by making two changes: by introducing a discount factor which affects the amount at stake, and/or by introducing litigation costs which are a function of time (they increase e.g. linearly as time proceeds). We would refer to the section on modelling issues for further discussion.

1.2.7. Extension 6: risk

Most studies in the economic analysis of litigation make the simplifying assumption that litigants are risk-neutral. Due to the substantial variation in possible pay-offs from a claim, in practice the presence of a pay-off lottery rather than a certain reward should have important consequences, especially for litigants who cannot diversify their holdings. Risk-averse individuals generally prefer a certain, safe amount to an uncertain gamble that offers the same expected value. In other words, a risk-averse plaintiff does not value a trial at its expected value $E(x)=pJ-Cp$, but discounts the expected value of trial taking into account the accompanying risk. Technically, if $u(x)$ is the utility function of the plaintiff, risk-averseness means that $E(u(x)) < u(E(x))$. From the point of view of the plaintiff, the expected value needs to be discounted with a risk premium Rp. This risk premium is the additional amount that he or she would need to receive to make him or her indifferent between an uncertain adjudication with the average result $E(x)$ and a certain sum $E(x)$. The plaintiff is indifferent between settling for $E(x) - Rp$ and the risky trial. Two important questions are (a) how large the risk premium is in actual litigation and (2) how one can model risk-aversion. We come back to these issues in later sections.

1.2.8. Asymmetric information models

Bebchuk (1988) demonstrated that the presence of informational asymmetries can also explain the success of some NEV suits. The author (p. 445-446) derives the conditions under which a plaintiff with an NEV suit can extract a settlement offer due to informational asymmetry. Quite logically, the greater the defendant's expected litigation costs and the lower the probability attached by the defendant that the suit has NEV,

the more likely it becomes that the defendant will make a settlement offer.[14] For a detailed discussion of Bebchuk's screening model with NEV suits (and of a signalling model with NEV suits), we would refer to our discussion asymmetric information models in the settlement stage.

2. Consequences of filing suits on earlier stages

In several other sections of this project, we analyze the impact of various aspects of the litigation process on deterrence. For example, in the discussion of stage IV (the trial stage), we examine the impact of uncertain legal standards on *ex ante* incentives. Here, we focus on two basic but important issues: (1) the misalignment between the private and the social incentive to sue, and (2) the question of whether there are circumstances under which frivolous suits can improve social welfare.

2.1. The private versus the social incentive to sue

The economic theory of tort law and contract law is founded on the idea that the threat of (extra-contractual or contractual) liability provides potential injurers and contract breachers with efficient incentives (to take care to avoid accidents, not to breach a contract when breach is inefficient) by forcing them to internalize the risk that their behaviour creates. Early models in law and economics ignored the fact that liability is a private remedy that can only be imposed if victims of an accident or a contract breach are willing to sue and thus willing to spend resources. In other words, the early models did not shed light on the social value of lawsuits as a way to internalize harm.

Shavell (1982) was the first to show that there is no necessary alignment between the private and the social value of suit. The private value depends exclusively on the comparison a plaintiff makes of the expected payment at trial with the cost of suit. The social value on the other hand depends on the extent to which lawsuits induce the defendant to undertake socially desirable actions. Consequently, a suit may be privately but not socially valuable, and the reverse may also be true. There can thus either be too much or too little litigation from a social perspective.

We would stress that there can be too many suits, even given the fact that the threat of lawsuits generally underdeters injurers. The reason behind this last proposition is

14 Note that when the plaintiff's private information may be revealed through pre-trial discovery and disclosure and the defendants' costs of pre-trial discovery are sufficiently low, defendants may separate between NEV and PEV plaintiffs and make the litigation threat of NEV plaintiffs non-credible (Schwartz, 2003; Schwartz and Wickelgren, 2009).

that defendants do not face the full social cost of an accident or a contract breach for several reasons. First, defendants do not take into account the damages suffered by victims who do not file suit. Second, unless the loser at trial needs to pay for *all* the costs of the winner, defendants ignore (part of)[15] the filing and trial costs of victims who do file suit. Despite this underdeterrence, lawsuits may not be socially desirable.

When suits are socially desirable, the presence of underdeterrence can in principle be remedied either by subsidizing suits so that more victims file them, or by charging defendants damages in excess of the losses suffered by victims. Unfortunately, both instruments have the effect of increasing the number of suits, thus raising litigation costs. Total social costs may not be reduced. Another approach exists however, and can increase incentives for care without adding litigation costs: imposing a tax on defendants to be paid to the government. One could even improve incentives for defendants and reduce the number of lawsuits by simultaneously increasing the cost imposed on defendants and lowering the award to plaintiffs (Polinsky and Che, 1991). In practice however, policymakers may be limited to policies that award plaintiffs at least the level of their actual losses. More generally, it has been recognized that corrective policies are difficult to apply, because they require information that cannot easy be obtained by policymakers.

The discussion above focuses on a rule of strict liability. The analysis is slightly different under a rule of negligence. Under a perfectly functioning negligence rule,[16] if it is privately optimal for the potential injurer to comply with the due care standard, the first-best outcome can be achieved. No victims will file suit, so there is no risk of an excessive incentive to sue. If however it is privately optimal for the injurer to take less than due care, then the outcome is identical to that under the rule of strict liability.

In the discussions above, the possibility of settlement was ruled out. Several things change when settlement is possible. First, more suits will be filed because in cases that settle, the defendant overcompensates those plaintiffs whose damages plus filing costs are less than the amount the defendant ends up paying them.[17] Second, the possibility of settlement can either increase or decrease social costs. While settlement causes some cases to be filed that otherwise would not have been filed, it also allows some trial costs to be saved. Third, the possibility of settlement will tend to reduce the injurer's incentive to take care. The intuition behind this is that the defendant

15 Depending on the precise rule of cost allocation.

16 In reality of course, negligence rules never function perfectly.

17 Note that this is based on the assumption that the filing costs are smaller than the defendant's optimal settlement offer.

always has the option to take all cases to trial by not offering a settlement amount. Consequently, his costs cannot be higher when he has an option to settle. Fourth, with the possibility of settlement, the defendant can be either over- or underdeterred. As before, the defendant ignores the plaintiff's filing and trial costs, potentially leading to underdeterrence. However, in cases that settle, the defendant overcompensates some plaintiffs, which may lead to overdeterrence.[18] In conclusion, whether suits are socially more valuable with settlement than without, depends on whether the reduced incentive to take care is smaller than the (potential) decrease in social costs.

The theories above did not account for an essential social benefit of trial: its lawmaking function. Judges have the opportunity to assess existing rules and replace them with more efficient rules, which lower social costs by creating better incentives (e.g. for individuals considering breaching a contract or potential injurers deciding how much care to take to prevent an accident). Settlements of course do not perform this function. So if trials do indeed promote the selection of more efficient rules over time, they are socially valuable. Whether this selection actually occurs is debated in the economic literature. One possibility, originally advanced by Posner, is that judges consciously or unconsciously promote efficiency by selecting more efficient rules. This point of view has been criticized however because it depends on the motivation and incentives of judges, about which little is known. Other scholars have – more in line with the standard method in economics – sought to identify *a process* that could explain that judge-made law evolves towards efficiency. Several models of efficient legal evolution were created (Rubin, 1977; Priest, 1977). The essence of these models is that inefficient laws will be litigated more frequently than efficient laws because the former impose larger costs on victims. Consequently, inefficient laws will come before the court for re-examination more often, resulting in a trend toward efficiency. Other scholars soon began to critically examine these models. For example, Landes and Posner (1979) argued that litigation might strengthen or weaken a precedent without overturning it completely and showed that this consideration weakens the evolutionary pressures for efficiency. Hirshleifer (1982) showed that the law could come to favour whichever party can most easily organize and mobilize resources for litigation of unfavourable precedents and that this movement would be independent of efficiency. Zywicki (2003) added a "supply side" evolutionary model (i.e. competition between several court systems) to the existing "demand side" evolutionary models (i.e. litigants demanding efficient rules). He argued that when competition is lacking, only weak or even non-existing tendencies for efficiency will exist. More recent

18 Miceli (2009) notes that underdeterrence seems more plausible than overdeterrence.

literature seeks to combine the selective litigation argument and biased judges. Gennaioli and Schleifer (2007a and 2007b) assume that judges are principally efficiency-seeking, but that they may hold a bias for one or the other side in a dispute. They examine the direction of legal change when judges are bound by precedent (in which case legal change can only occur if judges distinguish their decision from existing precedent) and when judges are not bound by precedent (they can change the law directly by overruling previous decisions). In the first case, judicial bias distorts the law away from efficiency, but it can improve the precision of law in the long run. In the second case, they find that selective litigation cannot counteract the detrimental effect of judicial bias when the bias is strong enough. Miceli (2010) criticizes this result however because Gennaioli and Shleifer assume that the divergence of beliefs by litigants about trial is unrelated to the nature of judicial bias. This treatment of the impact of selective litigation does not allow a careful examination of its interaction with judicial bias in determining the direction of legal change. Miceli (2010) considers a unilateral care accident model in which potential injurers and victims interact a fixed number of times over a given time interval and each interaction is a potential accident. Injurers choose a level of care in anticipation of each interaction. There are two rules for allocating liability: strict liability and no liability. Initially, both rules exist in the population of legal rules in some arbitrary proportion. Each potential accident involves just one of these rules. Both parties know with certainty the prevailing rule that applies to their particular interaction. Judges have biases (either pro-plaintiff or pro-defendant) and are imperfectly bound by precedent, so they will occasionally overturn laws based on their preferences. The probability that a judge will overrule a precedent is exogenously determined. Since there are only two rules, overruling means replacing strict liability with no liability and vice versa. Consequently, as time unfolds, the process of litigation will cause the distribution of legal rules to evolve as cases come before the court to be adjudicated. One of Miceli's main goals is to find the ultimate distribution of legal rules (more formally, the steady state equilibrium distribution). Miceli finds that the ultimate proportion of strict liability rules depends on the relative number of trials that arise under each of the two rules, and on the distribution of judicial bias. Logically, a larger fraction of pro-defendant judges leads to a smaller ultimate proportion of strict liability rules. Likewise, a larger relative number of trials under strict liability leads to a smaller ultimate proportion of strict liability rules. Intuitively, the more often a rule comes to trial, the more chances it has to be overturned by a judge and replaced by the other rule. The strength of precedent affects the rate of legal change (how long it takes to reach the steady state), but not the

long-run distribution of legal rules. Only in some special cases will the law converge to a single rule (to strict liability when all judges are pro-plaintiff and to no liability when all judges are pro-defendant). Generally, the common law will embody multiple rules, reflecting the biases of judges and the selection of cases for trial.

2.2. The impact of frivolous suits on deterrence

Stone and Miceli (2013) examine whether there are circumstances under which frivolous suits may increase social welfare. They start with the observation that there is generally a problem of underdeterrence in tort law,[19] because potential injurers do not fully internalize two social costs: the litigation costs of the plaintiffs they harm (unless the winner at trial is compensated by the loser for all his costs) and the magnitude of the harm suffered by victims who do not file suit because the costs outweigh the benefits. They focus on frivolous lawsuits that "piggyback" on genuine claims (and thus not on lawsuits which are initiated to simply harass the defendant, or which are filed to further an ulterior motive). These piggyback lawsuits are either brought by (1) actual accident victims whose injuries were caused by someone other than the defendant, or (2) actual accident victims whose injuries were caused by "nature", or (3) uninjured plaintiffs. Stone and Miceli assume that there is asymmetric information with respect to the quality of the plaintiff's claim, with the plaintiff being the informed player. Due to this asymmetric information, defendants face a dilemma when formulating a settlement offer. When they try to settle all cases, they allow piggyback plaintiffs to obtain positive settlements, which increases total settlement costs for the defendant. If however defendants litigate all cases, piggyback plaintiffs get nothing, but defendants incur litigation costs with genuine plaintiffs. The authors show that this leads to two distinct equilibria. The first is a pure strategy equilibrium in which all piggyback victims file suit and the defendant settles with all plaintiffs. This equilibrium emerges when the relative number of piggyback plaintiffs in the population of all plaintiffs is sufficiently small. The second is a mixed strategy equilibrium in which only a fraction of piggyback victims file suit, and the defendant settles a fraction of these cases. This second equilibrium emerges when the fraction of piggyback plaintiffs is relatively large. Thus, in both equilibria, at least some piggyback plaintiffs obtain settlements. Now it is easy to see the influence on the injurer's activity level and choice of care. With respect to the former, the injurer will choose a lower activity level compared with a situation without frivolous suits because (1) as an injurer's activity level increases, the number of piggyback plaintiffs also increases, and at least some

19 On a similar note, see also Ordover (1978) and Hylton, (1991).

succeed in receiving settlements, and (2) in the second equilibrium, the average liability per suit is higher because some cases go to trial. Consequently, the cost per unit of activity is higher. The resulting reduction in activity level may or may not enhance the deviation from the social optimum. The injurer may (still) over-engage because he does not internalize a genuine victim's filing cost or does not fully internalize a genuine victim's damages when settling. But the injurer may also under-engage due to the costs of paying off piggyback plaintiffs or because the injurer incurs litigation costs for those cases filed by genuine plaintiffs that end up going to trial. With respect to the level of care, the injurer will choose a higher level of care when frivolous suits are possible, at least in the second equilibrium. The reason is that by exercising additional care, the injurer reduces the frequency of cases that end up at trial with genuine plaintiffs. Whether this leads to an improvement from a social perspective is once again ambiguous, for the same reasons related to the activity level discussed above.

3. Modelling issues

3.1. Private versus social incentive to sue

A model which compares the private and the social incentive to sue can shed light on the key parameters that empirical research needs to take into account when trying to estimate whether policymakers should stimulate lawsuits or not. Miceli (2009) offers a tractable model. In this model, a potential injurer can invest x (dollars, euros) in care to reduce the probability of an accident $p(x)$, with $p'<0$ and $p''>0$. In the case of an accident, the victim suffers a loss of L. At the time of the accident, this loss is only observable to the victim. The injurer only knows the distribution function of the loss conditional on an accident, $F(L)$. The plaintiff can file a suit at cost k, and the trial costs of the parties equal C_p and C_d respectively. Liability is strict and all cases go to trial (in a later section we examine what changes when settlement is allowed).

Once an accident occurs, suit will be privately valuable for the plaintiff if $F(L) \geq k + C_p$. Consequently, the probability of a suit equals $1 - F(k + C_p)$. Now we look at the social value of suits. This involves several steps. In the case of an accident, the defendant's expected loss equals:

$$A=\left(1-F\left(C_p+k\right)\right)E\left(L+C_d \text{ , given that } L\geq k+C_p\right)=\int_{k+C_p}^{\infty}(L+C_d)dF(L)$$

Given this expected loss in the case of an accident, the defendant will choose a level of care x to minimize $x + p(x)A$. The first-order condition equals $1 + p'(x)A = 0$. Solving it delivers the injurer's optimal care x°.

The expected social costs conditional on an accident are:

$$H = E(L) + \Big(1 - F\big(k + C_p\big)\Big)\big(k + C_p + C_d + C_{soc}\big)$$

with E(L) being the plaintiff's expected loss in the event of an accident and Csoc the costs associated with a trial which the parties do not bear, but society has to pay for.

We have previously discussed the fact that lawsuits underdeter injurers. We can indeed see that A < H.

The expected social costs, evaluated at the defendant's privately optimal choice of care, are $x^\circ + p(x^\circ)H$.

If lawsuits were however prohibited, no victims would file and injurers would take no care. In that case, expected social costs would be p(0)E(L).

Lawsuits are socially valuable if $x^\circ + p(x^\circ)H < p(0)E(L)$, thus if:

$$p(x^\circ)\Big(1 - F\big(k + C_p\big)\Big)\big(k + C_p + C_d + C_{soc}\big) < p(0)E(L) - (x^\circ + p(x^\circ)E(L))$$

The left-hand side are the expected litigation costs of allowing lawsuits, the right-hand side represents the benefits of lawsuits.

This inequality shows the several elements empirical researchers need information on to be able to determine whether a given volume of suit is socially valuable or not, and what the optimal level of suit would be:

1. the distribution of the losses for the plaintiffs in the case of an accident: this determines, together with the plaintiff's filing and trial costs, the volume of suits;
2. the total costs of litigation (filing costs and the trial costs of the plaintiff and defendant and the costs society pays for);
3. the prevention costs of the defendant;

4. the accident technology, $p(x)$. This may be the most difficult element to obtain information on. As far as we were able to check, there are no empirical studies estimating the function $f(x)$ for specific types of accidents.

3.2. Modelling the impact of frivolous suits on *ex ante* incentives

As discussed above, Stone and Miceli (2013) examine the influence of frivolous suits on the defendant's activity level and his care level. From a modelling perspective, it is interesting to see how frivolous suits are introduced. Stone and Miceli introduce q, which represents the exogenous probability of a piggyback suit being filed per unit of the activity. The other notations are:
z = level of the risky activity in which the injurer (defendant) engages;
$V(z)$ = gross value of the activity, where $V(0)=0$, $V'>0$, and $V''<0$;
x = dollar spending on care by the injurer per unit of the activity;
$p(x)$ = probability of an accident per unit of the activity, where $p'<0$ and $p''>0$;
L = harm suffered by a genuine victim in the event of an accident;
k = victim's cost of filing suit;
$C\pi$ = cost of a trial for victims (plaintiffs);
$C\Delta$ = cost of a trial for the defendant;
S = settlement amount.

As discussed above, Stone and Miceli find two possible equilibria. The injurer's choice of care x and activity z will depend on which type of equilibrium he expects to arise in the settlement game. In the equilibrium in which all piggyback lawsuits are filed and all cases settle (for the genuine plaintiff's expected value of trial), the injurer seeks to maximize:

$$V(z) - z[x+(p(x)+q)(L-C\pi)]$$

In the second equilibrium, in which all genuine plaintiffs and a fraction of piggyback plaintiffs file suit and the defendants offer no settlement amount to a fraction of plaintiffs, Stone and Miceli show that the defendant's expected costs are equivalent to the cost he would incur if only genuine plaintiffs filed suit and all went to trial. Thus, the injurer seeks to maximize:

$$V(z) - z[x+p(x)(L+C\Delta)]$$

The socially optimal care and activity choices are the ones which would maximize the objective function of a social planner:

$$V(z) - z[x + p(x)(L+k)]$$

Note that only the filing costs play a role here, but not the trial costs of the plaintiff and the defendant. The reason is that, assuming that the defendant can perfectly distinguish between genuine and piggyback suits, the defendantwould settle with all genuine plaintiffs, and no piggyback plaintiffs would file suit.

3.3. Modelling risk

The way risk is modelled will crucially depend on the data available with respect to plaintiff wealth. Because data related to plaintiff wealth are often lacking, models frequently work with the assumption of constant risk aversion, and the insurance premium (or risk premium) R is modelled as proportional to the variance of court awards[20] and to the Pratt-Arrow measure of absolute risk-aversion. Viscusi (1988) uses the formula $R \approx (1/2)\sigma^2 A$. In reality, the measure of absolute risk-aversion might be decreasing rather than constant, so the willingness to sue might increase with wealth. If however there is no substantial heterogeneity in the wealth of the plaintiffs, the assumption of constant absolute risk-aversion will provide a good approximation of what would be found with a wealth-dependent measure.[21][22]

The simple representation of the plaintiff's risk premium captures two fundamental properties of risk-aversion. First, if litigation risk rises (thus if the variance of the trial outcome increases), then the plaintiff's risk premium rises as well. Second, the risk premium rises disproportionally as the stakes at trial increase. This property follows from the fact that variance hinges on the squared differences.[23]

20 If the court award is represented by x, than $\sigma^2=var(x)=E(x-E(x))^2$.

21 The constant risk aversion assumption implies that the utility function of the plaintiff has the form $-be^{-dX}$, where X is the plaintiff's asset level (Keeney Raiffa,1976).

22 Also, if the pay-offs are distributed lognormally, and if preferences do not satisfy constant absolute risk aversion but constant relative risk aversion – thus allowing wealth-related variations in risk aversion – then one finds a similar formulation.

23 $\sigma^2 = var(x) = E(x - E(x))^2$; $var(2x) = E(2x - E(2x))^2 = 4E(x - E(x))^2 = 4var(x)$.

3.4. Modelling case quality and *ex ante* decisions

Endogenous expenditure theories have the advantage that they explicitly model the quality of the case (F).[24] A larger F means a stronger case. However, the literature generally does not explicitly address the relationship between the defendant's *ex ante* behaviour (e.g. did he perform his contractual obligations correctly? Did he take adequate care to prevent an accident?) and the quality of the case. With respect to F, where do we draw the line between a case in which the defendant behaved in an *ex ante* efficient manner, and a case in which he did not? De Mot and Miceli (2015) solve this issue as follows in the context of the decision to take care: if the defendant took just enough care (e.g. to respect his contractual duties; to avoid an accident etc.), F is equal to 1/2. F decreases below 1/2 as the defendant took more than due care. Conversely, F increases above 1/2 as the defendant took less than due care (or when the defendant intentionally caused the harm). By relying on this insight, one can make more concrete assessments of how the defendant's *ex ante* decisions will influence his pay-offs at trial.

One may argue that when F is only slightly larger than 1/2, the endogenous models predict a win rate of "only" slightly more than 50 percent for the plaintiff, at least when the unit costs of the parties are the same. However, in reality, it is more plausible that when the plaintiff has quite a strong case (F is considerably larger than 1/2), his unit costs will be relatively low compared with the unit costs of the defendant. The reason is that the litigant with the truth on his side often[25] finds it easier to gather evidence in his favour than the party who does not have the truth on his side. In other words, on average, there is an inverse relationship between F and Cp/Cd. This has an important consequence: if the defendant took just enough care, the probability of a plaintiff victory will generally be considerably lower than 50 percent; if he failed to take sufficient care, that probability will generally be considerably higher than 50.

24 Note that models with exogenous trial effort can also make the distinction between weak and strong claims (e.g. weak claim: the defendant took due care; strong claim: the defendant took no care; probabilities of plaintiff victory can be derived by introducing fixed probabilities of type 1 and type 2 errors). However, such models have the disadvantage that they do not recognize that trial effort varies with the amount at stake and the merits of the case.

25 But not always. Even a plaintiff with a very strong case (e.g. the defendant took much too little care) may encounter evidentiary problems.

3.5. Modelling cost divisibility and gradual information flow

Here we sketch the contours of a model with two stages which takes into account cost divisibility, gradual information flow, the possibility to settle before each stage and the possibility to drop the case before each stage. As discussed before,[26] such a model can shed light on the importance of the variance of the court award. In the model, each litigant is risk-neutral and tries to maximize their expected wealth. The litigants confront one uncertain variable (e.g. a third-party witness testimony; a court's choice between two potential interpretations of a statute). In stage 1, the court resolves this uncertainty, but only after the parties have incurred trial costs (e.g. discovery costs; legal research costs). We now give a more detailed overview of the various stages.

Before stage 1: The plaintiff decides whether or not to file the suit. If the suit is filed, the parties either settle or go to stage 1.

Stage 1: The plaintiff spends an amount Cp1 to initiate the lawsuit, the defendant is forced to spend $Cd1$ to defend the lawsuit. If he fails to do so, he will incur a default judgment with consequences more severe than the litigation costs of stage 1. At the end of stage 1, additional information related to the litigation is revealed to both parties (e.g. through a third-party witness testimony).

Between stage 1 and stage 2: After information has been revealed in stage 1, the plaintiff has the option to abandon the litigation. If the plaintiff decides to continue, either the parties settle and thereby avoid the trial costs of stage 2, or they do not.

Stage 2: The plaintiff incurs a cost of $Cp2$ and the defendant a cost of $Cd2$. If the defendant does not incur this expense, once again he will incur a default judgment with consequences more severe than the litigation costs of stage 2. At the end of this stage, the court announces its verdict.

Let further J be the parties' expectations of the initial expected value of the court award if the lawsuit is pursued to its conclusion. At the end of stage 1, uncertainty will be resolved. With a probability p $(0<p<1)$, at the end of stage 1 the parties will know that the court will ultimately grant an amount of $J1$, and with probability $1\text{-}p$, an amount of $J2$, with $J1>J2$. Since information revealed at the end of stage 1 is constrained to have an initial expected value J, we can write that $J=pJ1+(1\text{-}p)J2$.

26 See the discussion of Grundfest and Huang (2006).

Finally, given the possibility for settlement at various occasions, one needs to introduce a parameter (e.g. β) representing the relative bargaining power of the parties. This makes it possible to pinpoint the exact settlement amounts.

This model can then be solved by using backward induction.

3.6. Modelling time

As we mentioned before, the basic model described in Cooter and Rubinfeld (1989) does not take into account the element of time, and this can be amended by making two changes. We now look at this from a slightly more formal perspective:

1. It can be amended by introducing a discount factor which affects the amount at stake (e.g. e^{-rt}. In that case, the expected judgment will become $e^{-rt}\mathrm{J}$ for both parties. Of course, if it is more likely that the plaintiff and the defendant weigh the future differently, subjective discount factors can be introduced.
2. It can also be amended by specifying that the litigation costs of the parties are a (e.g. linear) function of time. For example: one could assume that as times passes, litigation consumes an increasingly large part of the amount at stake: $\mathrm{Cp} = \alpha\mathrm{tJ}$ and $\mathrm{Cd}=\beta\mathrm{tJ}$, with α and β being a positive parameter. The time variable can be standardized ($t \in [0; 1]$) in order to avoid that different time unit measurements would have different effects.

Introducing these changes in the plaintiff's expected pay-off from a trial shows that his pay-off is a decreasing function of the length of a trial. When t reaches a critical value, the plaintiff's expected value becomes negative.

4. Empirical issues

4.1. Empirical studies on the number of suits relative to the number of legal problems

With respect to the United States, Trubek *et al.* (1983) estimated that 90 percent of civil disputes are settled without a suit being officially filed. Dutch empirical research conducted in 2003 shows that about two out of three Dutch citizens encountered a legal problem during the previous five years. In almost half of the cases (48 percent), the parties reached an agreement themselves and in 7 percent of cases a third party made a decision. In 35 percent there was no agreement or a decision of a third party, though some form of action was undertaken initially. Finally, in 10 percent no action at all was undertaken to solve the legal problem (Van Velthoven and Voert 2003).

4.2. Empirical studies on the volume of frivolous suits

From an empirical perspective, there is little reliable evidence that frivolous suits are a serious problem. Of course, empirical work faces some fundamental obstacles in this field. First, suits often settle and settlements are often kept confidential. Second, even if lawyers sometimes were to file frivolous suits, they would not be keen on admitting it. Third, the abovementioned lack of a common view on exactly what constitutes a frivolous suit further complicates matters. A very well known empirical study of medical malpractice litigation refused to draw any conclusions about frivolous litigation partly because of difficulties in determining which suits were frivolous (Harvard Medical Practice Study, 1990). In some areas of the law however, empirical studies do exist. Some empirical work relies on surveys of judges and lawyers. These surveys are of course sensitive to problems of respondent bias and framing, and cannot be considered as strong evidence. Several studies conclude that most judges do not find groundless litigation to be a problem in counselled cases (Wiggins *et al.*, 1991).[27]

We have mentioned before that frivolous suit issues in the contract setting are less likely to be as serious as in other litigation contexts. These predictions are in accord-

27 According to a Harris survey, 21% of U.S. federal judges and 14% of state judges believe frivolous litigation is a "major cause" of delays (Harris, 1989).

ance with the results of one American empirical study that found less Rule 11 activity[28] in contract cases than in any other case types (Marshall, *et al.*, 1992).

4.3 Empirical studies on the influence of costs on the quality of filed cases

Gross and Syverud (1991) studied 529 California civil jury trial cases in different areas of the law. They observed settlement offers and judgments for cases that go to trial. They find that when plaintiffs pay their own litigation costs, the plaintiff win rate at trial increases, showing that higher quality suits are brought on average when plaintiffs bear their own costs of filing (and trial).

4.4. Empirical studies on risk-aversion

Viscusi (1988) empirically examined the role of risk-aversion on the decision to drop the suit, on the decision to settle and on the settlement amount, by studying a sample of more than 10,000 closed product liability cases coming from 23 insurance companies and covering all of the 50 states in the U.S. The data were gathered through the Insurance Services Office (ISO) Product Liability Closed Claims Survey. Viscusi finds that risk-aversion is larger for plaintiffs than for defendants, but also that risk-aversion does not appear to be overly weighty. The expected rewards variable is dominant. Depending on the particular issue involved (the decision to drop the suit, the decision on whether or not to settle, and the settlement amount), the expected award had an effect 2 to 9 times greater than the effect of the variance in awards. Specifically, with respect to the decision to drop the suit, the effect of the expected reward is more than two times as great as the effect of the variance in awards.

28 In the United States, Rule 11 of the Federal Rules of Civil Procedure (and similar state rules) require that an attorney performs a due diligence investigation of the factual basis for any claim or defence. Sanctions may be imposed by a court upon the party or the lawyer who presents the frivolous defence or claim.

3

Stage III. The settlement stage

A. THE SETTLEMENT RATE

1. General

1.1. Description

After the plaintiff has asserted his legal claim (officially or informally), the parties may either settle the dispute or go to trial. The parties reach a settlement when funds are transferred from the defendant to the plaintiff without the involvement of the courts.

1.2. Causes of settlement failure

1.2.1. The basics of the relative optimism model

In the basic model of litigation (Landes, 1971; Posner, 1973; Gould, 1973), the parties to a dispute can choose between settling the matter themselves and having the case decided by a court. Clearly, a settlement has the advantage that it saves the costs of litigating the case. Put otherwise, settling the case generates a surplus for the parties in the form of saved litigation costs.[1]

When the parties have the same beliefs about the likely outcome of the trial, it is easy to see that the parties will always settle (unless they cannot agree on the division of the settlement surplus). The minimum settlement amount the plaintiff wants to receive (his "threat point") is what he expects to gain from litigating the case. This is equal to

1 One can assume zero settlement costs for simplicity. What matters is the difference between the trial costs and the settlement costs and not their absolute values. So in what follows, "trial costs" should be considered as trial costs minus settlement costs. The influence of settlement costs thus follows in a straightforward manner.

the expected judgment[2] minus his trial costs. The maximum settlement amount the defendant is willing to give (his "threat point") is what he expects to lose if it comes to a trial. This is equal to the expected judgment plus his trial costs. Since the parties have the same beliefs about the outcome of the trial, the maximum the defendant is willing to give is larger than the minimum the plaintiff is willing to accept. Formally, when Expected Judgmentpl = Expected Judgmentdef, then:

Expected Judgmentpl – Trial Costspl Expected Judgmentdef + Trial Costsdef

A fortiori, the parties will always settle when the defendant is more convinced than the plaintiff that the plaintiff will win at trial (this is often called "relative pessimism"). Formally, when Expected Judgmentpl ≤ Expected Judgmentdef, then:

Expected Judgmentpl – Trial Costspl ≤ Expected Judgmentdef + Trial Costsdef

Results may change when the plaintiff is more convinced than the defendant that he will win the trial (this is often called "relative optimism"). In that case, the plaintiff's estimate of the expected judgment is greater than the defendant's estimate of the expected judgment. Formally, when Expected Judgmentpl > Expected Judgmentdef, the parties settle if:

Expected Judgmentpl – Trial Costspl ≤ Expected Judgmentdef + Trial Costsdef

which can be rewritten as:

(Trial Costspl + Trial Costsdef) – (Expected Judgmentpl – Expected Judgmentdef) ≥ 0

or as:

Settlement Surplus ≥ 0

In words, when the plaintiff is more optimistic than the defendant, the parties settle if the sum of the trial costs exceeds or equals the difference between the parties' estimation of the expected judgment.

2 The expected judgment equals the plaintiff's probability of winning multiplied by the amount at stake.

1.2.2. Determinants of the settlement frequency in the relative optimism model

We can deduce the determinants and their influence from the discussion above:

The amount at stake: The parties will settle less often when the amount at stake increases. The plaintiff will want to receive a higher settlement amount because he expects a larger pay-off at trial. The defendant will be prepared to make a larger settlement offer, because he expects to lose more at trial. But since the plaintiff estimates his chances of winning as higher than the defendant's chances, the additional amount the plaintiff wants is larger than the additional amount the defendant is willing to give. Consequently the settlement surplus decreases when the amount at stake increases.[3]

The difference between the parties' estimates that the defendant will be held liable: When the difference between the parties' estimates increases, the parties will settle less often. Obviously, the divergence between the minimum amount the plaintiff is willing to accept and the maximum offer the defendant is willing to make, increases.[4]

3 For example, suppose the plaintiff thinks he has a chance to win in court of 80 percent. The defendant thinks the plaintiff only has a chance of 70 percent to win. The trial costs of the plaintiff and the defendant are 10 for each. If the amount at stake equals 100, the expected value of the trial for the plaintiff is 0.8x100 – 10 = 70. The expected loss of the defendant is 0.7x100 + 10 = 80. The parties will settle because the maximum the defendant is willing to pay is larger than the minimum the plaintiff is willing to accept. In other words, there is a settlement surplus (of 10). What happens now if the amount at stake is not 100 but 300? Will the parties still settle? The answer is no. The expected value for the plaintiff is now 0.8x300 – 10 = 230. The expected loss of the defendant is 0.7x300 + 10 = 220. The settlement surplus has vanished. Although the defendant is willing to pay more (220 instead of 80, this is an increase of 140), the plaintiff still wants more due to the fact that he estimates his chances of winning as being higher than the defendant's chances (from 70 to 230, this is an increase of 160).

4 For example, suppose the amount at stake equals 100 and the trial costs of the plaintiff and the defendant are 10 for each. If the difference in opinion between the parties equals 10 percent (e.g. 80 versus 70 percent), the parties will settle. The expected value of the trial for the plaintiff equals 0.8x100 – 10 = 70 and the expected loss of the defendant equals 0.7x100 + 10 = 80. There is a settlement surplus. If the difference in opinion now equals 30 percent (e.g. 80 versus 50 percent), the parties will not settle. The expected value for the plaintiff equals 0.8x100 – 10 = 70 and the expected loss of the defendant equals 0.5x100 + 10 = 60. There is no longer a settlement surplus.

The trial costs of the parties: When the total trial costs of the parties increase, the parties will settle more often.[5] The reason is very straightforward. Parties have an incentive to settle precisely because they want to avoid the costs of trial (see above). When these costs increase, the settlement surplus increases.[6]

1.2.3. Criticism of the relative optimism model

1. The parties settle in the relative optimism model whenever there is a settlement surplus. This model tells us little about the actual settlement amount the parties are expected to agree on (Shavell, 2004). Any amount between the threat point of the plaintiff and the threat point of the defendant is theoretically possible. This ambiguity can theoretically be solved in different ways: by introducing a parameter for the relative bargaining power of the parties, by looking for bargaining solutions which satisfy some criteria we suspect any solution should satisfy, and by introducing specific bargaining structures (e.g. the plaintiff is allowed to make a take-it-or-leave-it proposal).
2. In the relative optimism model, the parties may have different estimates about the outcome of the trial. At the same time, the plaintiff knows the defendant's estimate and the defendant knows the plaintiff's estimate[7] and the parties know that the other party is making a sincere estimate. According to some economists, such "agreeing to disagree" is in contradiction with rational choice theory (Aumann, 1976). A rational party will revise his optimistic assessment downward when he learns how optimistic his adversary is. The parties' beliefs should converge until a settlement is possible. Intuitively, the optimism of the other party conveys information and each party should place equal weight on the evidence they have per-

5 What is important are the total trial costs. Suppose that the trial costs of the plaintiff increase and those of the defendant decrease and that the increase is larger than the decrease. Then the settlement surplus will increase unambiguously because the *total* trial costs have increased.

6 Suppose the amount at stake is 100, the plaintiff estimates he has a probability of winning of 80 percent and the defendant estimates that the plaintiff has a probability of winning of 50 percent. If the trial costs of the parties are 10 for each, the parties will not settle. The expected value for the plaintiff equals 0.8x100 – 10 = 70. The expected loss of the defendant equals 0.5x100 + 10 = 60. There is no settlement surplus. If the trial costs of the parties are 20 for each, the parties will settle. The expected value for the plaintiff decreases to 0.8x100 – 20 = 60 and the expected loss of the plaintiff increases to 0.5x100 + 20 = 70. Now there is a settlement surplus.

7 More precisely, these estimates are assumed to be common knowledge.

sonally gathered *and* on the evidence the opponent has gathered.[8] However, several authors have argued that "agreeing to disagree" is not necessarily in contradiction with rationality when the initial beliefs of the parties are different (e.g. Gilboa (1992)).

3. Most authors who rely on relative optimism do not consider how the divergences between the parties arise (Shavell, 2004). They simply assume that the divergent expectations are exogenously determined. There are at least three different explanations for the divergences between the parties. First, the beliefs of the parties may be seen as being affected by randomly distributed, unbiased errors of interpretation when the parties weigh up the strength of the case (Priest and Klein, 1984). Second, it is also possible that parties have a tendency to interpret information in a "self-serving" manner that is favourable to them (Babcock and Loewenstein, 1997; Farmer and Pecorino, 2002). Third, divergences can be the result of the parties having different information. This last possibility has been examined very thoroughly in the literature. We will discuss it in detail below.

1.2.4. The basics of the asymmetric information model(s)

A. What does it mean that "parties have asymmetric information" and why does asymmetric information cause trials?

Parties have asymmetric information when they have different information from each other.[9] For example: the plaintiff may know the exact amount of damages while the defendant only knows that the damage is situated between 1,000 and 10,000 euros; the defendant may know whether he acted negligently while the plaintiff only knows that there is a 50 percent chance that the defendant acted negligently; the plaintiff may know how much a trial will cost him while the defendant only knows that the plaintiff's costs are situated between 2,000 and 5,000 euros; the plaintiff who suffered bodily injuries may know his costs associated with settlement delay while the defend-

8 A simplified example can make this clearer. Suppose the plaintiff and the defendant originally think that the plaintiff has a 50 percent probability of winning (they have the same "priors"). After the plaintiff talks to witness number one and the defendant talks to witness number two, the plaintiff now thinks he has an 80 percent probability of winning and the defendant thinks the plaintiff has a 60 percent probability of winning. After talking to the witnesses, they share their beliefs with each other. Then both parties will figure out the information the other party received from the new beliefs of the other party. Each party will use this information to update his own beliefs. Ultimately, the parties will have the same beliefs.

9 It is not so much the fact that one party has better information than the other but rather the fact that parties have different information that is important.

ant only knows that 80 percent of plaintiffs would prefer a settlement offer of 100,000 euros now over a settlement offer of 120,000 euros in one year's time etc.[10]

It is easy to see that asymmetric information can cause trials. If the uninformed party were never willing to go to trial, the informed party who is of a "bad type" would always have an incentive to misrepresent himself as a "good type" during settlement negotiations.[11] The reason is simple: the pay-off for a "bad type" is higher when he successfully misrepresents himself as a "good type". For example, a plaintiff who has suffered little harm ("bad type") will try to be mistaken for a plaintiff who has suffered a lot of harm ("good type"); a defendant who has a high chance of being found liable ("bad type") has an incentive to misrepresent himself as a defendant who has a small chance of being found liable ("good type"); a plaintiff with high trial costs ("bad type") can have an incentive to misrepresent himself as a plaintiff with low trial costs ("good type") etc. Consequently, an uninformed plaintiff who will never go to court will always receive the smallest possible settlement amount and an uninformed defendant who does not want to risk litigation will always have to offer the maximum possible settlement amount. Clearly, this cannot always be the best strategy for the uninformed party. Rational settlement behaviour by the uninformed party may create a chance of trial to prevent misrepresentation.[12]

10 We will discuss the several types of asymmetric information models in more detail later on.

11 This is called a "pooling equilibrium" in game theory.

12 The following example serves as an illustration. Suppose the plaintiff is privately informed about the exact amount of damages he suffered: 1,000 or 10,000. The defendant only knows that half of the plaintiffs suffered damage of 1,000 and the other half damage of 10,000. The plaintiff can make a take-it-or-leave-it proposal. The trial costs of the plaintiff and the defendant equal 500 for each. Consequently, each type of plaintiff will go to trial if the defendant refuses the settlement proposal (1,000-500 = 500>0 and 10,000-500=9,500 >0). If the defendant were never prepared to take the plaintiff to trial, the plaintiffs who have suffered harm of 1,000 will claim they have suffered harm of 10,000. All plaintiffs will make a settlement proposal of 10,500 (damages plus trial costs of the defendant). Since the defendant does not want to go to court, he will accept the proposal (if he were not to accept it, there is a 50 percent risk of trial since half of the plaintiffs suffered harm of 10,000).

B. Is asymmetric information always problematic?

It is not because the parties initially have different information that trial is inevitable. The informed party may be able to share the information with the uninformed party. This raises an important question. Why would a litigant who has information that is detrimental to himself ever be willing to share that information with the other litigant? More specifically: why would a defendant with a high chance of being found negligent ever be willing to inform the plaintiff? Why would a plaintiff who suffered only a small amount of harm want to inform the defendant? The answer is simple: because the litigant wants to avoid any negative inference that the uninformed party would draw from the failure to disclose information. The uninformed party could think that the hidden information is even more negative than it actually is. This is known as the "unravelling phenomenon" (Grossman, 1981).[13]

The following example illustrates the unravelling phenomenon. Suppose the plaintiff knows how much harm he suffered. The defendant only knows that the plaintiff suffered harm of 100 with a probability of 1/3, harm of 200 with a probability of 1/3 and harm of 300 with a probability of 1/3. A plaintiff who suffered harm of 300 clearly has an incentive to disclose his private information. If he remains silent, he knows that the defendant will conclude that he suffered less than maximal harm. Given the fact that a plaintiff who suffered harm of 300 will reveal his type, a plaintiff who suffered harm of 200 also has an incentive to do so. Otherwise, the defendant will conclude that he only suffered harm of 100.

The fact that it is often rational to disclose information voluntarily does not imply that information will *always* be shared and that parties will never go to trial (Shavell, 1989). First, some information may be difficult to share in a credible way.[14] If information cannot be shared, the "unravelling phenomenon" may be hampered in a straightforward[15] but also in more subtle way. If certain relatively good types cannot reveal their information credibly, other relatively worse types who are able to produce the infor-

13 For a survey, see (Gertner, 1998).

14 For example: the degree of risk aversion of one of the litigants. In general, documentary evidence would seem to be the best candidate for voluntary sharing.

15 The uninformed party cannot distinguish between the different types who cannot reveal their information.

mation may find it beneficial to remain silent as well. In that way they can try to be mistaken for good types who cannot reveal information. Second, a party may decide not to reveal information during pre-trial negotiations because it would give the other party the time to find a strategy to counter it at trial. For example, a plaintiff might not reveal he has suffered high damages because the defendant could then have a stronger incentive to devote more resources to casting doubt on the evidence of the plaintiff.

C. Basic assumptions of asymmetric information models

It would be impossible to discuss all asymmetric information models in great detail (although we do so for the most important ones later on). However, we can give an overview of the assumptions that almost all of these models have in common. Most asymmetric information models make the following assumptions:[16]

Assumption 1: The judge finds out the truth at trial: The judge is assumed to convict guilty parties, to acquit innocent parties and to award the correct amount of damages. This assumption is made to keep the models as simple as possible. Only a few models have incorporated a parameter for the degree of judicial error (e.g. Hylton (2002)).

Assumption 2: The uninformed party knows that the other party has private information: Although the uninformed party does not know the exact content of the information in possession of the other party, he does know that the other party has private information.

Assumption 3: The uninformed party makes a probability distribution for the private information of the other party: For example, a defendant who does not know the exact amount of damages the plaintiff has suffered may think that there is a 1/3 probability that the damage equals 100,000, a 1/3 probability that the damage equals 200,000 and a 1/3 probability that the damage equals 300,000.[17]

Assumption 4: The informed party knows the probability distribution function of the uninformed party: Referring to the example above, the plaintiff knows that the defendant thinks there is a 1/3 probability that the damage equals 100,000, a 1/3

16 When we specify the assumptions of some specific asymmetric information models in later sections, we will no longer mention these particular assumptions; they are used in all the models we will discuss.

17 Note that most models use continuous probability distributions.

probability that the damage equals 200,000 and a 1/3 probability that the damage equals 300,000.[18]

Assumption 5: Interiority of equilibria: In some circumstances, the parties will always settle even when there is asymmetric information. For example, this can be the case when the trial costs of the plaintiff are so high that he finds it worthwhile to make such a low settlement demand that all types of defendants accept the demand. Most models disregard this possibility and focus on situations in which there is a possibility of trial.

D. Types of asymmetric information models

Over the years a vast amount of asymmetric information models have been developed. In some models, the plaintiff has private information, in other models this applies to the defendant or even both parties. In some models, the informed party can make a proposal; in others the uninformed party can do so. Some models allow for more than one proposal. Sometimes bargaining is assumed to be costly and sometimes costless. In some models, the plaintiff will always go to trial when bargaining fails; in other models he might drop his claim etc. Obviously, a huge amount of combinations are possible. Many of them have been examined in the literature. It is not possible to discuss every single one in detail. We will briefly discuss the importance of some distinctions and then continue to examine the basic canons in greater detail.

D.1. Who has private information?

We can make a distinction between models in which one party has private information and models in which both parties have private information.

One-sided models: Here only one party has private information about a relevant parameter during bargaining.

Example 1: The plaintiff knows exactly how much harm he has suffered, while the defendant only knows that there is an 80 percent probability that the harm equals 1,000 and a 20 percent probability that the harm equals 5,000. The defendant has no private information (for example about the likelihood of being found negligent).

18 More specifically, the probability distribution is common knowledge.

Example 2: The defendant knows whether or not he will be found liable. The plaintiff only knows that there is an 80 percent probability that the defendant will be found liable and a 20 percent probability that he will not. The plaintiff has no private information (for example about the amount of damages).

Two-sided models: Here both parties have private information about the same parameter or a different parameter during bargaining.

Example 1 (same parameter): Only the plaintiff knows what witness number one saw and only the defendant knows what witness number two saw before a traffic accident occurred (e.g. whether the defendant was driving too fast). Here both parties have private information relevant for the liability issue.

Example 2 (different parameter): The plaintiff knows his level of harm and the defendant knows whether he has a high or a low probability of being found liable.

The importance of the distinction. When one party has private information, it can be either the informed party or the uninformed party who makes the proposal. If the uninformed party makes the proposal, the proposal itself does not convey any information. The uninformed party *screens* the informed party (see below). If the informed party makes the proposal, the proposal itself may reveal some information. The informed party may *signal* his type (see below).

When both parties have private information, the proposal will be made by a party who has private information and who knows that the adversary also has some private information. Such a situation is obviously much more complex. The question arises of whether either signalling or screening dominates in such a model or whether both phenomena are likely to be observed at the same time. Daughety and Reinganum (1994) find that the latter is the case. The informed party signals his type and at the same time uses his signal to screen the other party (Daughety and Reinganum, 1994).[19]

19 Suppose for example that there are two types of plaintiffs: those who have suffered harm of 100 and those who have suffered harm of 200. There are also two types of defendants: those who have a 30 percent probability of being found liable and those who have a 70 percent probability of being found liable. Then in equilibrium, plaintiffs who have suffered harm of 100 will ask for an amount that is lower than the amount asked by plaintiffs who suffered harm of 200. And both demands will screen the defendant: the offers will be oriented to the defendants with a higher probability of being found liable.

Note on the literature. Most asymmetric information analyses concern one-sided models. They are usually built on the foundational work of Bebchuk (1984) and Reinganum and Wilde (1986), who considered models in which respectively the defendant and the plaintiff are privately informed. However, some articles analyze two-sided models (Sobel, 1989; Schweizer, 1989; Daughety and Reinganum, 1994).

D.2. What is the content of the private information?
Parties can have private information about any parameter that may be relevant for the pay-offs during bargaining or trial: not only the level of damages, the probability of winning or court costs, but also the degree of risk-averseness, the costs of time, the amount of effort a lawyer spends on a case etc.

The importance of the distinction. We have seen that privately informed parties often have an incentive to reveal their private information voluntarily (the "unravelling phenomenon"). The content of the private information is important because some types of private information, especially documentary evidence, can often be revealed voluntarily in a credible way. For other types of private information, like the degree of risk-averseness, this might be more difficult.

Note on the literature. When the private information concerns liability, it is usually assumed that the defendant is the one with private information. When the private information concerns the level of damages, the plaintiff is usually assumed to be the one with private information. However, this need not always be the case. Consider an example concerning misappropriation of intellectual property. The defendant might be privately informed about damages because he is more familiar with the profits he made.

D.3. Is there one or more than one proposal?
Some models assume that there is only one proposal (by either the plaintiff or the defendant); other models assume that there is more than one proposal (either all made by one party (Spier, 1992) or by both parties in some ordered fashion).[20]

The importance of the distinction. When there is just one proposal, the party that can make the proposal makes a so-called take-it-or-leave-it offer. If the other party does not accept the proposal, a trial will follow. Needless to say, this gives substantial bargaining power to the proposer. The distinction is also important because models with

20 For example: first the plaintiff may make a proposal. If the defendant does not accept it, he can make a counterproposal. If the plaintiff does not accept the counterproposal, the plaintiff may make another proposal and so on (Wang, Kim and Yi, 1994).

just one proposal cannot shed light on the dynamics of the settlement process. If we want to make statements about the timing of settlement agreements, we need to introduce several bargaining rounds.

Note on the literature. Spier (1992) was the first to devise a formal model that generates valuable insights in the dynamics of settlement bargaining. She developed a model in which the defendant has private information and the plaintiff can make several proposals. If the first proposal is rejected, the parties move to the next bargaining round in which the plaintiff can make a new proposal. Spier finds that when bargaining costs are low, most settlement agreements are made just before the start of the trial.

D.4. Is the bargaining horizon infinite?
In models with an infinite bargaining horizon, the bargaining process continues as long as the parties do not reach an agreement, unless the plaintiff decides to stop the negotiations and goes to court. There is no known date in advance at which the trial, with its associated costs, will commence. A finite horizon means that the parties know in advance when the trial will start.

The importance of the distinction. When the bargaining horizon is finite, the parties know that a trial will not commence until a certain date. Until that date, a party can refuse the offers made by the other party without the immediate threat of a trial (and its associated costs). When the bargaining horizon is infinite, refusing an offer carries with it the immediate risk of a trial. This may soften the bargaining strategy of a party.

Note on the literature. Some models investigate bargaining with an infinite horizon, some bargaining with a finite horizon and some examine both situations (e.g. Spier (1992)).

D.5. Is bargaining costly?
Sometimes bargaining is (virtually) costless for a party (e.g. Spier (1992)), and sometimes significant costs are incurred.

The importance of the distinction. If bargaining is costless, the parties do not have an incentive to settle until the actual start of the trial. Making offers and counteroffers is completely free of costs then. Such models are not suited to shedding light on the

duration of settlement negotiations. Some kind of friction is necessary to induce the parties to settle before the actual start of the trial.

Note on the literature. Many models make the assumption that the bargaining process is costly for just one party, for instance the defendant (e.g. Spier (1992)). This assumption is made only for the sake of simplicity. What really matters is the difference between the bargaining costs of the parties.

D.6. Who makes a proposal?

In models in which there is only one proposal either the plaintiff[21] or the defendant[22] makes it. In models with more than one proposal many combinations are theoretically possible: all proposals can be made by either the plaintiff or by the defendant, the proposers may alternate (first the plaintiff, then the defendant, then the plaintiff again and so on) etc.

The importance of this distinction. Whether or not a party is able to make a proposal is important because the ability to make proposals affects the bargaining power of each party (Kennan and Robert, 1993). When only one party can make proposals, we could say that this party has some kind of monopoly power. When a party may make a counterproposal, this may also give him some bargaining power. Moreover, it is important to know whether the informed or the uninformed party is able to make a proposal. If the informed party makes a proposal, the proposal itself may transfer information. This is not the case when only the uninformed party can make a proposal.

Note on the literature. In most models, either the plaintiff or the defendant is allowed to make one or more proposals.[23] More exceptionally, both parties can make proposals (Wang, Kim and Yi, 1994).[24]

D.7. Will the plaintiff go to trial if bargaining fails?

There are two possibilities regarding the plaintiff's claim: either the uninformed party knows whether the claim has positive expected value or he does not know this.

21 This is sometimes called a "P-proposer model".

22 This is sometimes called a "D-proposer model".

23 In Spier (1992), only the plaintiff can make several proposals one after the other.

24 In one article the bargaining structure is considered to be endogenous. The authors give both parties the possibility to (simultaneously) make a proposal and investigate whether they have an incentive to do so (Daughety and Reinganum, 1993).

Example 1: Suppose the defendant knows the probability that the court will hold him liable but the plaintiff does not. The plaintiff only knows that there are two types of defendants and the probability distribution of these types: half of defendants have a 30 percent probability of being found liable and half of defendants have a 60 percent probability of being found liable. Suppose further that the damages equal 10,000 and that the plaintiff's trial costs are 2,000. It is easy to see that the plaintiff will go to trial whenever bargaining fails. His claim always has positive expected value. Even when the defendant has a rather low (30 percent) probability of being found liable, it is still worthwhile to go to trial: $0.3 \times 10{,}000 - 2{,}000 = 1{,}000 > 0$.

Example 2: We return to the previous example and change it so that half of the defendants have a 10 percent probability of being found liable and half of the defendants have a 60 percent probability of being found liable. Now the plaintiff does not know whether his suit has positive or negative expected value. His suit has negative expected value when the defendant is the type with a low probability of being found liable: $0.1 \times 10{,}000 - 2{,}000 = -1{,}000 < 0$.

The importance of this distinction. When the plaintiff does not know whether his suit has positive or negative expected value, he may need to update his beliefs if the defendant rejects his settlement proposal. The fact that the defendant did not accept the proposal can be seen as "bad news": maybe his claim is not worth pursuing. Of course, informed defendants may have an incentive to refuse proposals if they think the plaintiff will consequently drop the suit. More generally, the analysis becomes more complex when the suit may have negative expected value (see below).

Note on the literature. The early models dealing with asymmetric information all assumed that both parties know that the suit has positive expected value. Nalebuff (1987) was the first to relax this assumption for the situation in which the uninformed party makes the proposal (Nalebuff, 1987). Farmer and Pecorino (2002) later did the same for the situation in which the informed party makes the proposal.[25]

25 We discuss both models in later sections.

1.2.5. The basic screening model[26]

A. Assumptions of the model

1. The defendant has private information about the probability that the plaintiff will prevail at trial.[27] The plaintiff only knows the distribution of defendant types. If the defendant is found liable, the judge awards a positive amount in favour of the plaintiff. This amount is common knowledge.
2. Trial is costly for the plaintiff and for the defendant.[28] There are no settlement or filing costs.[29]
3. If the parties do not settle, the plaintiff goes to trial. Even if the defendant is of the "lowest type",[30] going to trial is still worthwhile for the plaintiff because the expected judgment will still be larger than the costs of trial (no NEV suit).
4. he plaintiff demands a settlement amount from the defendant on a take-it-or-leave-it basis.[31] [32] If the defendant accepts the demand, the dispute ends. If he does not, the parties go to court (see assumption 3).

B. The reasoning process of the parties

We start with the defendant because he is the last one to make a decision.[33] For any given settlement demand the plaintiff makes, the defendant knows that there will be a trial if he rejects it (see assumption 3). Thus, he will only reject if his expected loss at trial is smaller than the settlement demand. Otherwise he will accept the demand. The plaintiff takes this decision process of the defendant into account when he makes his demand. For any settlement demand the plaintiff makes, he knows that all defendants with an expected loss at trial that is greater than the settlement amount will accept the demand and all defendants with an expected loss smaller than the settlement amount will reject the demand. Since the plaintiff knows the distribution of defendant types, he knows exactly how likely it is that a certain demand will be ac-

26 The model described below was developed in Bebchuk (1984).

27 The model can easily be extended to a situation in which the defendant has private information not about his liability but about the size of the judgment (Bebchuk, 1984).

28 It is not necessary to view the litigation costs as fixed. These costs can be viewed as the expenditure levels the parties will find optimal in the event that a trial takes place.

29 This is merely a simplifying assumption.

30 Here "lowest type" means that he has a low probability of being found liable.

31 The model can easily be adjusted to accommodate cases in which the defendant makes the offer and the plaintiff is the one with private information.

32 This is merely a simplifying assumption.

33 Remember that the plaintiff will always go to trial if the parties do not settle. The plaintiff is thus not the last one to make a decision is this model.

cepted or rejected. Based on this information, he will determine his optimal settlement demand. Increasing the demand has the advantage that he receives a larger settlement amount if the demand is accepted but has the disadvantage that it reduces the probability that the demand will be accepted (and thus that the probability of a trial is increased). The plaintiff will increase his demand until the marginal cost of doing so equals the marginal benefit. The marginal cost is the additional probability that the defendant will reject the demand multiplied by the total costs of trial.[34] The marginal benefit is the probability that the case will not go to trial multiplied by the increase in the settlement demand. It follows that the settlement offer by the plaintiff will be accepted by defendants who have a relatively high probability of being found liable and rejected by defendants who have a relatively low probability of being found liable. In that sense, the demand of the plaintiff *screens* the defendant.[35]

C. Determinants of the probability of settlement

The amount at stake: An increase in the amount at stake will decrease the likelihood of settlement. The plaintiff will orientate his settlement demand more to defendants who have a high probability of being found liable because the increase in the amount at stake affects these defendants most. An example can illustrate this. Suppose there are three types of defendants: those with a probability of being found liable of 20 percent, 50 percent and 80 percent. If the amount at stake is 100, the expected judgments equal 20, 50 and 80. If the amount at stake doubles to 200, the expected judgments equal 40, 100 and 160. While the original difference between different types of plaintiffs was 30, the difference doubled to 60 when the amount at stake doubled. Focusing the settlement offer towards defendants with a relatively low chance of being found liable becomes more costly. Of course, this implies a lower settlement rate because more (low probability) defendants will refuse the higher demand.

The litigation costs of the parties: An increase in the litigation costs of the plaintiff or the defendant will increase the likelihood of settlement. Intuitively, an increase in the litigation costs of either party makes going to trial costlier for the plaintiff relative to

34 If the parties settle, the plaintiff's demand fully incorporates the defendant's trial costs (since the plaintiff can make a take-it-or-leave-it demand). In that sense, the plaintiff's costs of trial are his own trial costs *and* those of the defendant.

35 Suppose for example that there are five types of defendants. Those with a probability of 20, 30, 40, 50 and 60 percent of being found liable. The trial costs of the defendant are 10 and the amount at stake is 100. Suppose further that the plaintiff makes a demand of 50. This demand will be accepted by all defendant types with a probability of being found liable equal to or greater than 40 percent (since 0.4x100+10 = 50) and will be rejected by the defendant types with a probability of being found liable of less than 40 percent.

settling. If the plaintiff's own costs increase, this is very obvious. And also when the defendant's costs increase, the plaintiff loses more by going to trial. This is because if the parties settle, the plaintiff will get all of the gains from settlement,[36] including the litigation costs saved by the defendant. When the parties go to trial, the plaintiff does not get any part of the defendant's costs.

The distribution of types: An upward shift in the range of types has no effect on the likelihood of settlement. For example, it does not matter whether all defendants have a probability of being found liable ranging between 30 and 70 percent or between 40 and 80 percent, as long as the shape of the distribution remains the same. Intuitively, although such an upward shift increases the absolute levels of the expected judgment for all types of defendants, it does not change the difference in these levels between any two given types.

An expansion of the range of types will decrease the likelihood of settlement. For example, there will be less settlement when the probability that the defendant will be found liable changes from an interval of 40 to 60 percent to the interval of 30 to 70 percent (holding the mean constant). Intuitively, the likelihood of settlement decreases because the differences between the different types of defendants increase. The uninformed party needs to be more wary of misrepresentation by the informed party.

1.2.6. The screening model with NEV suits[37]

A. Assumptions of the model

The first two assumptions are exactly the same as in the basic screening model. The last two assumptions are different, with the purpose of incorporating the possibility of NEV suits.

1. The defendant has private information about the probability that the plaintiff will prevail at trial. If the defendant is found liable, the judge awards a positive amount in favour of the plaintiff. This amount is common knowledge.
2. Trial is costly for the plaintiff and for the defendant. There are no settlement or filing costs.
3. For some defendants, the probability of being found negligent is so low that going to trial with them has a negative expected value for the plaintiff. In other

36 Since he can make a take-it-or-leave-it demand and thus has all the bargaining power (see assumption 4).

37 The model described below was developed by Nalebuff (1987).

words, if the plaintiff were able to distinguish between all types of defendants, he would never go to trial with some of them.[38]

4. The plaintiff first makes a settlement demand. If the defendant accepts, the dispute ends. If he rejects, the plaintiff decides whether to proceed to trial or to drop the claim.

B. The reasoning process of the parties

The fact that the plaintiff could have a NEV suit has a fundamental influence on the reasoning process of the parties. The plaintiff knows that for any settlement demand he makes, he will have to update his information in case the defendant rejects his demand. More precisely, the plaintiff will have to decide whether it is still worthwhile to proceed to court after a rejection. Making a relatively small settlement demand may not be rational from the plaintiff's point of view. If the defendant rejects such a demand, this can be seen as a signal that the defendant has a strong case. In light of this "bad news", the plaintiff should then revise downwards what he can expect to win in court and possibly decide that it is no longer worthwhile pursuing the case. The defendant anticipates this and may have an incentive to undermine the credibility of the plaintiff's threat by refusing relatively low demands. If the plaintiff "inflates" his settlement demand and the demand is rejected, this does not necessarily imply that the plaintiff has a weak case: rejecting a large demand does not convey a lot of information.[39] Consequently, the plaintiff will still have a credible threat to go to trial after the rejection of such a demand. In more concrete terms, the optimal settlement demand for the plaintiff is the minimum demand at which the threat to litigate with certainty is credible.

C. Determinants of the probability of settlement

The amount at stake: The likelihood of settlement increases when the amount at stake increases. The reason is that the increase in the amount at stake allows the plaintiff to learn more "bad news". Rejection by the defendant of a demand is less detrimental for the plaintiff than before, since the expected value of a trial increases. Conse-

38 However, Nalebuff does assume that the case has positive expected value given the prior distribution of defendants. In other words, the mere fact that there is a possibility that the case might have negative expected value does not prevent the plaintiff from filing suit because the probability that the case has positive expected value is sufficiently large.

39 For example: the case in which the plaintiff demands $100,000 and this demand is rejected is very different from the case in which the plaintiff demands $100 and the demand is rejected. A rejection in the latter case obviously carries a larger signal that the plaintiff's case is weak. If the plaintiff asks for $100,000 and the defendant refuses, the case may still be worth $90,000 or $80,000 etc. This is not the case when a demand of $100 is refused.

quently the plaintiff will be under less pressure to inflate his settlement demand. This leads to more settlement.

The litigation costs of the parties: An increase in the litigation costs of the plaintiff leads to a lower settlement frequency. The reason is that the plaintiff has to make a higher settlement demand to maintain a credible threat if the demand is rejected. Fewer defendants will accept this higher demand. An increase in the litigation costs of the defendant has no influence on the likelihood of settlement. The litigation costs of the defendant do not affect the credibility of the plaintiff's threat to go to trial.

1.2.7. The basic signalling model[40]

A. Assumptions of the model

1. The plaintiff has private information concerning the damages: he knows how much the court will award in the case of a verdict in his favour. The defendant only knows the distribution of the plaintiff's type (and thus of the damages). The parties both know the probability that the defendant will be found liable.
2. Trial is costly for the plaintiff and for the defendant. There are no settlement or filing costs.
3. The plaintiff always proceeds to trial when the parties do not settle: going to trial is always worthwhile, even for those plaintiffs who have suffered the minimum amount of damages. In other words, all plaintiffs have a positive expected value suit (and they know it).
4. The plaintiff demands a settlement amount from the defendant on a take-it-or-leave-it basis. If the defendant accepts the demand, the dispute ends. If he does not, the parties go to court (see assumption 3).

B. The reasoning process of the parties

We start with the defendant, since he is the last one to act.[41] For any settlement demand the plaintiff makes, the defendant has to decide whether to accept or reject it. If he were always to reject, he would always be taken to court (see assumption 3). If he were always to accept, all plaintiff types would ask for the maximum settlement amount. Clearly, always rejecting or always accepting is not an optimal strategy for the defendant. Rather, he will reject any given settlement demand with a certain probability. The problem a defendant faces is to find the optimal probability of rejection for each possible settlement demand. The problem the plaintiff faces is to demand an

40 The model described below was developed by Reinganum and Wilde (1986).

41 Note that the plaintiff will always go to trial if the parties do not settle. In that way, not the plaintiff but the defendant is the last one to act.

optimal settlement amount given his actual damage and the rejection function of the defendant. Reinganum and Wilde (1986) show that there exists an equilibrium in which all plaintiffs reveal their types and defendants reject higher demands more frequently than lower demands.[42] The more the plaintiff asks for, the greater the probability of a trial. In more concrete terms, every demand is rejected with a probability that is exactly sufficient to make all plaintiffs reveal their true type. This is a so-called separating equilibrium, because the plaintiff's settlement demand signals his true damages.[43]

C. Determinants of the probability of settlement

The probability of a plaintiff victory: An increase in the probability of a plaintiff victory reduces the likelihood of settlement. The reason is that plaintiffs have an increased incentive to bluff by submitting a higher demand than the one associated with their type. If such a demand is accepted, they gain more than before.[44] The defendant has to increase the probability of rejecting any given demand to counter the increased incentive to bluff.

The litigation costs of the parties: An increase in the litigation costs of the plaintiff increases the likelihood of settlement. Intuitively, since going to trial becomes more costly for the plaintiff, his incentive to make a demand that is higher than the one associated with his type gets smaller.[45] Likewise, an increase in the litigation costs of

42 The authors also show that there exist other equilibria (pooling equilibria). They exclude these equilibria by using an equilibrium refinement developed by Banks and Sobel (1987), the so-called "universally divine equilibrium".

43 Clearly this is a Nash equilibrium because neither party can do better by choosing another strategy, given the strategy of the other party. If the plaintiff makes a larger demand than the one associated with his type, the probability that the defendant will take him to trial increases in such a way that the plaintiff will be worse off. The plaintiff will consequently not "defect". The defendant will also not "defect" given that all plaintiffs reveal their type. If he were to lower his rejection rate, some plaintiffs would find it worthwhile to pretend to be of a higher type. If he were to increase his rejection rate, he would go to trial more often. This would not leave the defendant better off.

44 Put otherwise, the adverse selection problem becomes more severe because the difference between the expected outcome of any two given plaintiff types increases. Suppose there are two types of plaintiffs: those who have suffered harm of 100 and those who have suffered harm of 200. If the probability of a plaintiff victory is 0.5, the difference between the types equals 0.5x200 – 0.5x100 = 50. If the probability of a plaintiff victory increases to 0.8, the difference between the types equals 0.8x200 – 0.8x100= 80. Consequently, the plaintiff who suffered harm of "only" 100 has an increased incentive to pretend he suffered harm of 200.

45 A higher demand by the plaintiff is answered by a higher probability of trial by the defendant.

the defendant leads to an increase in the likelihood of settlement. Settling becomes more worthwhile for the plaintiff because the settlement amounts increase.

The distribution of types: An upward shift in the range of types has no effect on the likelihood of settlement. Although such an upward shift increases the absolute levels of the expected judgment for all types of plaintiffs, it does not change the difference in these levels between any two given types. Consequently, plaintiffs do not have an increased incentive to pretend they are of a higher type.

An expansion of the range of types will decrease the likelihood of settlement. The reason is that the differences between the different types of plaintiffs increase.

1.2.8. The signalling model with NEV suits.[46]

A. Assumptions of the model

Assumption 1 is the same as in the basic signalling model. Assumptions 2, 3 and 4 are (partly) different:

1. The plaintiff has private information concerning the damages: he knows how much the court will award in the case of a verdict in his favour. The defendant only knows the distribution of the plaintiff's type (and thus of the damages). The parties both know the probability that the defendant will be found liable.[47]
2. Trial is costly for the plaintiff and for the defendant. There are no settlement costs. The plaintiff bears a positive cost of filing.
3. Some plaintiffs will not go to trial when the parties do not settle. For these plaintiffs, the damages are so low that their expected value of trial is negative.
4. The plaintiff demands a settlement amount from the defendant on a take-it-or-leave-it basis. If the defendant accepts the offer, the dispute ends. If he does not, some plaintiffs will go to trial and some will drop their case (see assumption 3).

B. The reasoning process of the parties

Farmer and Pecorino (2002) first examine what would happen if there were no filing costs (as in the basic signalling model). Would every plaintiff still make a demand that matches his type? The answer is no. A plaintiff with an NEV suit will never reveal his type through his demand. If he were to do so, the defendant would refuse his demand because he knows that the plaintiff will drop the case after a refusal. Plaintiffs with NEV suits have an incentive to make a demand associated with a plaintiff of a

46 The model described below was developed by Farmer and Pecorino (2002).

47 More precisely, it is common knowledge.

better type,[48] since there is no cost attached to making such a demand (the costs of filing are zero). If the defendant refuses the demand, the plaintiff simply drops his claim and will not incur any costs. As a consequence, the signalling equilibrium will come unravelled. The defendant will prefer to go to court with all plaintiffs rather than settle with some plaintiffs with NEV suits.[49]

Farmer and Pecorino (2002) show that the signalling equilibrium can be restored if the plaintiff has to incur a positive filing cost.[50] In that case, the defendant can increase his general rejection rate in such a manner that only plaintiffs who have a credible threat to proceed to trial will find it worthwhile to sue and make an offer.[51]

C. Determinants of the probability of settlement
Since a signalling equilibrium exists when there are positive filing costs, we can refer to 1.2.7. for the determinants.

1.2.9. Criticism of the asymmetric information model

A. Telling the truth
As we described above, asymmetric information causes trials because "bad types" may have an incentive to present themselves as "good types" during settlement negotiations. In other words, trials occur under the assumption that people will lie about their type if that would give them a larger pay-off.

How plausible is it that people lie during settlement negotiations? Although we are not aware of any empirical research that sheds light on this topic, there is some evidence that parties sometimes lie during trial (Sanchirico, 2004). And if parties some-

48 Thus a plaintiff with larger damages.

49 The reason is that the defendant is indifferent between a trial and a settlement in relation to plaintiffs with a positive expected value suit, since these plaintiffs make an offer equal to the defendant's expected loss of trial.

50 Or any other significant expenditure prior to the actual trial.

51 Remember that the defendant is indifferent between a trial and a settlement in relation to plaintiffs with a positive expected value suit, since these plaintiffs make an offer equal to the defendant's expected loss of trial.

times lie during trial, it seems very unnatural to assume that they would never lie during the settlement stage.[52]

B. Different bargaining structures, different results?

1. It is clear from the analyses above that there is no such thing as *the* asymmetric information model. Instead, there are many variants. In some models the plaintiff can make a proposal; in others the defendant. In some models, the informed party makes a proposal; in other models the uninformed party makes a proposal. In some models, there is only one proposal, in other models there can be many proposals and so on. As we have seen above, these variants may lead to different predictions about the likelihood of settlement.[53] However, the various models have more in common than one would think at first sight:
2. In all models, the presence of asymmetric information generally leads to some use of the courts by the parties.
3. For many models, several determinants influence the settlement rate in the same way. For example, the settlement rate increases in screening *and* signalling models (without NEV suits) when the amount at stake decreases, the litigation costs of the parties increase or the distribution of types gets narrower.
4. The insights from models with one-sided information are helpful in understanding the two-sided model. This model has a signalling and a screening element (Daughety and Reinganum, 1994).
5. Dynamic models of the settlement stage are not as different from the static ultimatum models as one may think: a lot of bargaining takes place in the last period in dynamic models (see also section B, The Duration of the Settlement Negotiations) (Spier, 1992).

In conclusion, many insights derived from the simplest models of settlement bargaining also have explanatory power in more complex bargaining environments. These simple models which provide very precise characterizations of the equilibrium strategies and outcomes can be seen as prototypes that reveal the main effects to be expected in more complex situations (Kennan and Wilson, 1993).

52 Kennan and Wilson (1993, p. 49) note: "…a long tradition of social and ethical commentary argues for the morality of forthright honesty, not only for one's personal integrity but also for the social welfare" and further "Nevertheless, in the tradition of economics as the 'dismal science', the prediction adopted as an assumption is that each party tries relentlessly to maximize its own private gain from trade".

53 Consider the differences between screening and signalling models and differences between models with and without negative expected value suits etc.

1.2.10. A comparison of the results of the relative optimism model and asymmetric information models

A. Similarities between the models

A.1 Similar influence of the basic determinants on the settlement frequency
In the relative optimism model, the basic screening model, the basic signalling model and the signalling model with NEV suits, an increase in either the trial costs of the plaintiff or the trial costs of the defendant leads to a higher settlement frequency. Likewise, an increase in the amount at stake decreases the settlement frequency in all these models. Next, in the relative optimism model, the settlement frequency drops when the parties' estimates about the likely outcome of trial tend to diverge more. In the abovementioned asymmetric information models, the settlement frequency drops when the range of types expands.

A.2 Tried cases are a selection of filed cases
The models have in common that the cases that are *tried* are not representative of the population of *filed* cases. In other words, filed cases pass through a filter in the two models, leading to a sample of tried cases that is atypical for the group of filed cases.[54] It is hard to take to overestimate the importance of this result: it implies that it is not possible to directly infer characteristics of filed cases from the more visible pool of tried cases. For example, it is not because plaintiffs win X percent of tried cases that they would win X percent of filed cases if they were all tried. Likewise, it is impossible to say whether legal standards favour plaintiffs, defendants or neither party by looking at the plaintiff win rate: low (< 50 percent) or high (> 50 percent) win rates indicate that especially weak or strong cases go to trial. We will describe how the filters of the two models work in the next section.

B. Differences between the models

B.1 Different influence of the basic determinants on the settlement frequency
In the relative optimism model, an increase in either the trial costs of the plaintiff or the trial costs of the defendant always leads to a higher settlement frequency. An increase in the amount at stake always leads to a smaller settlement frequency. This is not the case in all asymmetric information models. In the screening model with NEV suits, an increase in the litigation costs of the plaintiff lowers the settlement frequency and an increase in the amount at stake increases the settlement frequency.

54 Note that any model of settlement behaviour, not just the relative optimism or the asymmetric information model, is also a model of the selection of tried cases (Wittman, 1985; 1988).

B.2 Different selection of filed cases

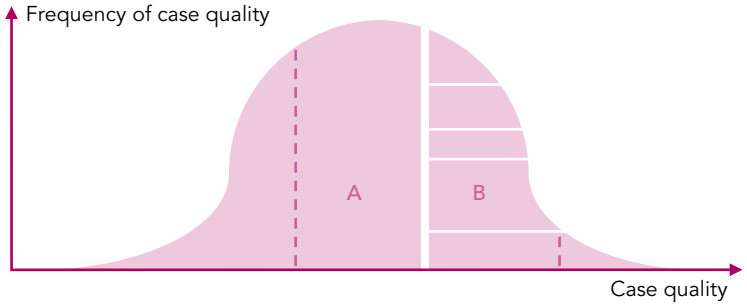
The cases that proceed to trial in the relative optimism model are quite different from the cases that proceed to trial in the asymmetric information model. In the relative optimism model, the disputes most likely to be litigated are those whose outcome is most problematic (given the decision standard). Under some specific assumptions (see below), plaintiff win rates show a tendency towards 50 percent as the number of trials decreases. In asymmetric information models,[55] informed parties only proceed to trial when they expect to win. Consequently, plaintiff win rates show a tendency towards either 0 (informed defendants) or 100 percent (informed plaintiffs) as the number of trials decreases (Waldfogel, 1998). We will first discuss the details of the selection process in the relative optimism model and then in the asymmetric information model.

Priest and Klein (1984) were the first to present a formal model of case selection based on the relative optimism model. Their model is based on the following assumptions:

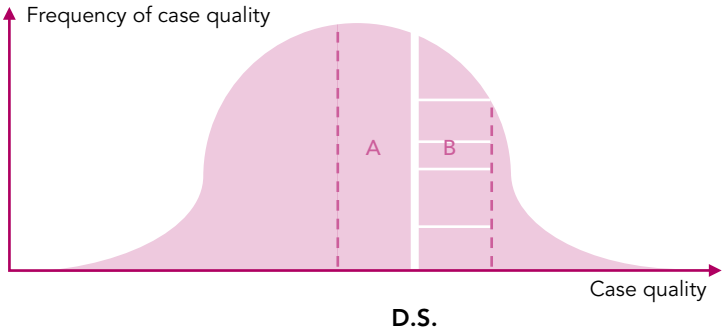
1. It is possible to array the pool of all disputes according to the relationship of each dispute to the decision standard (D.S.) that governs the disputes. Both parties know the exact location of the decision standard. This standard divides the distribution of all disputes into two sections: the section to the right of the decision standard represents the disputes the plaintiff would win if litigated and the section to the left of the decision standard represents the disputes the defendant would lose if litigated (see the figure below).
2. The stakes are symmetric.
3. The parties may have different opinions about the plaintiff's probability of winning. More specifically, the parties form an estimate of case quality that may contain an error. Both parties predict case quality with equal precision (with "equal error variance"). The error however is not biased (parties do not systematically over- or underestimate the true case quality) but random and not correlated across parties (e.g. when the plaintiff overestimates the true case quality, the defendant does not automatically overestimate the true case quality as well).
4. The plaintiff and the defendant form their settlement demand or offer non-strategically: the parties settle whenever the plaintiff's expected net benefit of trial is smaller than the defendant's expected loss of trial.

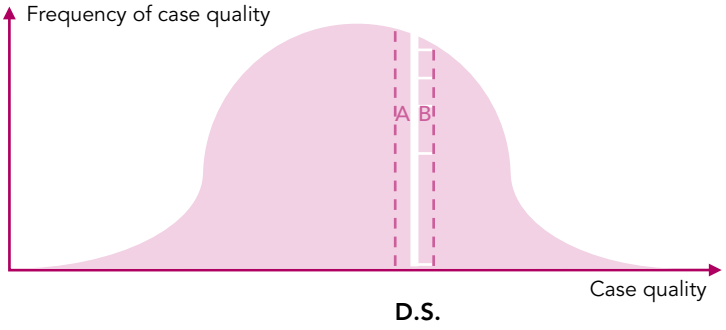
55 At least one-sided asymmetric information models.

By looking at the figures below, we can see that the plaintiff win rate moves to 50 percent when the number of trials becomes smaller. In the first figure, the parties estimate case quality with a rather large error. Litigated cases are confined to the interval between the dotted lines; all disputes outside of the striped lines are settled. The surface A to the left of the decision standard indicates the disputes decided for the defendant; the surface B to the right of the decision standard indicates the disputes decided for the plaintiff. Clearly, surface A is larger than surface B, so the plaintiff win rate will be lower than 50 percent. In the middle figure, the parties estimate case quality with a smaller error than before. We can see that the areas A and B converge. The plaintiff win rate moves closer to 50 percent. In the figure below, the parties estimate case quality with a very small error. There will only be very few trials. The areas A and B are almost the same now, which means that plaintiff and defendant win rates should be almost the same: 50 percent.[56]



56 When we drop the assumption of equal stakes, we reach different results. When the stakes are greater for plaintiffs than for defendants, smaller differences between the parties' expectations of success will be sufficient to cause a trial, since plaintiffs in general will increase their settlement demands (relative to defendants' offers). The rate of litigation thus increases. Moreover, plaintiffs will increase their settlement demands in particular in disputes in which they have a relatively greater chance of winning. As a consequence, the plaintiff win rate at trial will exceed 50 percent. Alternatively, when the stakes are greater for defendants than for plaintiffs, greater differences between the parties' expectations are necessary to cause a trial, since defendants in general will increase their settlement offers (relative to plaintiffs' demand). The rate of litigation thus decreases. Moreover, defendants will increase their settlement offers in particular in disputes in which the plaintiffs have a relatively greater chance of winning. As a consequence, the plaintiff win rate at trial will be lower than 50 percent.





In asymmetric information models, the selection of tried cases from filed cases is very different. Take for example the basic screening model (see above). In this model, the uninformed plaintiff makes an offer that is accepted by defendants with a relatively high probability of being found liable and is rejected by defendants with a relatively low probability of being found liable. Consequently, when the number of trials becomes very small, in the limit the plaintiff only takes those defendants to trial which have a zero probability of being found liable. The plaintiff win rate goes to zero as the number of trials decreases (on a similar note see Waldfogel (1998)). The situation is quite different in the basic signalling model (see above). In that model, the informed plaintiff makes an offer to the uninformed defendant, who rejects higher offers more frequently than lower offers to prevent misrepresentation. Consequently, plaintiffs who have a higher probability of winning go to court more often. In the limit, as the number of trials goes to zero, the plaintiff win rate converges to one.

2. Consequences for earlier stages

2.1. The influence of settlement on the incentive to file

Holding everything else constant, an increased settlement rate leads to an increase in the number of filings. The intuition is simply that plaintiffs do not settle for less than their expected value of trial. Of course, in reality, not everything else is held constant when the settlement rate increases. When an increase in the settlement rate goes hand in hand with other changes, results can be ambiguous (see also section 1.2. of the filing stage chapter).

2.2. The influence on *ex ante* behaviour

An increased settlement rate can be advantageous for the defendant. Depending on his relative bargaining power, the amount he has to pay in settlement can be considerably lower than his expected loss at trial. This will lead the defendant to take less care *ex ante*, to not fulfil his contractual obligations more often etc. Ultimately, the effect of an increased settlement rate on the defendant's *ex ante* incentives is ambiguous, given that two effects working in different directions are at play (more plaintiffs can file suit; the defendant loses less in the case of a settlement than with a trial). However, the greater the bargaining power of the defendant, the more likely it becomes that an increased rate of settlement will negatively influence his *ex ante* incentives.

2.3. Should settlement be encouraged?

2.3.1. Shavell's theory applied to the decision to settle

Previously, we discussed Shavell's theory regarding the social versus the private incentive to sue. Shavell found that, from a social point of view, there could be either too much or too little suit. Likewise, there may also be a divergence between the social and the private incentive to settle. First, the litigants may have an insufficient incentive to settle because (often) they do not bear all of society's trial costs (e.g. the costs of the public courts are partly subsidized). Second, the incentive to settle may also be inadequate because of asymmetric information. As Shavell puts it, while asymmetric information leads parties to fail to settle because they may misgauge each others' situations, the fact that a party may incorrectly estimate the other's situation does not

mean that social resources should be expended on trial. Third, it is possible that parties may settle too often. Parties will usually not consider deterrence as an important factor in deciding between settlement and trial. More specifically, when settlement reduces deterrence, that means that there is too much settlement.

2.3.2. Social benefits and costs of settlement

From a social point of view, settlement has several important benefits:

1. It reduces the transaction costs of the legal system. One of the major reasons parties settle is to avoid the (often large) costs of trial. While settlement also entails transaction costs, the costs of trial are usually thought to be larger than the costs of settlement.
2. It reduces the costs (of risk averse parties) due to the inevitable uncertainty of the trial outcome.
3. It may bring swifter relief for injured plaintiffs.
4. It frees up space on crowded dockets.

However, an increased settlement rate can have negative implications as well:

1. It may increase the number of filed suits (see above).
2. Some pertinent legal questions will not be answered by the court, or it will take longer before an opportunity arises for these questions to be answered. From a dynamic perspective, the original increase in the settlement rate may be followed by a decrease in the settlement rate in the future, because litigants will have fewer judge-made rules to base their decisions on, and consequently the risk of divergent expectations may increase.
3. Settlement secures privacy. Information that would have emerged at trial may never come to the attention of the public. While in some cases the litigants' desire for privacy may be socially beneficial, at many times society could benefit from the information that a trial would have revealed (e.g. whether a product is defective).
4. If the defendant has sufficiently large bargaining power, his/her *ex ante* incentives may decrease (see above). However, several authors have noted that one can always change the rules by increasing damages so that settlement does not lead to less deterrence than trial (e.g. Spier (1997)). In other words, according to these authors, there are better ways to maintain deterrence than by discouraging settlement.

5. Friedman and Wickelgren (2008) however argue that in the presence of asymmetric information, this last statement is not necessarily true. Their argument is structured as follows:
   a. Ideally, the legal system should discourage actions that cause more harm to others than the benefits the actions provide. Also, the legal system should encourage or at least not discourage activities with benefits larger than the harm they cause (so that socially desirable activities are not *chilled*).
   b. Suppose the defendant has superior information about her liability, and thus the probability that she will win in court. Allowing settlement in this instance tends to decrease the accuracy of the legal system and leads to either less deterrence of harmful activity or more chilling of beneficial activity that could be mistaken for harmful activity. The reason that settlement can reduce the accuracy of the legal system is that culpable defendants have a larger incentive to settle than blameless defendants. Consequently, when the litigants do not settle, the plaintiff will realize that he probably faces a blameless defendant. This will cause the plaintiff to expend less effort on trial preparation (or not go to trial at all), and will reduce the accuracy of the trial process. The less accurate the trial process, the less effective the legal system is at deterring harmful behaviour without chilling beneficial activity. Even when the rules for awarding damages can be changed, when settlement is allowed there may be no way to change the rules to maintain deterrence without creating undesirable chilling.

3. Modelling issues

We have described several models and extensions in the section on the filing stage. We will also discuss several models in the section on the trial stage. To avoid duplication, we would refer to those sections for further discussion.

4. Empirical issues

4.1. Empirical research on the voluntary transmission of information

The unravelling phenomenon has been confirmed in an experimental setting in which costless and credible transmission of information is possible. Pecorino and Van Boening (2004) find that the great majority (80 percent) of (individuals assigned the role

of) plaintiffs with strong cases indeed reveal their information. When the plaintiff remains silent, the defendant interprets this as an indication that the plaintiff has a weak case and makes a low offer.

4.2. Empirical research on who has private information

There are some empirical indications that defendants are more often privately informed than plaintiffs about the plaintiff's chances of success in court. Osborne (1999) reasons that a privately informed party should be better able to predict the outcome of a trial. He finds that defendants are significantly better able to predict the outcome of non-settled cases than plaintiffs.

4.3. Empirical research on the number of offers during settlement negotiations

In the U.K., most claims receive only one offer before settlement or trial. Swanson and Mason (1998) found that only 0.3 offers are made in each year of a claim.[57] Kritzer (1985) found that there were three or more exchanges of offers and counteroffers in only 15 percent of all U.S. cases that he studied. This figure is more or less the same across case complexity and stakes.

4.4. Do parties bargain with a finite or infinite horizon?

In reality, parties often bargain with a finite horizon (also called a "fixed date"). The reason is that court availability is generally irregular. This makes it necessary to schedule a court date in time.

4.5. Who usually makes the settlement offer?

Fenn and Vlachonikolis (1990) find that the defendant is typically the party that makes the offer in British cases.

4.6. Is the outcome of the screening model realistic?

In an experimental setting in which the plaintiff has private information and the defendant makes a take-it-or-leave-it offer, Pecorino and Van Boening (2004) find that defendants engage in a screening strategy under which they settle (very frequently) with plaintiffs who suffered relatively small harm and go to trial with plaintiffs who suffered relatively large harm.

57 They use data from the U.K. Taxing Masters.

4.7. The extent of settlement costs

Although there is only a small amount of data available on the extent of settlement costs, the data that do exist show that these costs can be quite high. Salop and White (1988) examined about 2,000 U.S. antitrust cases filed between 1973 and 1983 (that is about 1/6th of all antitrust suits during that period). They estimated that the litigation costs of cases that were settled amounted to roughly 75 percent of the litigation costs of tried cases.

4.8. The influence of asymmetric trial cost

Most theoretical models predict that the incidence of trial remains the same as long as the sum of the trial costs of the parties does not change. Experimental research by Pecorino and Van Boening (2006) however shows that not only the total trial costs are important, but also how these trial costs are distributed between the parties. They find that the dispute rate increases when the trial costs of the party making the offer are larger than the trial costs of the other party. The reason is that fairness plays a role in settlement negotiations and that it is more difficult for parties to agree on what is a fair offer when trial costs are asymmetric.

4.9. Empirical findings in favour of the optimism model

Priest and Klein (1984) found evidence that the plaintiff win rate at trial was roughly 50 percent in different contexts. They examined the proportion of plaintiff verdicts in contested civil cases tried before juries between 1959 and 1979 in the courts of Cook County, Illinois.[58] They found a proportion of plaintiff victories varying between 44.1 (1972) and 55.3 (1963) percent. The composite proportion for the 21 years was 48.47 percent.[59] The authors also looked at the proportion of plaintiff verdicts for three courts separately: the Municipal Courts, the Circuit Courts and the U.S. District Courts. Once again, the composite figures are close to 50 percent: 45.60, 49.54 and 52.99 percent respectively. Finally, Priest and Klein also looked at jury verdicts by subject for the same time period: the composite figures were 47.4 for traffic cases, 47.8 for injury on property cases, 55.6 for street hazard cases, 66.3 for worker injury cases, 42.8 for product liability cases and 39.6 for malpractice cases.[60]

58 Default judgments, directed verdicts and cases in which liability by the defendant was admitted were excluded.

59 With the exception of 1977: the plaintiff win rate was only 37.7 percent in that year.

60 Note that plaintiff win rates close to 50 percent are not real proof since the 50 percent result is reached only when the number of trials converges to zero.

Above we have seen that the Priest and Klein model predicts a specific relationship between the trial rate and the plaintiff win rate at trial: the plaintiff win rate moves to 50 percent as the trial rate approaches zero. Waldfogel (1995) examines this relationship across case types and across judges within a US federal court district. He finds that plaintiff win rates move towards 48 percent as the trial rate for case types approaches 0. He also finds that for tort cases, plaintiff win rates approach 54 percent as the trial rate for judges approaches zero.

The same author also uses the Priest and Klein model in reverse: given the variation in the trial rate and the plaintiff win rate across groups of cases, he derives estimates of the decision standard, the standard error of the parties and the degree of stake asymmetry between the parties for intellectual property cases, contract cases and tort cases. He finds that if all filed cases were to be tried, plaintiffs and defendants would win an equal amount of intellectual property cases, the majority of contract cases would be won by plaintiffs and the majority of tort cases would be won by defendants. In other words, the decision standard roughly equals zero for intellectual property cases, is way below zero for contract cases and way above zero for tort cases. Second, standard error is greatest in tort cases and lowest in contract cases. Finally, defendants have higher stakes in tort cases and plaintiffs have higher stakes in intellectual property and contract cases.

Siegelman and Waldfogel (1999) compare these estimates derived from the Priest and Klein model with independent evidence on the degree of stake asymmetry and error.[61] They reason that if the independent evidence matches the estimates, then this provides strong evidence for the Priest and Klein model. The authors find that the independent evidence is indeed consistent with the estimates.[62] However, when they make the same comparison for three other case types (labour cases, civil rights cases and prisoner cases), they find that the estimates are far less consistent with the inde-

61 Independent evidence means data that is independent of the litigation outcomes themselves. The authors use the relative fraction of repeat plaintiffs and repeat defendants and the relative fraction of institutional parties among plaintiffs and defendants as a measure of stake asymmetry. They use the fraction of plaintiffs that are not represented by a lawyer, the fraction of published opinions, the fraction of adjudicated cases that are adjudicated early and the fraction of fully tried cases heard by juries as a measure of error.

62 The independent evidence shows that tort defendants and intellectual property plaintiffs are often repeat players and that most plaintiffs and defendants in intellectual property and contract suits are institutions, while not even 5 percent of tort plaintiffs are institutions. Furthermore, the independent evidence shows that plaintiffs who are not represented by lawyers are most frequent in tort cases, just like published opinions and fully tried cases (but not early adjudication).

pendent evidence (especially for the last two case types). The inconsistencies disappear however when the authors relax the assumption that plaintiffs estimate case quality as well as defendants.

Farmer, Pecorino and Stango (2004) find support for the relative optimism model in data from major league baseball disputes about salaries.[63] These disputes are well suited to empirical research since all relevant information is available: the offers by both sides, the *ex post* salary received by the player and objective data about the value of individual players (e.g. the history of games played, home runs, saves, runs, hits etc.). The authors first use an econometric model that predicts player and club offers using the publicly available information about a player/club pair. When an actual player offer is higher than the predicted player offer, this is called an aggressive player offer. Likewise, a club offer below the predicted club offer is called an aggressive club offer. In the optimism model, a party who is making an aggressive offer is making a mistake (in this instance, because the offer exceeds the expected value by the econometric model). A more aggressive offer by a player should thus be associated with a lower *ex post* player salary and a more aggressive offer by a club should be associated with a higher *ex post* player salary. On the contrary, in an asymmetric information model, an aggressive offer reflects the possession of private information. A more aggressive offer by a player means that he has private information beneficial to him, a more aggressive offer by a club means the club has private information that is harmful to the player. If the parties' offers reflect their private information, players who make aggressive offers should obtain higher salaries and clubs that make aggressive offers should pay lower salaries. The results these authors find are consistent with a model in which players display excessive optimism: in player/club pairs that go to arbitration, more aggressive player offers lead to lower player salaries.

Finally, many laboratory experiments have shown that individuals exhibit self-serving biases.[64] Individuals who are given identical facts in a civil litigation case interpret the facts more favourably for the defendant when they are given the role of defendant and interpret the facts more favourably for the plaintiff when they are given the role of plaintiff.

63 The parties may resort to final offer arbitration in such disputes.

64 For a survey of this experimental literature, see Babcock and Loewenstein (1997). Self-serving bias is not only found to be exhibited by inexperienced individuals, but also by lawyers and other experienced negotiators, see Babcock and Loewenstein (1997).

4.10. Empirical findings in favour of the asymmetric information model[65] [66]

Many authors have reported strong deviations from the 50 percent rule. Eisenberg (1990) for example examines plaintiff win rates by case category for the U.S. federal district courts during the period 1978-1985.[67] He finds plaintiff win rates of 69 percent

65 We would stress that there are other possible reasons for settlement failure. First, trials can also be caused by agency problems between a client and his lawyer. Often, the lawyer is much better informed about the relevant laws and the possible strategies than his client. For the client it is often difficult to verify whether his lawyer is acting in his best interests. Lawyers, knowing this, can have an incentive to act in their own personal interests rather than in their clients' interests (unless there are strong reputation sanctions). One consequence of this agency problem is that lawyers who are paid by the hour will go to trial too often (or settle not often enough) seen from the client's perspective (at least if trial work is more attractive than other available work): if the lawyer settles the case, his income is usually less than when the case goes to trial. Another reason for settlement failure is external effects. Even when parties agree on the likely outcome of the case and there are no principal-agent problems, trials may occur because of the existence of external effects. Litigation may result when a current case has consequences for the future pay-offs of the parties. More specifically, trials occur when the external benefits to one party are sufficiently larger than the external losses to the other party. The asymmetries in external effects then simply prevent the existence of a positive settlement range. Example 1: A defendant may refuse to settle because he wants to establish a reputation as a tough negotiator. This may benefit the defendant in future negotiations with other plaintiffs. The future plaintiffs may be willing to accept rather low settlement amounts because the defendant's threat to go to court if they do not accept low offers may be conceived as credible given the fact that the defendant went to court with other plaintiffs in the past. Suppose that the plaintiff and the defendant agree that the plaintiff has a 70 percent probability of winning 10,000 and that the trial costs of each party equal 2,000. Clearly, a settlement is possible: the bargaining range equals 4,000. However, suppose now that going to court would create an external benefit for the defendant of 5,000 because future plaintiffs will accept low settlement offers and that there are no external losses for the plaintiff (for instance because he is a "one-shot player"). This asymmetry in external effects causes the settlement range to shrink from 4,000 to –1,000. A settlement is no longer possible. Example 2: A defendant may refuse to settle because he wants to elicit a precedent from the court stating that his behaviour was compliant with the law. This may benefit the defendant in future negotiations with other plaintiffs if the defendant expects similar claims from other plaintiffs (Galanter, 1974; Rubin, 1977; Che and Yi, 1993; Kobayashi, 1996).

66 Some authors have made efforts to synthesize relative optimism and asymmetric information models into an informational model in which the parties may each exhibit a self-serving bias (Farmer and Pecorino, 2002). The insight that informational asymmetry and relative optimism both play a role can also be found in Genn (1995) and in Goriely (2002).

67 A total of almost 65,000 cases.

for marine contract cases, 77 percent for defaulted student loan cases, 57 percent for contract product liability cases, 12 percent for land condemnation cases, 68 percent for rent, lease and ejectment cases, 64 percent for tort to land cases, 60 percent for motor vehicle tort cases, 38 percent for medical malpractice cases, 63 percent for Fair Labor Standards Act cases and so on.[68] In the Netherlands in 1986, a plaintiff success rate of 87 percent was measured for cases in first instance (except family law cases; default cases were excluded) for the courts of The Hague, Rotterdam, Breda, Arnhem and Alkmaar for cases in which plaintiffs and defendants were both individuals and the plaintiff did not enjoy financial assistance from the state (Klijn, 2007). When plaintiffs who did not receive financial assistance from the state sued a company, the plaintiff win rate was 63 percent. On average, plaintiffs had an 80 percent win rate. Plaintiffs who enjoyed financial assistance from the state had a 76 percent win rate against individuals and a 54 percent win rate against companies; on average, they had a 67 percent win rate.[69] [70]

Osborne (1999) reports evidence in favour of the asymmetric information model that does not rely on plaintiff win rates different from 50 percent. He reasons that if asymmetric information exists and one party is better informed than the other, then one party should be able to predict the outcome of the trial much better than the other party. In the relative optimism model on the contrary, trials are the result of an unusual amount of optimism on the part of either the plaintiff or the defendant. As a result, the accuracy of predictions should be more or less equal among parties. Osborne uses the data set from the Civil Litigation Research Project (CLRP) (Trubek, 1983). The set contains data from US court (and alternative dispute) cases in the late 1970s and covers a wide variety of areas of law. In contrast to many other data sets, the CLRP data set contains a measure of the plaintiff's and the defendant's expectation of the outcome of the trial.[71] The author finds that defendants are significantly better able to predict the outcome of non-settled cases than plaintiffs. However, this is only true for

68 Deviations from 50 percent are also reported in Hylton (1993).

69 Note that a plaintiff was assumed to win his case even if his claim was only partially granted.

70 Remember that the 50 percent rule is only true when the number of trials goes to zero. In that sense, deviations from the 50 percent rule cannot be seen as evidence against the optimism model and in favour of the asymmetric information model.

71 More precisely, the attorney's estimate of the maximum the client should have taken to settle the case. Note that the estimate was only gathered after the lawsuit was concluded. The attorneys were however specifically asked to place themselves in the *ex ante* perspective and were allowed to consult case records. The author also points out that the *ex post* assessment was a good predictor of the attorney's own investment in the case, making plausible that the *ex post* assessment can be used as an alternative for a genuine *ex ante* assessment.

cases in which the plaintiff pays his lawyer on the basis of a contingency fee contract. A plausible explanation for this last result is that plaintiffs opt for contingency fee contracts when they are unable to predict case outcomes well, because such a contract can lower the risk for the plaintiff.[72]

Waldfogel (1998) finds some evidence for the presence of asymmetric information early in litigation: he finds one-sided plaintiff win rates in cases adjudicated prior to trial. Pre-trial litigation and settlement filters out likely plaintiff winners and losers from the pool of filed cases.[73]

Sieg (2000) had access to a data set on medical malpractice disputes in Florida. The set contains information about the insurer's assessment of the severity of the injury, the percentage of claims that are settled, the financial terms of the settlement, the number of settlements that are favourable to the plaintiff, the average compensation in cases that go to court, the defence costs of settled and litigated cases and so on. The technique Sieg uses is as follows. First, he constructs a formal model based on the asymmetric information model by Nalebuff (1987). Then he introduces some assumptions on the distribution of the exogenous variables (e.g. the damages). Finally, he estimates the parameters of the model (litigation costs, compensation awarded in settlement and at trial etc.) and compares them with the empirically observed data. The estimates of most parameters are remarkably close to the observed data. For example, the model predicts bargaining costs for cases settled out of court to be $29,000. This is only 2.3 percent higher than the costs actually observed. The model predicts an average settlement amount of $233,000. The data show a settlement amount of $210,000. Other predictions are less accurate. For example, mean jury awards are $416,000, while the model predicts a mean jury award of $348,000 (Sieg, 2000). All in all, it seems that the asymmetric information model fits the data well.[74]

72 Since under a contingency fee contract, costs are always proportionate to recovery, while under an hourly fee, costs are the same whether the plaintiff wins or loses.

73 With as a consequence rather balanced plaintiff win rates (and not extreme ones) at trial.

74 Froeb (1993) also finds some results consistent with the asymmetric information model in the criminal law context. He finds a positive relationship between the trial rate and the guilty rate at trial, which is consistent with a model in which the defendant is better informed than prosecutors about their probability of being found guilty at trial. Plea bargains will be rejected mostly by defendants who know they are unlikely to be found guilty.

4.11. The influence of the amount at stake on the trial/settlement decision

Most models predict that trial becomes more likely as the amount at stake increases. Everything else being equal, this should mean that average trial awards are larger than average settlement amounts. Many empirical studies do indeed find that average compensation in tried cases is higher than in settled cases. Kakalik *et al.* (1984) for example studied a sample of 513 (U.S.) asbestos injury-related lawsuits in the period January 1980-August 1982. They conducted a survey of defendants, insurers and plaintiff's attorneys for all tried cases during that period, as well as a random sample of claims that settled or were dropped before trial. On average, trials resulted in payments that were more than twice as high as the compensation received by plaintiffs in settled cases. Kakalik *et al.* (1988) studied aviation accident lawsuits for more than 2000 cases filed between 1970 and 1984. Claims that went to trial resulted in an average award of $599,000. Claims that resulted in a lawsuit but were settled before trial averaged $387,000. Claims that settled without a suit settled for $256,200 on average. Note that aviation litigation rarely deals with liability issues, but usually focuses on damages. Black *et al.* (2005) studied medical malpractice litigation in Texas. They had access to a database including all closed medical liability insurance claims over a period of 15 years. The average settlement ranged between $303,000 and $410,000, while the average award in tried cases was $1,528,525. Chandra *et al.* (2005) studied medical malpractice litigation as well. Their data came from the National Practitioner Data Bank. Four percent of all claims resulted in a trial. The average award was roughly twice as large as the average settlement.

4.12. Gaps in empirical research

Empirical research shows that asymmetric information may play a certain role in certain types of disputes and in certain stages of a dispute. However, we do not know the nature of most private information, e.g. information about liability, the amount at stake, the costs of the parties, the degree of risk aversion etc.

B. THE DURATION OF THE SETTLEMENT NEGOTIATIONS

1. General

1.1. Description

When a person suffers harm, he may or may not decide to assert a legal claim. He can do this officially or informally by private communications with the other party. Usually, the parties will attempt to settle the dispute privately. This section deals with the time between the moment that the claim is asserted and the moment that the parties reach a settlement agreement (if at all).[75]

1.2. Basic theories on settlement duration

1.2.1. The duration when parties have the same information

A. The parties have the same expectations: do different incentives lead to delay?
A plaintiff typically has an incentive to settle as soon as possible: he prefers to receive money sooner than later and he wants to keep his costs of negotiation as low as possible. A typical defendant on the other hand has mixed incentives. On the one hand, he has a natural incentive to delay payment as long as possible.[76] On the other hand, he wants to keep his negotiation costs low and he does not want the plaintiff to go to trial. An important question is whether these different incentives of the parties can cause delay? The answer is no. If the parties have the same expectations from the start, they will agree on the settlement amount they would settle for in the future. Consequently, they can settle *right now* as well for a time discounted settlement amount (Wang, Kim, and Yi, 1994).

B. The parties do not have the same expectations: the case of relative optimism

B.1. Can relative optimism cause settlement delay?
Can the existence of relative optimism cause settlement delay? Suppose the parties start negotiating on the 1st of March and that the trial is set for the 1st of November. Can excessive optimism cause the parties to settle at an intermediate date, for example the 1st of June? Intuitively one may reason that each party may be so optimistic about what he will get in the case of a delay that an agreement will not be reached at

75 At least in the theoretical part. Empirical data (see below) sometimes give the duration between accident and date of settlement.

76 At least when damages are awarded with no or with low interest.

the beginning. It is important to note however that excessive optimism does not lead to a delay in settlement when the parties know that both of them will remain sufficiently optimistic for a sufficiently long time. In such a case, the parties will either settle immediately or they will not settle at all. The intuition is simple. By settling early, the parties avoid the costs of settlement delay (negotiation costs etc.). The earlier they settle, the more of these costs are avoided. So if the parties do not settle immediately, they will certainly not settle tomorrow because the costs of delay will be smaller tomorrow,[77] while the degree of optimism remains more or less the same. If the costs of delay are large enough, the parties will settle at the start. Otherwise they will go to trial.[78]

But what if optimism is not necessarily persistent during the settlement negotiations? What if information is generated during settlement negotiations so that the parties can learn and may change their beliefs as they move forward? In that case, delay in settlement becomes possible. The intuition goes as follows. When party A believes he has a strong case, he also believes that the events are likely to proceed in such a way that it will be proven that he has a strong case. In that case, party B will be convinced that A is right and he will be persuaded to agree to A's settlement terms. If the initial beliefs of party B are not too firm, this will happen rather soon and A will find it worthwhile to wait and persuade B. Party B may reason in the same way and find it worthwhile to wait and persuade A. This can cause a delay in settlement.[79]

B.2. The determinants of settlement duration

A model by Watanabe (2005) shows the following elements are the determinants of settlement duration:

The bargaining costs of the parties: Parties will settle sooner when their bargaining costs increase. Intuitively, waiting for valuable information becomes less attractive when the cost of doing so increases.

The severity of the injury: The consequences of differing initial beliefs are more serious when the injury of the victim is more severe. For example, suppose that the plaintiff initially thinks he has a 70 percent chance of winning and the defendant initially thinks the plaintiff has only a 40 percent chance of winning. If the harm caused equals 100, there is an initial difference of 30 between the parties. If the harm caused equals

77 Since part of the costs have already been incurred.

78 For a formal analysis when optimism is sufficiently persistent, see Yildiz (2003).

79 For a formal analysis when the parties' initial beliefs are not too firm and they may learn, see Yildiz (2004).

200, the initial difference is 60. It will take longer for the expectations of the parties to converge. Settlement delay increases with the severity of the injury.

The defendant's estimation of his liability: The more certain the defendant is that he is liable, the faster the parties will settle. Logically, the plaintiff will disagree less often with a defendant who thinks he is liable than with a defendant who thinks he is probably or certainly not liable.

The initial beliefs of the parties: When the initial beliefs of the parties diverge more, it will take longer before the information released causes the expectations of the parties to converge sufficiently for an agreement. More divergent initial beliefs increase settlement delay.

The rate at which new information arrives during bargaining: The faster new information is released, the faster the gap between the initial beliefs of the parties will shrink. Faster revelation of information decreases settlement delay.

1.2.2. The duration when one of the parties has private information relevant for the outcome of the case

A. General

We have seen that asymmetric information is a source of trial. The question to be answered in this section is whether it is also a source of settlement delay. Spier (1992) was the first to formally examine this issue. She built a model in which pre-trial negotiation takes place over several periods.[80] It is assumed that:

1. The trial date is fixed.[81]
2. The uninformed plaintiff makes a finite series of settlement offers to the informed defendant.[82]

80 And this in contrast with the traditional strategic literature on litigation, which assumed that there is just one bargaining period before the possible onset of trial. Logically, such models cannot shed light on the factors which influence the timing of settlement (Spier, 1992).

81 This implies that the trial date is given and thus that the parties have no influence over it. Spier also considers the case in which the trial date is endogenous – the plaintiff has the option to go to court or to make a settlement offer in each round (Spier, 1992).

82 More specifically, there are T period of bargaining prior to the trial. Trial takes place in period T+1. In each period, the plaintiff makes a settlement offer which the defendant must either accept or reject. After a rejection, the game goes on and the plaintiff makes another settlement offer in the next period. If the parties do not agree during the last period, the case proceeds to trial.

3. Both parties discount time at the same rate.
4. The plaintiff never drops the case: the plaintiff goes to court even if the defendant has rejected every settlement offer.

The author finds that when there are no bargaining costs, all settlement occurs just before trial ("on the courthouse steps"). The intuition behind this deadline effect is rather subtle: if the plaintiff makes an offer in the first round and this offer is rejected by the defendant, a rational plaintiff will update his estimate of the defendant's type: since the defendant rejected the offer, he must be among one of the "better" types. A rational plaintiff will consequently lower her settlement offer after a rejection. The defendant, who knows this, would wait to accept so he can receive a better offer. To avoid this, the plaintiff simply waits until the last period to make a take-it-or-leave-it offer, hereby optimally extracting rents from the defendant.

When there are bargaining costs, not all settlement occurs on the courthouse steps. On the one hand, the plaintiff wants to wait until she can make a take-it-or-leave-it offer, but on the other hand she wants to make an acceptable offer so that bargaining costs can be avoided. Consequently, it is not optimal to delay all settlement until the final period. Some agreements will be reached immediately, others on the courthouse steps and the rest at some point in between. In conclusion, asymmetric information is indeed a source of settlement delay.[83]

83 Miceli (1999) offers another explanation for settlement delay that is also based on asymmetric information. In his model, one party has private information about his cost of waiting (and not about something relevant for the outcome of the trial). Like in Spier's model, the pre-trial period has a fixed length, trial is costly (for the plaintiff and the defendant) and so is waiting for trial (for the plaintiff and the defendant). In contrast to Spier, (1) the costs of waiting differ among plaintiffs and defendants do not know which plaintiffs have high and which have low costs of waiting. Moreover, (2) all negotiations take place immediately after the accident. More specifically, the defendant offers a set of contracts combining a settlement offer and the moment in time at which this sum of money will be transferred. For example: the defendant offers 10,000 to all plaintiffs who want to settle immediately and 12,000 to all plaintiffs who are willing to settle just before trial. Although all settlement takes place immediately (or not at all), clearly the transfer of money is delayed. (3) The defendant is capable of making a take-it-or-leave-it offer. He can credibly commit not to re-open negotiations after the plaintiff has selected a contract. The model is thus only applicable to defendants who are repeat players or who can develop a reputation for making binding commitments in some other way.

B. The determinants of the settlement duration

The bargaining costs of the parties: When the bargaining costs increase, the plaintiff will be willing to make offers that are acceptable to more types of defendants more quickly. In this manner, he can avoid the bargaining costs of the next stage(s). Settlement delay decreases when the bargaining costs of the plaintiff increase.

The severity of the injury: Settlements take longer when the severity of the injury increases.[84] The reason is that the difference between the different types of defendants increases. Put otherwise, the adverse selection problem gets worse. An example can make this clear. Suppose there are two types of defendants: those who have a 20 percent chance of being found liable and those who have a 70 percent chance. If the severity of the injury equals 10,000 euros, the difference between the two types equals 5,000 (0.5x10,000). If the injury equals 20,000 euros, the difference between the two types equals 10,000 (0.5x20,000). For the plaintiff it is more risky to try and settle early when the severity of the injury increases. If his offer is rejected, he has to update his estimate on the defendant's type. The larger the differences between the types, the more detrimental such an update is for the plaintiff.

The liability of the defendant: Settlement takes longer when the liability of the defendant increases.[85] Once again, the adverse selection problem gets worse. The difference between different types of defendants increases. For example, suppose there are two types of defendants: those who have caused harm of 10,000 euros and those who have caused harm of 20,000 euros. If the probability of being held liable equals 20 percent, there is a difference between the two types of 2,000 (0.2x10,000). If the probability of being held liable equals 70 percent, there is a difference between the two types of 7,000 (0.7x10,000). For the plaintiff it is more risky to try and settle early when the liability of the defendant increases.

2. Consequences for earlier stages

In the discussion of the next stage (the trial stage), we examine the influence of the duration of trial on previous stages. We can refer here to that analysis. For the incentive to file and the *ex ante* incentives, it does not matter in qualitative terms whether there is delay in settlement or delay in trial.[86]

84 We assume here that the parties have different information about the probability of liability.

85 We assume here that the parties have different information about the extent of damages.

86 Unless of course there are legal rules which treat both types of delay differently.

3. Modelling issues

Most models focusing on the duration of settlement are asymmetric information models (see e.g. Spier (1992)). One interesting model *not* focusing on asymmetric information is the model by Watanabe (2005). The model is quite elaborate and complicated, so here we can only focus on its basic setup. Watanabe constructs a model in which the plaintiff and the defendant initially have different beliefs about the probability of the plaintiff prevailing at trial.[87] New information is revealed during bargaining. The parties learn from this information and update their expectations of obtaining a favourable verdict. Every period of bargaining is costly[88] and costs differ depending on whether or not the plaintiff has filed a lawsuit.[89] Learning has the effect of drawing the expectations of the parties closer together and thus increases the probability of settlement. The amount of delay in equilibrium depends on the following trade-off: on the one hand delay is costly because of the per-period lawyer fees and on the other hand delay may be valuable because of the possibility that new information may be learned.

4. Empirical research

4.1. The deadline effect[90]

Some empirical studies have found a deadline effect in settlement negotiations. For example, a study of a sample of civil cases from Arizona (U.S.) found that 70 percent of all cases were settled during the last 30 days before trial, and 13 percent were settled on the day of the trial itself (Williams, 1983).

87 His model is inspired by a more general model by Yildiz (2004) which does not explicitly focus on litigation.

88 The author assumes that bargaining is costly only for the defendant because he focuses on medical malpractice cases, in which most plaintiff lawyers work on the basis of a contingency fee. With a contingency fee, the plaintiff's legal costs are a fixed fraction of the compensation payment.

89 Costs are generally higher during the litigation phase because it entails additional legal procedures.

90 We would stress that other models of settlement behaviour (especially the principal-agent model) may also explain delay in settlement and possibly the deadline effect. When a lawyer is paid per hour, he may have an incentive to prolong the case. When the case is strong, he will have an incentive to prolong the case until a court verdict. When the case is weak, he may have an incentive to prolong but not to litigate the case. Since the case is weak, litigating causes a high risk of losing in court and this may create negative reputation effects for the lawyer. So, a large fraction of cases (the weak ones) will be settled just before trial.

The existence of a deadline effect does not shed much light on the question of whether bargaining is impeded more by relative optimism or by asymmetric information. The reason is that a deadline effect may emerge in both models. In relative optimism models, the parties may be so optimistic about their bargaining power at the deadline that both of them are willing to wait and reach an eleventh-hour agreement just before the deadline (Yildiz, 2004). In the asymmetric information model, the party who can make an offer may be willing to make his offer on the courthouse steps so that he can optimally extract rents from the other party.

4.2. Empirical research on learning during settlement negotiations

Some empirical research shows that the parties learn during settlement negotiations. Fenn and Rickman (2005) had access to a dataset of all personal injury claims brought against a group of hospitals in one geographic region of the U.K. The dataset contains a total of 3,749 cases, of which 1,542 cases had closed (the rest were still active). The dataset contains periodic information on perceived strength of the claim[91] (not liable – probably not liable – unclear – probably liable – liable) and the severity of the injury (in pounds).[92] The data clearly show that litigation over liability and damages is associated with a learning process in which information about the circumstances of the event and the standard of care adopted by the hospital becomes available and uncertainty is diminished. For example, most cases that begin with an assessment of "unclear" evolve to an estimate that the hospital is "liable" or "not liable". Furthermore, claims have an initial average expected value of £40,783 and a final average expected value of £48,854.

Cremers and Schliessler (2015) use a dataset consisting of around 80 percent of all patent litigation cases in Germany between 2000 and 2008 to examine the within-trial settlement decisions of the parties. They find that the availability of detailed new information through an expert's opinion on the case increases the settlement rate by about 10 percentage points. Such reports make the litigants' estimates of their chances in trial converge.

91 Perceived by the claims manager.

92 Modern claims management procedures force hospital claims managers to record and update their expectations about the settlement prospects for each case.

4.3. The determinants of settlement duration

We have previously examined the determinants of the duration of settlement in the presence of relative optimism and in the presence of asymmetric information respectively. Several studies have examined the empirical relationship between these determinants and the duration of settlement. Before turning to the results of these studies we will briefly describe their data sets.

4.3.1. The data sets of the empirical studies

1. Fournier and Zuehlke (1996) use data from a survey of U.S. civil lawsuits during the fiscal years 1979-1981.[93] All federal courts participated. For every suit, a form was filled in that contains objective descriptions of the nature of the suit[94] and its final disposition.
2. Kessler (1996) uses a sample from a database of 17,588 automobile insurance bodily injury claims. The data come from 34 insurers in the U.S. passenger automobile insurance market and cover the period 1977-1987.[95] For each claim, information is available about the settlement amount, the number of days between the accident and the settlement, the state in which the accident occurred and many of the claim characteristics (weeks lost from work, claimant age, types of injuries etc.).
3. Fenn and Rickman (1999) use data from English health care providers (N.H.S. Trusts) who contributed to a database of clinical negligence and employee claims during the period 1974-1995. From this database, they studied data relating to 759 claims (80 percent clinical negligence cases and 20 percent employer liability cases) which were settled or abandoned between 1990 and 1995. A lot of information was recorded for these cases: the date of initiation of the case, the date of settlement or abandonment of the claim, the defence costs and the damages paid, an estimate by the claims manager of its likely value if the claim was settled and an estimate of the probability that the plaintiff would win if the case went to court.
4. Fenn and Rickman (2001) collected data about 301 closed personal injury claims against the policyholders of a major motor vehicle insurer in 1998 and 1999. Information was available about the payments made at the different stages in the set-

93 The survey was performed by the Division of Information Systems of the Administrative Office of the U.S. Courts.

94 A classification system with 98 categories of cases was used. The categories are organized around the statutes involved in the case (Administrative Office of the U.S. Courts, 1980).

95 This amounted to a 60 percent participation level.

tlement process. These payments were divided into personal injury damages, medical expenses, third-party costs and own costs. Moreover, the data contain estimates by the claims manager about the likely costs and damages at different intervals and the severity of the plaintiff's injury and also contain some information about the likely source of the plaintiff's finance for the claim[96] (e.g. legal aid, legal expenses insurance etc.).

5. Watanabe (2004) examined 3,845 claims against physicians in Florida during the period 1985-1999. For each case, information was available about important dates (calendar date of occurrence, initial claim, filing of lawsuit, resolution either by settlement or by court judgment), settlement payments and court awards, total legal costs incurred by the defendants and characteristics of the injury (e.g. severity).[97]
6. Fenn and Rickman (2005) had access to a database of 3,749 personal injury claims brought against a group of hospitals in one geographic region of the United Kingdom. Roughly stated, the database covers the time period 1980-2000. It contains information about the payments made to claimants and defence lawyers in the settlement process and the dates of these payments. It also contains estimates (at several moments in time) by the hospitals' claim managers about the likelihood that their hospital would be found liable. In addition it contains estimates on the likely costs and damages in the event of the hospital being held liable. The database also provides information about the severity of the patient's injury and about the likely source of funding of the patient's claim (e.g. private funds, legal aid funds etc.).

4.3.2. Empirical results for the determinants

The bargaining costs of the parties: Fenn and Rickman (1999) find that higher defence costs are associated with shorter delays. They also find shorter delays for plaintiffs who are not legally aided compared with longer delays for plaintiffs who are legally aided. Fournier and Zuehlke (1996) also find an inverse relationship between costs and settlement duration. However, they only have information on the public costs of various case types[98] and use the public costs as a proxy for the private costs of litigation (they thus assume that public and private costs are correlated). Fenn and Rickman (2001) find that cases believed by the insurer to be more costly for the plaintiff are as-

96 The senior solicitor in the insurer's claims department was asked to make a judgment about the likely source of finance in light of the evidence in the claim file.

97 In Florida, medical malpractice insurers are required by statute to file a report on all of their closed claims, regardless of their outcome. Reports are collected by the insurance regulator of the Florida state government (the Florida Department of Financial Services).

98 They get these data from Kakalik (1983) and Kakalik and Robyn (1982).

sociated with shorter delays. Fenn and Rickman (2005) find that a greater personal responsibility for bearing the costs of litigation leads to faster settlement and earlier abandonment of a claim.

The severity of the injury: Fenn and Rickman (1999) find that when defendants have a higher *ex ante* expectation of the damage level, it takes longer before a settlement is reached. Kessler (1996) finds that prejudgment interest increases the duration of settlement. Fenn and Rickman (2001) find that cases with higher *ex ante* expected values result in longer delays. Fenn and Rickman (2005) find that claims believed to have a potentially high settlement value take longer to settle than average. They also find that claims perceived to be of high value last longer before being dropped. When the hospital revises its belief about the likely amount of damage upwards, the likelihood of abandonment decreases.

The liability of the defendant: Fenn and Rickman (1999) find that the more the defendant thinks he is liable, the shorter the delay; the less he thinks he is liable, the longer the delay Likewise, Fenn and Rickman (2001) find that cases in which the insurer believes its policyholder is fully responsible are associated with shorter delays. Finally, Fenn and Rickman (2005) find that cases in which a hospital initially believes it is not liable survive much longer before settling compared with cases where the hospital initially believes it is liable. Cases where the hospital initially believes it is liable are abandoned much later by comparison with cases in which the hospital initially believes it is not liable. Furthermore, the arrival of new information weakening the hospital's case speeds up the settlement process and leads to longer durations before a case is dropped.

C. THE SETTLEMENT COSTS

1. General

1.1. Description

The settlement costs are the costs the parties incur during settlement negotiations, mainly to see whether a settlement agreement is possible and if this is the case, to try to obtain the largest possible fraction of the settlement surplus. The costs include:

1. The fees for the services of lawyers: lawyers have to be compensated for negotiating, collecting information, sending letters, studying the law etc.
2. The costs due to the time spent by the parties themselves: interaction with the other party, travelling, instructing lawyers, consulting family and friends etc.
3. Psychological and emotional costs, e.g. stress.

1.2. The basic determinants of the settlement expenditures

Above we examined the basic determinants of the duration of settlement. Taking into account a per-period cost of negotiation, we can make predictions about the basic determinants of the settlement expenditures.

1.2.1. The parties have the same information but have different expectations

We have seen above that Watanabe (2005) constructed a model in which information is released during bargaining and the parties may learn and update their expectations of obtaining a favourable verdict.[99] Learning has the effect of drawing the expectations of the parties closer together and thus increases the probability of settlement. The amount of delay and settlement expenditures in equilibrium depends on the trade-off between the per-period costs and the benefits of new information.

The (per period) bargaining costs of the parties: Parties settle sooner when the (per period) bargaining costs increase. Whether the total settlement costs will increase or decrease is hard to tell. The number of settlement periods decrease but the per-period costs increase.

99 His model is inspired by a more general model by Yildiz (2004).

The severity of the injury: When the severity of the injury increases, it takes longer for the parties to reach an agreement. More settlement expenditures will be incurred.

The defendant's estimation of his liability: The more certain the defendant is that he will be held liable, the faster the parties will agree. Less settlement expenditures are incurred.

The initial beliefs of the parties: When the initial beliefs of the parties diverge more, it will take longer before the information released causes the expectations of the parties to converge sufficiently for an agreement. The settlement expenditures will rise.[100]

The rate at which new information arrives in bargaining: The faster new information is released, the faster the gap between the initial beliefs of the parties will shrink. The settlement expenditures will decrease.[101]

1.2.2. One of the parties has private information relevant for the outcome of the case

As we have seen in the context of the duration of settlement, in Spier's model the defendant has private information and the plaintiff can make a settlement demand at the beginning of each period (Spier, 1992). Spier finds that there are mixed incentives for the plaintiff: one the one hand, he wants to delay making an offer so he can optimally extract rents from the defendant, and on the other hand he wants to settle sooner rather than later to avoid the costs of bargaining.

The (per-period) bargaining costs of the parties: Parties settle sooner when the (per period) bargaining costs increase. Whether the total settlement costs will increase or decrease is hard to tell. The number of settlement periods decrease but the per-period costs increase.

The severity of the injury: Settlements take longer when the severity of the injury increases. This will increase the settlement expenditures.

The liability of the defendant: Settlement takes longer when the liability of the defendant increases. This will increase the settlement expenditures.

100 When the initial beliefs differ too much, the parties will go to trial immediately. In that sense, there is no settlement delay.

101 Holding per unit costs fixed. In reality, faster release of information may be accompanied by increased unit costs.

2. Consequences for earlier stages

2.1. The influence on the incentive to sue

Holding everything else constant, when the cost of settlement increases, potential plaintiffs will be less inclined to file suit because the expected benefits of doing so decrease. Note that in reality, the *ceteris paribus* assumption will not hold. Increasing settlement costs can affect other elements, and this in turn may affect the incentive to sue. For example, increasing settlement costs will lead to a higher trial rate. This will further reduce the plaintiff's incentive to file suit. Also, increasing costs of settlement decreases the surplus from cooperation. This can reduce the settlement amount, and once again lead to a lower incentive for the plaintiff to file suit.

2.2. The influence on *ex ante* incentives

Given that with increasing settlement costs plaintiffs will be less inclined to file suit and will receive smaller settlement amounts, one can expect a negative impact on the *ex ante* incentives of potential injurers. However, the fact that increased settlement costs may also increase the expected loss of the defendant can (partially) offset the other effects.

3. Modelling issues

The models which shed light on the duration of settlement negotiations (Watanabe, 2005; Spier, 1992) can also be deployed to gain insight in the cost of settlement. The reason is that these models assume a per-period cost of negotiation. The cost of settlement will then equal the duration of the settlement negotiations multiplied by the per-period cost.

4. Empirical issues

Very few data are available on settlement expenditures. What we do know is that the ratio of settlement costs to trial costs may be very different depending on the type of dispute. Salop and White (1988) examined about 2,000 (U.S.) antitrust cases filed between 1973 and 1983 (that is about 1/6th of all antitrust suits during that period). They estimated that the litigation costs of cases that were settled amounted to roughly 75 percent of the litigation costs of tried cases. Sieg (2000) provides information on (U.S.) medical malpractice cases. He finds mean defence settlement costs of $28,390 and mean defence trial costs of $60,366.

D. THE OUTCOME OF THE SETTLEMENT NEGOTIATIONS

1. General

1.1. Description

By the outcome of the settlement negotiations we mean the amount of money the plaintiff receives from the defendant as the result of an agreement between them.[102]

1.2. Theories on the outcome of settlement negotiations

1.2.1. Determinants of the settlement amount in relative optimism models

The parties always settle in the relative optimism model whenever there is a bargaining surplus. By introducing a parameter for the relative bargaining power of the parties, more can be said about the settlement amount and its determinants. The settlement amount can be divided into two parts. As a minimum, the plaintiff receives his expected value of trial.[103] On top of that, he receives a fraction of the settlement surplus,[104] unless he has zero bargaining power. The settlement amount will increase 1) when one of these two parts increases while the other part remains unchanged or 2) when one of the parts increases and one decreases, but the increase is larger than the decrease. We now discuss the determinants:

The plaintiff's estimation of his probability of prevailing at trial: The settlement amount increases when the plaintiff's estimation of his probability of victory increases. Although the settlement surplus decreases,[105] the expected value of trial for the plaintiff increases by a larger amount. The reason is that the decrease in the settlement surplus is not as influential because the plaintiff only receives part of this surplus.

The defendant's estimation of the probability that the plaintiff will prevail: The settlement amount increases when the defendant's estimation of the plaintiff's probability of victory increases. The reason is that the settlement surplus increases[106] while the expected value of trial for the plaintiff remains the same.

102 Or the amount that the potential plaintiff receives from the potential defendant.

103 If he would receive less, he would prefer to go to trial.

104 More precisely, his bargaining power multiplied by the settlement surplus.

105 Since the plaintiff wants to receive a higher settlement amount while the defendant is not willing to give more.

106 Since the defendant is willing to pay more.

The amount at stake: The settlement amount increases when the amount at stake increases. Although the settlement surplus may decrease,[107] the expected value of trial increases more. The reason is that the (possible) decrease in the settlement surplus is not as influential because the plaintiff only receives part of this surplus.

The trial costs of the plaintiff: The settlement amount decreases when the trial costs of the plaintiff increase. Although the settlement surplus increases, the expected value of trial for the plaintiff decreases more. The reason is that the increase in the settlement surplus is not as influential because the plaintiff only receives part of this surplus.

The trial costs of the defendant: The settlement amount increases when the trial costs of the defendant increase. The reason is that the settlement surplus increases[108] while the expected value of trial for the plaintiff remains the same.

The bargaining power of the plaintiff: The settlement amount increases when the bargaining power of the plaintiff increases. The plaintiff's expected value of trial remains the same, but the fraction of the settlement surplus that goes to the plaintiff increases.

1.2.2. The determinants of settlement amounts in the asymmetric information model

A. Introduction

Above we have described several types of asymmetric information models. These models shed light not only on the determinants of the settlement frequency but also on the determinants of the settlement amount. The asymmetric information literature often provides very precise predictions about the settlement amount. It has to be stressed that this preciseness is due to the specifications regarding the bargaining procedure. Without these specifications, many equilibrium settlement amounts would be possible.

B. The basic screening model

The amount at stake: An increase in the amount at stake will increase the settlement amount for two reasons. First, an increase in the amount at stake leads to less settlement because the differences between several types of defendants become larger. The defendants that still settle are those who have a relatively high chance of being found liable and are thus those defendants who will accept a greater settlement de-

107 This is the case when the parties are relatively optimistic. When they are relatively pessimistic, the surplus will increase when the amount at stake increases.

108 Since the defendant is willing to pay more.

mand. Second, even if the plaintiff would still settle with the same pool of defendants if a higher amount were at stake, he would raise the settlement demand because these defendants will be willing to pay a higher amount (since they would also be forced to do so if the case went to court) (Bebchuk, 1984).

The litigation costs of the parties: An increase in the plaintiff's litigation costs will decrease the settlement amount. Since going to trial becomes costlier for the plaintiff, he will make a lower offer that will be accepted by a larger fraction of defendants.

An increase in the defendant's litigation costs may decrease, increase or have no effect on the settlement amount. First, if the plaintiff would be forced to settle with the same fraction of defendants, he would increase his demand. The reason is that those defendants are willing to pay a higher demand, since going to trial becomes costlier for them. Second, an increase in the defendant's trial costs makes it more interesting for the plaintiff to settle than go to trial. Only if he settles can he take advantage of the increased trial costs of the defendant. This will make him lower the settlement demand so he can settle with more types of defendants. Which effect dominates depends on the distribution of types (Bebchuk, 1984).

The distribution of types: An upward shift in the range of types raises the settlement amount. Intuitively, this is because raising the settlement offer becomes less costly since the expected value of trial increases (Bebchuk, 1984).[109]

C. The screening model with NEV suits

The amount at stake: An increase in the amount at stake reduces the incentive of the plaintiff to inflate his settlement demand. The credibility problem is less severe when the amount at stake increases: a greater amount at stake allows the plaintiff to learn more bad information and still preserve a credible litigation threat. On the other hand, asking for a higher settlement demand becomes more profitable because the expected loss of all defendants who have a positive chance of being found negligent increases. Because of these two conflicting tendencies, the settlement amount can either rise or fall when the amount at stake increases (Nalebuff, 1987).

The litigation costs of the plaintiff and the defendant: When the litigation costs of the plaintiff increase, he has to make a higher settlement demand to maintain a credible threat if the demand is rejected. An increase in the litigation costs of the defendant

109 Bebchuk also shows that expanding the range of types without changing the mean of the distribution of types might increase, decrease or have no effect on the settlement amount (Bebchuk, 1984, pp. 411-412).

leads to a higher settlement demand, because the defendant's expected loss at trial increases (Nalebuff, 1987).[110]

D. The basic signalling model

In the basic signalling model, each type of (informed) plaintiff makes a settlement demand equal to the (uninformed) defendant's expected loss at trial (Reinganum and Wilde, 1986). From this result, the influence of the different determinants can be easily deducted:

The probability of a plaintiff victory: Since the defendant's expected loss at trial increases when the probability of a plaintiff victory increases, the settlement amount will increase.

The litigation costs of the plaintiff and the defendant: The settlement amount does not change when the litigation costs of the plaintiff increase; the defendant's expected loss at trial does not depend on these costs. The settlement amount increases when the litigation costs of the defendant increase since the defendant's expected loss at trial increases.

The distribution of types: An upward shift in the range of types increases the settlement amount, since the defendant's expected loss at trial increases (for any type of plaintiff).

E. The signalling model with NEV suits

We have seen that a signalling equilibrium can exist in a model with NEV suits under the condition that there is a positive filing cost for the plaintiff (Farmer and Pecorino, 2002). For this reason we would refer to the basic signalling model above to identify the influence of the various determinants on the settlement amount.

1.2.3. A comparison of the relative optimism model and the asymmetric information model

There are some interesting differences between the relative optimism model and several versions of the asymmetric information model. For example, in the relative optimism model, increased stakes always lead to a higher settlement amount. This is also the case in most asymmetric information models, with the exception of the screening model with NEV suits. Another difference concerns the trial costs of the defendant. When these increase, settlement amounts increase in the relative optimism model but

110 Nalebuff does not examine the influence of the distribution of types.

not necessarily in all asymmetric information models (e.g. in the basic screening model settlement amounts may increase, decrease or remain the same).

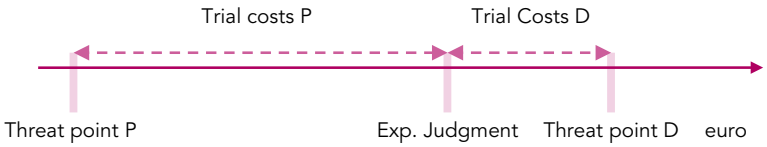
1.2.4. The accuracy of settlement amounts

Are settlements unbiased estimates of trial outcomes? This is an important question since settlements are allowed and judicially approved on the presumption that they are (Wilson, 1998). If (before the trial) a defendant has a 70 percent chance of being found liable by the court and the harm done equals 100, an accurate settlement amount would equal the expected judgment of 70 (0.7x100).[111]

A. Accuracy when parties have no private information

The threat point of the plaintiff equals the expected judgment minus his trial costs. The threat point of the defendant equals the expected judgment plus his trial costs (see figure below). The settlement amount will reflect the expected judgment when the trial costs of both parties are the same and both parties have equal bargaining power. It is unlikely that these two conditions will be systematically fulfilled. Although average settlement amounts may be close to the expected judgment,[112] in individual cases they will tend to be either lower or higher than the expected judgment.

Settlement amounts are more likely to be lower than the expected judgment when the trial costs of the plaintiff are larger than the trial costs of the defendant and also when the bargaining power of the defendant is greater than the bargaining power of the plaintiff. Settlement amounts are more likely to be larger than the expected judgment when the trial costs of the plaintiff are smaller than the trial costs of the defendant and also when the bargaining power of the plaintiff is greater than the bargaining power of the defendant.



111 We will discuss the accuracy of trial outcomes later on. Here we assume that trial outcomes are accurate. This will allow us to focus on inaccuracies due to bargaining itself.

112 This would be the case if trial costs and bargaining power are more or less randomly distributed across plaintiffs and defendants. If one of these conditions is not fulfilled, biases in favour of one party may exist. For example: it is possible that defendants generally have greater bargaining power than plaintiffs. If trial costs are randomly distributed across plaintiffs and defendants, on average settlement amounts will be lower than the expected judgment.

B. Accuracy in asymmetric information models

B.1. The basic screening model

When an uninformed plaintiff makes an offer to an informed defendant, the settlement amount can be either smaller or larger than the expected judgment. Settling defendants with a relatively low chance of being held liable always settle for an amount that is larger than the expected judgment; defendants with a relatively high chance of being held liable may settle for an amount that is smaller than the expected judgment.

This can be illustrated with an example. Suppose there are four types of defendants: those who have a 50, a 60, a 70 and an 80 percent chance of being held liable by the court. The amount at stake is 100. The trial costs of the defendant equal 15. Suppose that the plaintiff's trial costs are such that he will make a demand that will be accepted by all defendants who have a chance of being held liable of 60 percent or more. In other words, he asks for a settlement amount of 75 (0.6x100+15). For a defendant who has a 60 percent chance of being held liable, the settlement amount is larger than the expected judgment (75>60). For a defendant with a 70 percent chance, the settlement amount is also larger than the expected judgment (75>70). For a defendant with an 80 percent chance, the settlement amount is *smaller* than the expected judgment (75 < 80).

B.2. The screening model with NEV suits

We would refer here to the basic screening model.

B.3. The basic signalling model

In the basic signalling model, settlement amounts do not reflect the expected judgment. If the informed plaintiff makes the offer, he will ask for a settlement amount equal to the defendant's expected loss at trial. Clearly, this amount is higher than the expected judgment: it is equal to the expected judgment *plus* the trial costs of the defendant. If on the other hand the informed defendant makes the offer, the settlement amount will be equal to the plaintiff's expected value of trial. This amount is less than the expected judgment: it is equal to the expected judgment minus the trial costs of the plaintiff. In conclusion, the settlement amount is smaller or larger than the expected judgment, depending on who has the opportunity to make a take-it-or-leave-it proposal.

The following example illustrates that settlement amounts do not equal the expected judgment. Suppose there are two types of defendants: those who have a 50 and those who have a 70 percent chance of being held liable by the court. The amount at stake is 100. The plaintiff's trial costs equal 20. If the informed defendant makes the offer, the first type of defendant will make an offer of 0.5x100-20=30. This is clearly lower than the expected judgment of 50 (0.5x100). The second type of defendant will make an offer of 0.7x100-20=50. This is clearly lower than the expected judgment of 70 (0.7x100).

B.4. The signalling model with NEV suits

We have seen that a signalling equilibrium can exist in a model with NEV suits under the condition that there is a positive filing cost for the plaintiff. For this reason, we would refer to the basic signalling model above.

2. Consequences for earlier stages

2.1. The influence on the incentive to file

When the settlement amount increases, holding everything else constant, the incentive of victims to sue will increase. This may not only stimulate strong claims, but also some relatively weak claims.

2.2. The influence on the *ex ante* incentives of injurers

When settlement amounts increase, holding everything else constant, defendants will take more care, will breach contracts less often etc. Not taking care, breaching a contract or more generally not obeying the law becomes more expensive for the potential injurer.

3. Modelling issues

In a divergent expectations setting, creating models which shed light on precise settlement amounts can be done in three ways. First, we can simply introduce a parameter for the relative bargaining power of the parties. Second, we can look for bargaining solutions which satisfy some criteria we suspect any solution should satisfy. Third, we can introduce specific bargaining structures.

A. Introducing bargaining power[113]

The simplest way to make more specific predictions about the settlement amount is to introduce a parameter for the bargaining power of the parties (similarly, Hay and Spier, 1998)). We can define the bargaining power of the plaintiff as the percentage of the settlement surplus that he will obtain. Likewise, we can define the bargaining power of the defendant as the percentage of the settlement surplus that he will obtain.[114] The sum of the bargaining powers should of course be equal to 1. For example, when the plaintiff's expected value of a trial is 80 and the defendant's expected loss is 100 and the parties settle for 95, then the plaintiff's bargaining power equals 0.75 (15/20) and the defendant's bargaining power equals 0.25 (5/20). Settlement amounts can thus be divided into two parts. As a minimum, the plaintiff receives his expected value of trial.[115] On top of that, he receives a fraction of the settlement surplus,[116] unless he has zero bargaining power. The settlement amount will increase 1) when one of these two parts increases while the other remains unchanged or 2) when one of the parts increases and one decreases, but the increase is larger than the decrease.

B. The Nash Bargaining Solution

Some authors use the Nash Bargaining Solution to make predictions about the settlement amount (Bebchuk, 1996). Under this solution, the parties divide the joint surplus from settlement equally.[117] However, the Nash Bargaining Solution is a *cooperative* bargaining solution. This means that it is only valid when the parties have agreed that the solution should be efficient. In reality, bargaining is typically *non-cooperative*: each party tries to maximize his own pay-off rather than the joint surplus.

113 On the various sources of bargaining power, see Muthoo (1999).

114 Of course, this tells us nothing about the bargaining power of the parties in reality.

115 If he would receive less, he would prefer to go to trial.

116 More precisely, his bargaining power multiplied by the settlement surplus.

117 Note that this equal division is not an assumption but is derived from a set of axioms that Nash felt any bargaining solution should satisfy. The axioms are the following: 1) The solution should not depend on how the utility scales of the players are calibrated, 2) the solution should be efficient, 3) if the players sometimes agree to one outcome when another is also possible, then they never agree to that other outcome when the first one is feasible and 4) when players are identical, both should receive the same pay-off (Nash, 1953).

C. Introducing a specific bargaining structure

C.1. One take-it-or-leave-it proposal

When a "fully rational" party can make a take-it-or-leave-it proposal, he will make a proposal that is very one-sided. Suppose the plaintiff expects to gain 50 from trial and the defendant expects to lose 100 at trial. If the plaintiff can make the only offer, he will ask for a settlement amount just below 100 (e.g. 99.9). A "fully rational" defendant will accept this. If he does not accept it the parties will go to trial and the defendant will expect to lose more (100 > 99.9). If the defendant can make the only offer, he will propose a settlement amount just above 50 (e.g. 50.1). A "fully rational" plaintiff will accept this. If he does not accept it the parties go to trial and the plaintiff will expect to win less (50 < 50.1).

C.2. More than one proposal

When both parties can make proposals, it is not always possible for a party – even the one who makes the first offer – to get the entire settlement surplus. For example, Wang, Kim and Yi (1994) created a model in which both parties can make settlement proposals one after the other. The defendant makes the first offer, after which the plaintiff accepts the offer, makes a counteroffer or goes to court. If the plaintiff makes a counteroffer and the defendant accepts, the dispute ends. If he does not accept, the parties move again to the next period in which the defendant makes a counter-counteroffer, and so on. Unlike in the case of a single take-it-or-leave-it proposal, settlement amounts are not necessarily one-sided. The party who makes the first proposal (here the defendant) cannot always settle at the other party's concession limit. If the costs of negotiation are relatively high for the offering party (compared with the costs of the other party), the parties will settle for an amount between the threat points of the parties. If the costs of negotiation are relatively low for the offering party, the party who makes the first offer can still settle at the concession limit of the other party.

4. Empirical issues

4.1. Do parties settle for the expected judgment?

Little is known about actual settlement amounts. Most importantly, we do not know whether the parties generally settle for an amount close to the expected judgment.

Such knowledge is important because most parties settle their disputes. Consequently, the *ex ante* behaviour of potential injurers (taking precautions, undertaking a certain activity) is largely driven by what they expect to pay as a settlement amount.

4.2. Experimental research on the importance of fairness

Experimental research shows that in simple ultimatum bargaining games,[118] the party who can make a take-it-or-leave-it offer typically offers one-third to one-half of the surplus to the other party. Many laboratory experiments found the equal split as the modal outcome. Offers of one-third or less of the surplus are frequently rejected.[119] These findings are in contrast with standard "fully rational" economic theory, which predicts that the offerer will offer and the offeree will accept a very one-sided offer. The experiments show the importance of fairness in determining how two parties divide joint surplus.

On the basis of this experimental research, one should not conclude too hastily that the parties in a dispute will usually split the settlement surplus when one party can make a take-it-or-leave-it proposal. In the simple ultimatum bargaining games, the players are explicitly asked to divide a given surplus. In settlement negotiations however, the parties negotiate implicitly rather than explicitly over the surplus. When the expected value of a trial for the plaintiff equals 100 and the expected loss of the defendant equals 150, the parties do not simply regard themselves as bargaining over a surplus of 50. Pecorino and Van Boening (2004) conduct an experiment in which the bargaining game is framed in greater accordance with actual settlement bargaining. They call this the "embedded ultimatum game". Their results differ strongly from the results of the simple ultimatum bargaining games. Offers in the embedded game are much more selfish (one-sided) than in the simple game. On average, the offerer keeps 87 percent of the surplus and grants only 13 percent of the surplus to the other party. The non-offering party has a greater willingness to accept low offers in the embedded game than in the simple game.

Some researchers have examined the settlement offers of uninformed defendants to informed plaintiffs in an experimental setting (Pecorino and Van Boening, 2014). They find that offers by the defendant contain about 27 percent of the settlement surplus.

118 In such games, two parties have to split a fixed amount of economic surplus. This is often called "the pie". One party can make a take-it-or-leave-it offer to the other. If that other party rejects the offer, the entire pie is lost (the parties get nothing).

119 Overviews of the literature on ultimatum bargaining games can be found in Thaler (1988), Roth (1995) and Fehr and Schmidt (2000).

Given the rejection behaviour of plaintiffs, they find that an optimal offer contains about 17 percent of the joint surplus from settlement. In other words, defendants make offers and plaintiffs accept offers that are close to the offers the rational models predict. The offers are somewhat higher than the rational models predict. The reason is probably that fairness plays a role in settlement negotiations.[120]

120 Fairness appears to be defined over the joint surplus from settlement and not over the plaintiff's dispute costs.

4

Stage IV. The trial stage

A. TRIAL EXPENDITURES AND TRIAL OUTCOME

1. General

1.1. The basics of endogenous litigation spending

Law and economics scholars have traditionally focused more on the reasons for trial than on the trial itself. More recently however, some studies have paid more attention to the legal battle itself. These studies have analyzed the impact of legal rules on the expenditure decisions of the parties at trial and have examined how these effects may change some standard results of the literature. Many of these articles use an explicit function for the probability of a plaintiff victory. These litigation success functions usually take the following form:

$$p(X,Y) = \frac{X^a F}{X^a F + Y^a (1-F)} = \frac{(\frac{X}{Y})^a F}{(\frac{X}{Y})^a F + (1-F)}$$

with X and Y being the effort levels of the plaintiff and the defendant respectively, F the degree of defendant fault (with $0 \leq F \leq 1$), and a the force exponent that indicates the relative importance of expenditures vis-à-vis the defendant's fault (or briefly, a productivity parameter of the effort level). This function obviously has the following features (Hirshleifer and Osborne, 2001):

1. The determinants of success include both the degree of defendant fault and the litigation effort of the parties;

2. The probability of a plaintiff victory depends on the ratio of the effort levels, hence the name "ratio form" that is used for this functional form;
3. For a very close case (F=1/2), the outcome at trial depends only on the litigation efforts;
4. Given equal efforts, the outcome depends only on the degree of fault;
5. If the defendant is totally in the wrong (F=1), she always loses as long as the plaintiff makes some effort (X>0);
6. If the defendant is totally without fault (F=0), she always wins as long as she makes some effort (Y>0).

Some authors have used a simplified version of this function such as

$p(X,Y) = \frac{XF}{XF + Y(1-F)}$, implicitly assuming that a=1, or $p(X,Y) = \frac{X}{X+Y}$, implicitly assuming that F=1/2 and a=1. Some other authors use a slightly different function,

$p(X,Y) = \frac{e^M X^a}{e^M X^a + Y^a}$, with M reflecting the objective merits of the case (with M varying between –∞ and +∞) (Katz, 1988). When $M > 0$, the merits favour the plaintiff, when M <0 the merits favour the defendant, and when M=0 they favour neither party. However, it is easy to see that both functions are equivalent. We can see this by setting F equal to $\frac{e^M}{e^M + 1}$. When M = –∞, F=0, when M=0, F=1/2, and when M=+∞, F=1.

A simple model using this function unveils the basic determinants of the effort levels of the parties and of the plaintiff's probability of victory. In that basic model, both parties select their level of legal effort once and simultaneously. The effort levels are investments made to persuade a fact-finder. The amount at stake equals J. Both parties are assumed to be risk-neutral. Each contender aims to maximize his expected income. Each party is assumed to be responsible for his or her own legal expenses regardless of the outcome (the U.S. rule of cost allocation applies). The actual merit of the claim is known by both litigants but is unknown to the court. The unit costs of legal effort equal Cp for the plaintiff and Cd for the defendant.[1] Consequently, the

1 Cp and Cd can for example be seen as the hourly fee of the litigant's lawyer.

plaintiff spends CpX and the defendant CdY. We obtain solutions for the litigation efforts under the Nash-Cournot protocol.[2]

The plaintiff wants to maximize his/her expected value:

$$EV_{pl} = \frac{X^a F}{X^a F + Y^a (1-F)} J - C_p X$$

The defendant wants to minimize his/her expected loss:

$$EV_{def} = \frac{X^a F}{X^a F + Y^a (1-F)} J + C_d Y$$

The first-order conditions are:

$$aF(1-F)X^{a-1}Y^a J = C_p (X^a F + Y^a (1-F))^2$$
$$aF(1-F)X^a Y^{a-1} J = C_d (X^a F + Y^a (1-F))^2$$

From these equations, it follows that in equilibrium $X = CdY$.[3] In other words, the plaintiff spends the same amount as the defendant, independent of the inherent quality of the parties. Now we know that $CpX = CdY$ in equilibrium, we can derive the effort levels of the parties:

$$X^* = \frac{F(1-F)}{C_d (F + \frac{C_p}{C_d}(1-F))^2} J$$

$$Y^* = \frac{F(1-F)}{C_p (\frac{C_d}{C_p} F + (1-F))^2} J$$

The equilibrium expenditures equal:

2 The Nash equilibrium is a solution concept of a non-cooperative game involving two or more players in which each player is assumed to know the equilibrium strategies of the other players, and no player can gain by altering only his or her own strategy. If each player has chosen a strategy and no player can benefit by changing his or her strategy while the other players keep their strategies unchanged, then the current combination of strategy choices and the corresponding pay-offs constitutes a Nash equilibrium.

3 We can easily see this by dividing the left-hand sides of the equations and the right-hand sides.

$$C_pX^* = \frac{C_pF(1-F)}{C_d(F+\frac{C_p}{C_d}(1-F))^2}J$$

$$C_dY^* = \frac{C_dF(1-F)}{C_p(\frac{C_d}{C_p}F+(1-F))^2}J = \frac{C_pF(1-F)}{C_d(F+\frac{C_p}{C_d}(1-F))^2}J = C_pX^*$$

From these conditions, it follows that the parties always spend an equal amount regardless of the level of fault ($C_pX^* = C_dY^*$ for all F).

The expenditures of the parties increase with the productivity parameter a and with the amount at stake J. They are highest for close cases ($F=1/2$) and lowest for cases with extremely high or low merit ($F=0$ and $F=1$).

The plaintiff's probability of winning equals $P_{pl}(X^*,Y^*) = \frac{F}{F+\frac{C_p}{C_d}(1-F)}$ and thus equals the exogenous quality of the case whenever the unit costs of the parties are equal.

1.2. Endogenous litigation expenditures in a divergent expectations framework

Recently, De Mot and Depoorter (2016) have expanded the endogenous expenditures framework to incorporate relative optimism. Conventional understanding in economic models of litigation holds that optimism unambiguously reduces litigants' willingness to settle.[4] Accordingly, if one or more parties are overly optimistic about the

4 Shavell (1982) set out the seminal analysis of the litigation-settlement decision, suggesting that litigation results from excessive optimism on the part of plaintiffs and defendants. For statements reflecting the general understanding of the effect of optimism on the bargaining range, see, e.g., Posner (1992) "case will be settled if one party is more pessimistic than the other… In general, litigation will occur only if both parties are optimistic about the outcome of the litigation"; Cooter and Ulen, (2004), "the greater the amount by which the plaintiff's estimate of the likelihood of winning exceeds the defendant's, the smaller the tendency toward settlement"; Bar-Gill, (2006) "A higher level of optimism entails fewer settlements (and consequently more costly trials)"; Shavell (2004) "the greater the amount by which the plaintiff's estimate of the likelihood of winning exceeds the defendant's, the smaller the tendency toward settlement". The more formal representation of this insight is as follows: if *Pp* (*Pd*) is the plaintiff's (defendant's) subjective probability of the plaintiff winning in court, *J* is the amount at stake and *Cp* (*Cd*) is the plaintiff's (defendant's) trial costs, then the settlement surplus equals *Cp* + *Cd* – (*Pp* – *Pd*)*J*. The settlement surplus increases if *Pd*>*Pp* and decreases if *Pp*>*Pd*.

outcome in trial, there may be no "mutually acceptable settlement amount".[5] Quite counterintuitively, De Mot and Depoorter show that optimism can also *expand* the settlement range. By increasing the perceived value at stake in litigation, optimism may induce parties to invest additional resources in a dispute, which increases the overall bargaining range. Because of the strategic nature of litigation expenditures, optimistic litigants may spend an amount that outweighs the negative impact of optimism on the bargaining surplus. In the equilibrium with optimism, the increase in total expenditures may be larger than the difference in the perceived expected judgment. In those instances, optimism increases the settlement range. Whether optimism increases or decreases the settlement rate, ultimately depends on whether, in concrete instances, the negative effects of optimism on the bargaining surplus outweigh the positive effects created by the additional investments in litigation. These insights have implications for the selection of disputes for litigation. De Mot and Depoorter show that optimism has a different effect on relatively weak and strong cases. Plaintiff optimism, for instance, is more likely to increase the settlement surplus when involving relatively weak claims.

1.3. Endogenous litigation expenditures in an asymmetric information setting

Farmer and Pecorino (2013) build an asymmetric information model with endogenous litigation expenditures. The main focus of the model is how endogenous expenditures can change some of the standard results concerning pre-trial disclosure and discovery. In the model, the uninformed party makes the offer. They find that the issue of endogenous spending at trial can either strengthen or weaken the bargaining position of the uninformed party with the player types who settle. When the bargaining position is strengthened, some standard results on information transmission prior to trial change. The receiver of the offer with a weak case may make a costly voluntary disclosure. Furthermore, the party making the offer may refuse costless discovery. Each of these results are in contrast with the standard results in the literature derived from models in which spending at trial is considered to be exogenous.

1.4. The plaintiff faces a budget constraint

The more substantial the costs of litigation, the higher the chance that individuals will not have sufficient wealth or liquid assets to pursue a claim. Unsurprisingly, experts estimate that four-fifths of people with low income in the United States have no access to an attorney when they need one (New York Times, 2011). In many European

5 Hay and Spier (1998) write "if both parties are sufficiently optimistic about their prospects in court… then there may be no mutually acceptable settlement amount".

countries, legal aid or legal expenses insurance may be helpful, but these systems are far from complete. Many households do not take out insurance and, despite having a relatively low income, are not eligible for legal aid. Note further that budget constraints are not only a problem for individuals, but also for some companies. Consider two companies that are litigating over a failed business venture. For one company, the business venture was one of many it pursues each year and the amount at stake is just a small fraction of the company's assets. For the other company however, the business venture was central to the company's development and all of its financial resources were devoted to it. While the large company will often be able to hire the best counsel money can buy, the smaller company will likely be unable to gather the financial resources necessary to hire top counsel to litigate the case.

There fact that litigation is (sometimes very) costly has two important consequences. First, many claims cannot be brought to court because their expected benefit (the probability of winning multiplied by the award in the case of a victory) is smaller than their costs.[6] In other words, these claims have negative expected value. From a societal point of view, this is not a problem for weak cases, but many meritorious cases also suffer from this problem. When deserving plaintiffs do not bring suit they go uncompensated. This lowers the expected cost of engaging in activities that pose a risk to individuals with low wealth, which results in suboptimal deterrence of wrongful behaviour in such activities. As discussed in the part on the *ex ante* stage, Rubin and Sheperd do indeed find that there is less deterrence of wrongful behaviour directed towards lower income groups.

Second, even when claims have positive expected value, a plaintiff may be unable to pursue the claim because he or she does not have the minimum funds required to do so. Similarly, a plaintiff may have enough funds to bring the claim to court, but not enough funds to invest in the case as vigorously as he or she would have done with-

6 Note that since lawsuits often involve a lot of uncertainty for claimants, even cases in which the statistically correct expected benefit of the case is larger than the costs may not be brought because of risk aversion.

out the budget constraint (X in the model above).[7] Suboptimal investment or a complete lack of investment may once again result in meritorious cases going undercompensated, with suboptimal deterrence as a consequence. De Mot and Faure (2016) show that the consequences of eliminating a budget constraint on the side of the plaintiff (e.g. through third party financing of litigation) are not as straightforward as one might think. Intuitively, one may think that with the elimination of a budget constraint, the parties will spend more. De Mot and Faure show that due to strategic interaction between the parties, total expenditures can either decrease or increase. Interestingly, the ability of the plaintiff to spend more is more likely to reduce total litigation expenditures for relatively strong claims than for weaker claims. Intuitively, this makes sense. For weaker claims, when the plaintiff spends more due to vanishing budget constraints, the case becomes closer, and it becomes more worthwhile for the defendant to spend additional resources as well. For relatively strong claims, when the plaintiff spends more, the case becomes even less close and it becomes less valuable for the defendant to spend more.

1.5. Risk

Despite the risky nature of rent-seeking activities such as litigation, most of the literature has studied rent seeking with risk-neutral agents. Only a few papers examine the impact of risk aversion on rent-seeking effort. Most of these papers find that risk aversion decreases rent-seeking efforts (e.g. Hillman and Katz (1984) and Skaperdas and Gan (1995)). Konrad and Schlesinger (1997) however find that a contest with risk-averse players may dissipate more rents than the same contest with risk-neutral players. More recently, Treich (2009) finds that risk aversion always decreases rent-seeking efforts compared with risk-neutrality, on the condition that the risk-averse rent seekers are also "prudent". Formally, being prudent means that the rent seekers have a convex marginal utility function within the expected utility model ($u''' \geq 0$). Risk aversion together with failure of prudence increases rent-seeking efforts. However, Treich assumes that if all rent seekers exert the same level of effort, they have the same probability to get the rent. This assumption of homogeneity is quite restrictive. In the litigation context, this would mean their model is relevant especially for disputes in which

7 This can explain the success of third party financing. With respect to Burford, one of the largest providers of capital for litigation and arbitration in the world, Molot writes: "[a]lthough Burford's capital has been used by different businesses for different purposes, as a general matter Burford's financing has enabled those businesses to retain higher-quality counsel and/or mount a more vigorous prosecution of a case than would have been pursued in some form even without Burford financing. Each of the lawsuits that Burford financed likely would have been pursued in some form even without Burford financing" (Molot, 2014, p. 179).

in the inherent merits of the case are balanced ($F=1/2$). The extent to which the model applies to relatively weak and relatively strong cases is uncertain (these are cases with heterogeneity).

2. Consequences of litigation costs

2.1. The influence of litigation costs on the settlement frequency

In a divergent expectations framework, when the sum of the litigation costs of the parties increases, the settlement surplus increases. Usually, a larger settlement surplus is associated with an increased chance of settlement. The intuition is straightforward: parties have an incentive to settle (*inter alia*) in order to avoid the costs of a trial. When those costs increase, the incentive to settle increases.

However, not all theoretical frameworks predict that increased litigation costs will also increase the settlement rate. First, when there is a settlement surplus, parties need to reach an agreement on how to divide it. They may fail to reach such an agreement (Cooter, Marks and Mnookin, 1982). When the settlement surplus is large, the conflict between the parties can in theory be larger than when there is less to divide between the parties. Also, although most models dealing with asymmetric information conclude that increased litigation costs are associated with a higher settlement frequency, not all such models lead to this conclusion. In Nalebuff (1987) for example, the opposite finding emerges. In this last model, increased trial costs for the plaintiff induce the plaintiff to raise the reservation settlement demand in order to maintain a credible threat of going to trial. We have discussed this model extensively in the section on settlement.

2.2. The influence of litigation costs on the settlement amount

When the trial costs of the plaintiff increase, the plaintiff will get a smaller settlement amount if the case settles. When the litigation costs of the defendant increase, the plaintiff obtains a larger settlement amount, at least if the plaintiff has some bargaining power.

2.3. The influence of litigation costs on settlement duration

As we have seen in the section on the duration of settlement, higher costs lead to shorter settlement delays.

2.4. The influence of litigation costs on the incentive to file suit

Increasing litigation costs can affect the decision to sue in different ways. First, increases in costs for the plaintiff can decrease the plaintiff's expected value of trial, as well as the settlement amount. This will tend to reduce the incentive to file. Second, the fact that increased litigation costs may increase the settlement frequency may (partially) offset this effect. Third, increased costs for the defendant may increase the settlement amount, and thus the plaintiff's incentive to sue.

We will now discuss the determinants of the decision to file which follow from an analysis of the plaintiff's expected value of suit given endogenous expenditures:

$$\text{EVpl} = \frac{F}{F + \frac{C_p}{C_d}(1-F)} J - \frac{C_p F(1-F)}{C_d (F + \frac{C_p}{C_d}(1-F))^2} J - C_p F$$

with F being the inherent merits of the case, J the amount at stake, Cp and Cd the unit costs of the plaintiff and the defendant respectively, and CF the costs of filing.

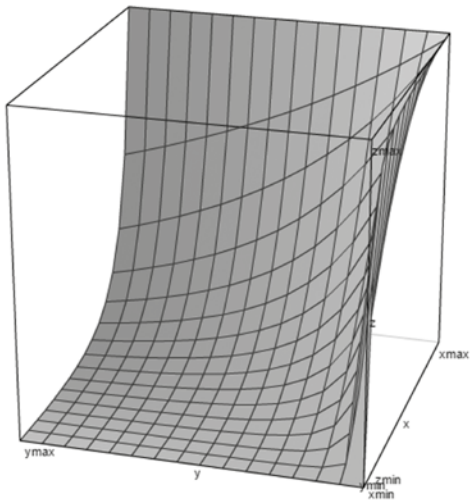
A comparative statics analysis shows how the plaintiff's expected value changes with the various key determinants. Here, we are especially interested in the influence of the inherent merits of the case and the ratio of the unit costs of the parties, because these elements are new compared with a model with exogenous litigation costs.

a. The plaintiff's expected value increases with the inherent quality of the case. Intuitively, this makes sense. An increasing case quality will always lead to a larger probability of a plaintiff victory. And although it may sometimes also increase the

litigation expenditures,[8] this increase does not outweigh the increase in the expected judgment for the plaintiff.

b. The plaintiff's expected value decreases with an increasing ratio of the unit costs of the parties. Intuitively, a higher ratio of unit costs can be interpreted in such a way that access to evidence becomes more difficult for the plaintiff. This must have a negative effect on his expected value and thus on his incentive to sue.

The figure below illustrates these findings. The x-axis represents the inherent quality of the case F (which varies between 0 and 1). The y-axis represents the ratio of the unit costs of the parties (which varies between 0 and 5 in the figure). The z-axis represents the expected value of the plaintiff. Clearly, the expected value increases with F and decreases with Cp/Cd.



2.5. The influence of litigation costs on deterrence

An increase in litigation costs may influence deterrence in several ways. First, when the plaintiff's costs increase, this may decrease deterrence, because fewer plaintiffs will file and the plaintiffs who do file may obtain a smaller settlement amount. More subtly, increasing costs for the plaintiff may increase the settlement frequency, which reduces the defendant's costs, which also decreases deterrence. When the defend-

8 For example, parties will spend more in relatively balanced cases (F=0.5) than in weak cases (F=0.2). See stage IV for a more elaborate analysis.

ant's costs increase, deterrence may be increased because settlement amounts may increase and because defendants bear larger total losses in the case of a trial.

3. Modelling issues

3.1. Optimism about the amount at stake

Here we briefly show how optimism about the amount at stake can be incorporated in a model with endogenous litigation spending. Once again, the following function represents the plaintiff's probability of success:

$$P(X,Y) = \frac{X^a F}{X^a F + Y^a (1-F)}$$

The same notations as before apply. The difference is that J now represents the amount at stake according to the defendant, and βJ (β times J) the amount at stake according to the plaintiff (with β>1 in the case of relative optimism). For the sake of simplicity, we leave out the unit costs of litigation Cp and Cd.

The expected value of litigation for the plaintiff equals:

$$EV_{pl} = \frac{X^a F}{X^a F + Y^a (1-F)} \beta J - X$$

The expected loss of litigation for the defendant equals:

$$EL_{def} = \frac{X^a F}{X^a F + Y^a (1-F)} J + Y$$

The interior Nash equilibria are:

$$X^* = \frac{a\beta^{a+1} F(1-F)J}{(\beta^a F + 1 - F)^2}$$

and

$$Y^* = \frac{a\beta^a F(1-F)J}{(\beta^a F + 1 - F)^2}$$

The settlement surplus (ELdef - EVpl) equals:

$$\frac{a(\beta+1)\beta^a F(1-F)J}{(\beta^a F + 1 - F)^2} - \frac{(\beta-1)FJ}{F + \frac{1-F}{\beta^a}}$$

The settlement surplus is larger with relative optimism than without it if:

$$\frac{a(\beta+1)\beta^a F(1-F)J}{(\beta^a F + 1 - F)^2} - \frac{(\beta-1)FJ}{F + \frac{1-F}{\beta^a}} > 2aF(1-F)J$$

The first term on the left-hand side of the inequality represents the trial expenditures when there is optimism. The second, negative term, represents the difference in the parties' perception of the expected judgment. The term on the right-hand side of the inequality represents the trial expenditures without optimism.

Note that the condition is more likely to be fulfilled for claims that have relatively little merit. For a given degree of optimism, low-quality cases are more likely to get settled. Formally, if we divide both sides of the inequality by $F(1\text{-}F)$, we can see that the first (positive) term on the left-hand side of the equality increases with F, the second (negative) term on the left-hand side decreases with F, while the right-hand side does not change with the merits of the case.

3.2. Optimism about the merits

Instead of introducing optimism about the amount at stake, one may also formally examine optimism about the merits of the case. This can be done by making the inherent quality of the case subjective: Fpl for the plaintiff and Fdef for the defendant, with Fpl > Fdef. However, as it turns out, this makes finding a closed form solution very complicated, and alternative methods may be required.

3.3. Plaintiff does not have sufficient funds

What changes when the plaintiff faces a budget constraint B? Starting from the general model, this means that plaintiffs may not be able to adopt an effort level of X^* and thus invest an amount of CpX^*.

Obviously, when $B \geq CpX^*$, the effort levels of the parties remain X^* and Y^*. In that case, the plaintiff has sufficient funds to adopt an effort of X^*. In the opposite case however, thus when $B < CpX^*$, the plaintiff cannot choose X^* and is forced to choose a lower effort. We can easily see that the plaintiff will choose an effort level equal to B/Cp: between an effort of zero and X^*, his expected value increases concavely for all values $\frac{\partial^2 EV_{pl}}{\partial X^2} = -2\frac{YF^2(1-F)}{(XF+Y(1-F))^3} < 0)$. Consequently, if X^* is not possible, the maximum possible effort level B/Cp is the next best thing for the plaintiff. Given that the plaintiff chooses an effort level of B/Cp, it follows that the defendant chooses an effort level equal to $\frac{\sqrt{\frac{F(1-F)BJ}{C_pC_d}} - \frac{BF}{C_p}}{1-F}$.[9] So when the plaintiff faces a budget constraint, the parties' expenditures equal B for the plaintiff, and $Cd\frac{\sqrt{\frac{F(1-F)BJ}{C_pC_d}} - \frac{BF}{C_p}}{1-F}$ for the defendant.

9 Technically, this follows from the second first-order condition.

What is important is that the sum of the litigation expenditures is not necessarily smaller when the plaintiff has a budget constraint. The total litigation expenditures are smaller without the budget constraint than with the budget constraint when the following conditions are satisfied:

$$\frac{2C_pF(1-F)J}{C_d(F+\frac{C_p}{C_d}(1-F))^2} < \mathrm{B} + \mathrm{Cd}\frac{\sqrt{\frac{F(1-F)BJ}{C_pC_d}} - \frac{BF}{C_p}}{1-F}$$

$$\mathrm{B} < \frac{C_pF(1-F)J}{C_d(F+\frac{C_p}{C_d}(1-F))^2}$$

With respect to the first inequality, the left hand side represents the sum of the litigation expenditures of the plaintiff and the defendant when there is no budget constraint. The right hand side represents the sum of the plaintiff's and the defendant's expenditures when the plaintiff faces a budget constraint. The second inequality represents the fact that the plaintiff will not lack funds for *all* cases. Whether or not the plaintiff lacks funds depends on the amount at stake, the unit costs of the parties and the merits of the case (see the right-hand side of the second inequality, which is simply equal to the plaintiff's optimal expenditure level). Suppose the plaintiff has funds equal to $20,000. When the merits of the case equal 0.8 (a rather strong case), the amount at stake equals $100,000, and the unit costs of the plaintiff and the defendant are the same, then the plaintiff would normally spend $16,000, and he can do so without third-party funding. However, if everything else is held constant, except that the merits of the case are now equal to 0.5, then the plaintiff would normally want to spend $25,000. This time, he is not able to spend this amount, and would need to acquire additional funding to reach his optimal expenditure level.

3.4. Complex legal rules

In the basic model of endogenous litigation spending, whether the plaintiff wins or loses depends on one single issue. And if he wins, he obtains full damages. In reality, things are often more complex. For example, in contract law, even if a defendant did not perform his contractual duties (the first issue), he may be fully excused in some

circumstances (a second issue). In tort law, when both parties were negligent and comparative negligence applies, each party will bear part of the damages. The question is how these more complex legal rules should be modelled.

A more complex model can deal with these issues as follows. At trial, both parties choose their level of spending on legal assistance. There are two choice variables on each side: the litigation effort regarding the first issue and the litigation effort regarding the second issue. Using a standard contest function, the probability that the defendant will lose on the first issue is:

$$P_d = \frac{X_1 F_1}{X_1 F_1 + Y_1 (1 - F_1)} = \frac{1}{1 + \left(\frac{Y_1}{X_1}\right)\left(\frac{1 - F_1}{F_1}\right)}$$

where X_1 equals the expenditures of the plaintiff, Y_1 equals the expenditures of the defendant and F_1 is the inherent quality of the issue ($0 \leq F_1 \leq 1$).

Similarly, the probability that the defendant will win on the second issue (which will for example excuse him) equals:

$$P_p = \frac{Y_2 F_2}{Y_2 F_2 + X_2 (1 - F_2)}$$

where $X2$ equals the expenditures of the plaintiff for the second issue, $Y2$ equals the expenditures of the defendant for the second issue and $F2$ is the inherent quality of the issue ($0 \leq F2 \leq 1$).[10]

Under *a rule in which the defendant is fully excused if he wins on the second issue*, the plaintiff's expected value equals:

$$EV_{pl} = P_d(1 - P_p) - X_1 - X_2$$

The first term, $P_d(1 - P_p)$, represents the situation in which the defendant loses on the first issue (probability P_d) and on the second issue (probability $1 - P_p$).

10 Note that there is no systematic relationship between Pp and Pd (e.g. their sum is not equal to one). The reason is that Pp and P_d are differently defined as is typical in the economic analysis of civil procedure. In this model, Pp and Pd concern the estimates of two different issues, not the overall estimates that the plaintiff will win at trial.

Similarly, the defendant's expected loss equals:

$$EL_{def} = P_d(1 - P_p) + Y_1 + Y_2$$

Under *a rule in which damages are shared when the plaintiff wins on the first issue and the defendant wins on the second issue*, the expected value of the plaintiff's claim equals:

$$EV_{pl} = P_d(1 - P_p) + P_pP_d\sigma - X_1 - X_2$$

The first term, $P_d(1 - P_p)$, represents the situation in which the plaintiff wins on both issues. In that case, the plaintiff is awarded the amount at stake. The second term, $P_pP_d\sigma$, represents the situation in which the plaintiff wins on the first issue and the defendant wins on the second issue. In that case, the plaintiff only receives partial compensation ($0 < \sigma < 1$). The last two terms represent the expenditures of the plaintiff (regarding the defendant's negligence and the plaintiff's negligence).

Obviously, the plaintiff's share σ in the event that both parties win on an issue will depend on the inherent quality of both issues and on the expenditures of the parties. For example, even though the inherent quality of the first issue may be large (F_1 is large) and the inherent quality of the second issue may be small (F_2 is small), the plaintiff's share in the event that both parties win on one issue may still be relatively small if the defendant spends a lot on convincing the court that the inherent quality of the first issue is not so high, and/or that the inherent quality of the second issue is high. One may assume that the anticipated division of the loss between the parties is a function of the relative probabilities of victory (which themselves depend on the exogenous merits and the expenditures of the parties):

$$\sigma = f\frac{P_d}{P_d + P_p} = f\frac{1}{1 + \frac{P_p}{P_d}}$$

Note that if we set f equal to zero, then $\sigma = 0$ and we get a rule in which the plaintiff needs to win on both issues to obtain a judgment. In other words, we can consider the case of the first type of complex legal rules as a special case of the second type of complex legal rules , which will make the mathematics easier: we can look for the

equilibrium expenditures for the second type of complex legal rules, and derive the equilibrium expenditures for the first type by setting f equal to zero.

Turning from the plaintiff's expected value to the defendant's expected loss, we obtain in a similar way:

$$EL_{def} = P_d(1 - P_p) + P_pP_d\sigma + Y_1 + Y_2$$

If we now fill in the formulas (contest functions) for P_p and P_d in the plaintiff's expected value and the defendant's expected loss, we obtain the following Nash equilibria:

$$X_1^* = Y_1^* = F_1(1-F_1)(1-F_2) + F_1(1-F_1)F_2 f\frac{F_1}{F_1+F_2}\left[2 - \frac{F_1}{F_1+F_2}\right]$$

$$X_2^* = Y_2^* = F_1F_2(1-F_2)(1 - f\frac{F_1}{F_1+F_2}(1 - \frac{F_2}{F_1+F_2}))$$

4. Empirical issues

4.1. Empirical research regarding the plausibility of endogenous litigation models

Existing empirical research sheds some light on the plausibility of the theories on endogenous litigation expenditures. The following empirical studies confirm several results of these theories:

1. A first study concerns "strategic reciprocality" (Katz, 1988). To describe this phenomenon, it is necessary to introduce some definitions. A party's effort is called *provocative* when a marginal increase in it leads to an increase in her opponent's effort. A party's effort is called *deterring* when a marginal increase in it leads to a decrease in her opponent's effort. The *favourite* is the party with a greater than 50 percent probability of winning and the *underdog* is the party with a less than 50 percent probability of winning. Strategic reciprocality means that at the (Nash) equilibrium one party's effort is provocative and one party's effort is deterring. The models using the ratio form lead to such a result: the effort of the favoured party is deterring and the effort of the underdog is provocative. In other words: the

underdog will spend less when the favoured party increases his effort and the favoured party will spend more when the underdog increases his effort. This can be explained intuitively. When the favoured party increases his effort, the case becomes (even) less close. It becomes less worthwhile for the underdog to spend more. When the underdog increases his effort, the case becomes a closer one. It becomes more worthwhile for the favourite to spend more. Sheperd (1999) has made an empirical study of the economics of pre-trial discovery.[11, 12] His findings are in accordance with strategic reciprocality. The author examined 369 U.S. federal cases in which the attorneys of *both* sides provided information about the number of days they devoted to seeking discovery.[13] Further information was available through court docket sheets and written questionnaires. Sheperd finds that plaintiffs and defendants behave very differently when conducting discovery. When the plaintiff engages in excessive discovery,[14] the defendant retaliates. The plaintiff on the other hand retreats when the defendant engages in excessive discovery.[15, 16]

2. A second result that has been empirically validated concerns the influence of an increase in the stakes on the expenditures of the parties. In theoretical models that use the ratio form, an increase in the stakes induces both parties to increase their effort since the marginal value of effort rises. Clearly, $aF(1\text{-}F)J$ increases with J. Empirical studies have indeed found a strong correlation between the stakes and the efforts of litigants. Kakalik *et al.* (1998) find that "higher stakes are associated with significantly higher total lawyer work hours, significantly higher lawyer work hours on discovery, and significantly longer time to disposition". More spe-

11 Sheperd uses data from a survey that researchers at Columbia University conducted in 1962 and 1963 to assess the consequences of the discovery provisions in the Federal Rules of Civil Procedure. The author argues that the data are still relevant despite their age (over 40 years). This is because the discovery rules have not changed substantially since the survey. Although approximately one-third of the states have introduced a requirement of automatic disclosure, discovery not within the scope of automatic disclosure is still subject to the "old" rules. Moreover, about two-thirds of states still have systems that do not require automatic disclosure.

12 During pre-trial discovery, the parties can force each other to disclose documents and other evidence relevant for the case.

13 With exclusion of the days the litigant devoted to responding to the requests of the other party.

14 Excessive according to the defendant.

15 Excessive according to the plaintiff.

16 Note that Sheperd's results imply that discovery may produce efficiency and justice since the plaintiff bases his discovery amount on the fundamentals of the case. At the same time, the fact that the defendant behaves strategically without looking at the fundamentals of the case directly leads to social waste and injustice.

cifically, "median total lawyer work hours were more than two and a half times larger for cases with monetary stakes over $500,000 than for cases with monetary stakes of $500,000 or less, while mean total lawyer work hours were almost four times larger". Willging *et al.* (1998) find that "the size of the monetary stakes in the case had the strongest relationship to total litigation costs of any of the characteristics we studied."

3. Third, in its simplest version ($P(X,Y) = \frac{X}{X+Y}$), the model predicts that each party spends one-quarter of the amount at stake in litigation expenditures. Also in the more complex case of $P(X,Y) = \frac{FX}{FX+(1-F)Y}$, each party spends approximately one-quarter of the amount at stake if the case is a close one (F≈1/2). In an experimental study, Eastman and Viswanath (2003) analyze the litigation expenditure decisions of parties in a setting which allows litigants to believe that they can outsmart (or be outsmarted by) the other litigant in making spending decisions. Litigants were told that their probability of victory was equal to the ratio of their effective litigation spending to that of the other side. The instructions explained that effective litigation spending could be equal to actual dollars spent, but could also be higher or lower than actual spending. The litigants were told that the effectiveness of their spending would be determined based on an outside evaluation. The case involved a car accident with stipulated damages of $100,000. Litigants were given the option of spending on several categories (e.g. legal research, hiring a detective etc.). They were not given the opportunity to settle the case. The authors find that the respondents spent an average of $30,000, which is relatively close to and statistically not different from the prediction of rational choice theory ($25,000).
4. Fourth, the model predicts that the parties will spend the same amount (no matter how low or high the quality of the case). This finds some confirmation in the same experimental study of Eastman and Viswanath (2003). In most cases, the parties spent a similar amount.

This is not to say that there are no criticisms of the ratio form:

5. Some predictions of the ratio form have not been confirmed. For example, the model predicts that the expenditures of litigants increase *linearly* with the amount at stake. However, empirical research by Kakalik *et al*. (1984) found that the expenditures of the parties do not rise in proportion to the amount awarded, but rather concavely. This cannot be seen as a fundamental flaw of the ratio form. If we were to introduce fixed costs on top of variable costs in the basic model, then the expenditures would no longer rise linearly with the amount at stake.
6. We have seen that the ratio form can be derived in several ways. However, other functional forms can also be derived in different ways: the logit function and the probit function. These three functions differ in their assumptions on the distributions of the error terms. In the Tullock model the error terms have an inverse exponential distribution, in the logit model they are extreme value distributed and in the probit model they are normally distributed. Empirical tests have not yet been successful in showing which model best captures the characteristics of a particular contest and gives the most accurate predictions (Jia, Skaperdas and Vaidya, 2013).
7. Finally, some other criticisms do not explicitly concern the type of function that is used for the plaintiff's probability of victory, but rather the solution concept that is employed. Most articles rely on the Nash-Cournot protocol: the parties choose their expenditures in ignorance of the opponent's simultaneous choice. However, social interactions are usually too complex to be captured by any simple model. In some contexts, other solution concepts may be more satisfactory. For example, under the Stackelberg solution concept, one side commits to a level of effort, to which the other side then makes an optimizing response. Using this solution concept, Hirshleifer and Osborne (2001) find some interesting differences with the model that relies on the Nash-Cournot protocol: the litigation efforts of the parties are almost always unequal and the side with the better case tends to fight harder. Outcomes tilt disproportionally in favour of the side with the more meritorious case.

In conclusion, specific litigation success functions can be helpful in analyzing the consequences of legal rules. The function most intensively used in the literature, the ratio form, has strong theoretical underpinnings and leads to results which have been empirically validated. However, more theoretical and empirical research is necessary to determine whether other functional forms better capture the expenditure decisions for some types of disputes.

4.2. Empirical research regarding the influence of litigation costs on earlier stages

We have seen that whether increased trial costs increase the settlement frequency depends on the theoretical models one uses.

A study by Stanley and Coursey (1990) let subjects bargain over the division of 100 valuable tokens. If the subjects were not able to settle on a division during the allotted negotiation period, all the tokens were given to one subject or the other, which depended on whether an urn with 100 red and white chips held more white or more red chips. If the parties did not settle, costs were imposed on both parties. These costs were varied in the experiments. Before the start of the negotiations, each subject privately sampled a certain number of chips from the urn. The study could not confirm that settlement rates increase with legal costs. Stanley and Coursey discussed some possible explanations for this result. One explanation is the possibility that increasing the settlement range, which would increase the scope for settlement, also increases the range of indeterminacy and the scope for bargaining.

Fournier and Zuehlke (1989), using data from a survey of U.S. federal civil filings during the fiscal years 1979-1981, find a negative influence of the litigation costs on the probability of settlement. However, a substantial limitation of their test is that they had only information on the estimated public costs of the case. This was then used as a proxy for the private costs of litigation.

Gross and Syverud (1991) study 529 California civil jury trial cases in different areas of the law. They observe settlement offers and judgments for cases that go to trial. They find that when plaintiffs pay their own litigation costs, the settlement rate increases. We need to stress however that selection effects may complicate things substantially. When trial costs decrease, the plaintiff will be willing to file some relatively weak suits that he or she would not have filed if the trial costs were higher. In other words, the population of filed cases may be quite different when costs are low than when costs are high. Gross and Syverud did indeed find that when plaintiffs pay their own litigation costs, the plaintiff win rate at trial increases, indicating that on average higher quality suits are brought.

Eisenberg and Farber (1997) showed that an increase in the *variance* of potential plaintiff's trial costs leads to a decrease in the settlement rate. They test this result using data from 200,000 civil suits in U.S. federal district courts filed between 1986 and 1994. They find that individuals, who have a more variable distribution of litigation costs than corporations,[17] settle less often as plaintiffs. Theoretically, the reason is that, as Eisenberg and Farber argue, for a fairly general class of distributions of litigation costs, increasing the dispersion of litigation costs among potential plaintiffs results in lower average litigation costs among plaintiffs who actually file lawsuits.

4.3. Empirical research about the effect of optimism on the litigation expenditures

Although it is extremely difficult to measure optimistic beliefs of the parties, optimism about the amount at stake is similar to one party having greater stakes than the other. Gross and Syverud (1991) study 529 California civil jury trial cases in different areas of the law. They observe settlement offers and judgments for cases that go to trial. They find that asymmetric stakes cause the settlement rate to decrease and the win-rate for the high-stakes player to increase. This could mean that high stakes litigants try to establish a tough reputation by going to trial more frequently, and more importantly in the context of expenditures, that they exert more effort to win cases at trial.

17 Eisenberg and Farber (1997, p. 93): "Individuals are relatively free to "indulge" their tastes for litigation either by refraining from filing suit where they have a strong case or by filing suit where they do not have a strong case. In contrast, corporations are relatively constrained by market forces and adhere more closely to the goal of profit maximization. Corporations can afford neither to avoid lawsuits expected to be profitable nor to pursue lawsuits not expected to be profitable."

B. UNCERTAINTY, AMBIGUITY AND INACCURACY

1. Description of uncertainty, ambiguity and inaccuracy

Legal systems are often accused of performing poorly because of the haphazard way in which courts decide contractual and extra-contractual liability issues and determine the amount of damages. Civil litigation regimes are perceived as far too random. Awards are said to be unpredictable and often unrelated to a reasonable notion of fault. Some have even argued that civil litigation can best be seen as a lottery system (Harris, 1992).

It is important to define the terms uncertainty, ambiguity and inaccuracy. In the literature, they are sometimes used interchangeably, which can create a lot of confusion. We start with inaccuracy (or error). Shavell stresses the need to define error in an economic way: "Errors may be factual in nature, and they may also occur in the determination of the applicable legal rule or in its use in combination with the found facts. The latter types of error are conceptually clear when the law is well articulated, but often the law will not be specified in a relevant aspect and must be amplified. If the reader takes the view that there exists a correct definition of the social good, then error consists in failing to extend the law in accord with this definition. Otherwise, error is not a well-defined concept" (Shavell, 1995, p. 53). A distinction can be made between different types of errors. Distinguishing between different types is important because they affect the behaviour of the parties in different ways. A first distinction is between Type I and Type II errors. The court makes a Type I error when it finds a violation where conformity occurred (the innocent are convicted). The court makes a Type II error when it finds conformity where a violation occurred (the guilty are acquitted). Another distinction can be made between errors concerning the law and errors concerning the facts. This distinction is important because errors concerning the law may have larger negative effects since they (generally) affect third parties to a larger extent than errors concerning the facts. Next, errors concerning whether a violation occurred are different from errors about the amount of damages. Errors about the amount of damages do not always affect the incentives of potential injurers negatively.[18] A final distinction leads us to the concept of uncertainty: while biases are errors that can be perfectly predicted in individual cases (e.g. the actual harm is 100, but judges are known to award only 50), uncertainty refers to errors that cannot be anticipated but their probability distribution is known and their average generally corresponds to the true value (e.g. the actual harm equals 100, and half of judges award 50 and the other

18 For example, potential injurers will not take too much care when courts grant too much damage under a negligence rule. The potential injurer can simply avoid liability by taking due care.

half 150) (Dari-Mattiacci and Deffains, 2007). Ambiguity is different from uncertainty in the sense that under ambiguity, the decision maker is not able to come up with a probability distribution based on available information. A decision-maker may simply not have sufficient observations of the realization of a random variable at stake to be able to apply the law of large numbers and thus use the empirical frequency as a trustworthy estimate of the true probability distribution. This occurs especially when the event is non-repeatable by nature. Ambiguity of the law can be linked to the inherent ambiguity of language itself, the use of vague notions (e.g. the reasonable man standard), and a natural process of obsolescence because of continual changes in society and technology.

2. Consequences of uncertainty, ambiguity and error

2.1. The effect on *ex ante* behaviour

2.1.1. General

The consequences of error and uncertainty on *ex ante* incentives have been examined quite extensively in the law and economics literature (Calfee and Craswell 1984; Craswell and Calfee 1986; Kahan 1989, Kaplow and Shavell 1994; Kaplow and Shavell 1996). Dari-Mattiacci (2005) systematizes the various theories and provides an integrated formal analysis, the results of which we summarize here.

Dari-Mattiacci distinguishes between biases and uncertainty. As discussed before, biases are errors that can be perfectly predicted in individual cases (e.g. the actual harm is 100, but judges are known to award only 50). Uncertainty refers to errors that cannot be anticipated but their probability distribution is known and their average generally corresponds to the true value (e.g. the actual harm equals 100, and half of judges award 50 and the other half 150). Furthermore, the author distinguishes between situations in which adjudicators rely on *rules* in which regulators define due care (and damages are separately awarded following adjudication) and situations in which adjudicators rely on *standards* in which the law defines the due level of care only in general terms (often by using vague criteria like the "reasonable man"). This distinction is important because with rules an error in due care does not necessarily imply an erroneous evaluation of the harm (and vice versa), while with standards, errors concerning damages may cause errors in the setting of due care (at least when judges determining due care in individual cases weigh the cost of care against the

reduced accident loss). Finally, the author makes a distinction between two types of negligence rules: one with full damages and one with incremental damages. In the latter case, due to the interaction of the causation rule, negligent injurers do not pay for damages that would have occurred even if they had been non-negligent. This difference causes the two rules to react in a different manner to errors (Calfee and Craswell, 1986; Kahan, 1989). The reason is that negligence with full damages creates a discontinuous cost function for the injurer given that he pays full damages if negligent and no damages otherwise. With the causation rule incorporated, the injurer's costs rise continuously when they reduce their precaution level below due care.

With respect to biases under rules, Dari-Mattiacci establishes the following results:

1. Under strict liability, over-compensation yields over-precaution, and under-compensation yields under-precaution.
2. Under negligence with full damages, if due care is optimal, over-compensation and moderate under-compensation produce no effect, while serious under-compensation yields under-precaution.
3. Under negligence with incremental damages, if due care is optimal, then over-compensation produces no effect, while under-compensation yields under-precaution.
4. Under negligence with full damages, if damages are equal to the harm, the injurer takes due care if it is lower or moderately higher than optimal care. If due care is seriously higher than optimal care, then the injurer takes optimal care.
5. Under negligence with incremental damages, if damages are equal to the harm, the injurer takes due care if it is lower than optimal care. If due care is higher than optimal care, the injurer always takes optimal care.
6. Under negligence with full damages, the effects of under-compensation can be mitigated by lowering due care below the optimal level. Under negligence with incremental damages, the effects of under-compensation cannot be corrected.
7. Under negligence with full damages, if due care is higher than optimal care, the injurer can be induced to take a lower level of care by lowering the damage award. If due care is lower than optimal care, then the injurer cannot be induced to take a higher level of care. Under negligence with incremental damages, the effects of higher-than-optimal and lower-than-optimal due care cannot be corrected.

With respect to biases under standards, the results are as follows:

1. Under negligence with full damages, if damages and due care are set jointly, the injurer always takes due care, which is equal to the level of precaution that the injurer takes under strict liability given the same damage award.
2. Under negligence with incremental damages, if damages and due care are set jointly, the injurer always takes due care, which is equal to the level of precaution that the injurer takes under strict liability given the same damage award.

Regarding uncertainty under rules, the following results are presented:

1. Under all rules, with efficient due care levels, uncertainty concerning the determination of damages does not affect the injurer's precaution.
2. Under negligence with full damages, if damages are equal to the harm and there is uncertainty concerning due care, the injurer may take under- or over-precaution.
3. Under negligence with incremental damages, if damages are equal to the harm and there is uncertainty concerning due care, the injurer takes under-precaution.
4. Under negligence with full damages and under negligence with incremental damages, the effects of uncertainty concerning due care can be completely offset by adjusting the damage award.

With respect to uncertainty under standards, the author finds that:

1. Under negligence with full damages, if damages and due care are set jointly and there is uncertainty, the injurer may take under- or over-precaution. If rules yield over-precaution, standards will lead to even higher levels of precaution. If rules yield under-precaution, the level of precaution under standards may be higher or lower than under rules.
2. Under negligence with incremental damages, if damages and due care are set jointly and there is uncertainty, the injurer takes under-precaution. Such a level is higher than the level of precaution taken under rules.

2.1.2. The filtering effect of uncertainty

De Mot, Miceli and Depoorter (2016) unveil a hitherto neglected benefit of uncertainty in adjudication. They show that uncertainty in adjudication has a filtering effect

on the frequency and types of accidents that occur in society. When accounting for imperfect decision making by courts, uncertainty regarding the correct application of liability standards has a beneficial filtering effect on harmful activities: it prevents the more harmful accidents while leaving less harmful accidents unaffected. To illustrate this filtering effect, consider an activity A that imposes harm of 100. The harm can be completely avoided with an investment in precautions of 60. Assume a setting where 25 percent of judges ("lenient judges") make errors and incorrectly excuse injurers as soon as precaution costs exceed half of the harm (50). The other 75 percent of judges ("efficient judges") do not err and forgive injurers only when the precaution costs exceed the accident costs. If potential injurers know ahead of time whether the case will be assigned to an efficient or lenient judge, 25 percent of potential injurers know that they will not be held accountable. These potential injurers will neglect to invest 60 in precautions to prevent the accident. The resulting social losses are 10 (0.25x40=10). Now compare this with a setting where there is more uncertainty about the legal errors being made. If there is uncertainty about judicial error, and potential injurers do not know *ex ante* whether their case will be handled by the more lenient judges, all potential injurers now face a 75 percent probability that they will be held liable if they do not invest 60 in precautions and a 25 percent probability that they will be excused even if they do not take precautions. Potential injurers will take care since the precaution costs are lower than the expected liability costs (60<75).

Interestingly, this beneficial filtering effect of uncertainty disproportionally impacts those accidents that are socially most wasteful. To illustrate, compare the effect of uncertainty on activity A with that on activity B where the precaution costs are 90 instead of 60 (activity A). Although the accident costs (100) exceed the precaution costs for both activities (60 and 90, respectively for A and B), it is more valuable socially to prevent activity A since harm can be prevented at a lower cost for activity A (60) than for activity B (90), resulting in social gains of 40 instead of 10. First, under conditions of certainty, 25 percent of potential injurers again know that they will not be held accountable if they fail to invest 90 in precautions to prevent the accident. The resulting social losses are 10 (0.25x10=2.5). Second, by contrast, if there is uncertainty about judicial error, activity B injurers face a 75 percent probability that they will be held liable if they do not invest 90 in precautions and a 25 percent probability that they will be excused even if they do not take precautions. Here, as opposed to activity A, uncertainty will not induce potential injurers to take care. The precaution costs

exceed the expected liability costs (90>.75x100+.25x0). Consequently, the resulting social losses (10) occur for activity B regardless of uncertainty.

Far from being abstract and exogenously determined, the level of uncertainty is influenced by various concrete institutional decisions and procedural rules. The authors show that their filtering insight helps inform discussions on the optimal strength of legal precedents. Legal systems generally differ with regard to the strength of precedent. In a strong precedent system, such as that employed by the U.S. Supreme Court, judges are legally bound by prior judicial decisions. In a weak system of precedent, such as those found in many civil law jurisdictions, prior decisions provide valuable guidance but are not formally binding upon subsequent decision makers.[19] How strong should the authority of precedent be on subsequent decision makers? Generally, strong precedent systems are heralded for providing legal certainty and stability. By constraining judges who might otherwise overrule preceding law, a strict precedent system increases judicial predictability, enhances economic efficiency, and enables a higher amount of settlements (Landes and Posner, 1976; Macey, 1989, 1998; Spier, 2005; Shapiro, 1972). A potential downside of a strong precedent system, however, is that an erroneous precedent may control subsequent judicial decisions or become inefficient over time due to changing circumstances.[20] For litigants, the strength of precedent impacts the level of uncertainty in adjudication. A weak precedent system allows for greater levels of uncertainty. Several contradictory rules may emerge from different cases, and potential injurers will select their care level by taking this mixed case law, with efficient and inefficient precedents standing side by side, into account. In a strong system of precedent, by contrast, the existence of an established precedent reduces (but of course does not fully eliminate) the level of uncer-

19 Although the strictness of precedents is often noted as a major procedural difference between civil and common law jurisdictions, most legal systems are mixed systems of precedent. Theoretically, a line is drawn between the doctrine of *stare decisis*, which prescribes that courts adhere to past legal precedent on issues of law when deciding pending cases and the doctrine of *jurisprudence constante*, which holds that judges are bound only when a consolidated trend of decisions has already been established. Judicial decisions do not become a source of law until they mature into a prevailing line of precedents (Dainow, 1974; Dennis, 1993).

20 A major drawback of a system based on precedent is the danger that an erroneous precedent will control subsequent judicial decisions (Kornhauser, 1989). Precedents that were once efficient may become inefficient ("depreciate") over time, for instance due to changing economic and social conditions (Posner, 1992). This cautions against stronger precedent regimes. On this account, economic analysis recommends a certain degree of leniency: judicial decision makers should ignore existing precedent when the value of creating a new rule outweighs the loss in predictability that is caused by overturning the old rule (Macey, 1998, p. 71).

tainty as to the case outcome.[21] Once a precedent is set, potential injurers select their level of care on the basis of this precedent – regardless of whether it is efficient or inefficient. Consider for instance a setting where two alternative legal doctrines might apply to the behaviour of a given type of agent with respect to her principal. Each doctrine imposes a different duty of care, e.g. gross negligence and ordinary negligence.[22] Suppose that the ordinary negligence duty is efficient. Under a strong precedent system, either the efficient doctrine (ordinary negligence) will become precedent or the inefficient one (gross negligence). From there on, the agent is certain as to the required duty of care. Under a weak precedent system, both doctrines may co-exist. The agent will be uncertain as to the doctrine that will be applied to her conduct in a given situation, and hence as to the duty of care to which she will be held.

Applying the analysis of De Mot, Miceli and Depoorter (2016), if judicial decision making is imperfect, the uncertainty imposed by a weak system of precedent benefits from the filtering effect described above. As a result, a weak precedent system is more likely to inhibit the filtering effect of preventing the most socially harmful accidents but without exercising any deterrent effect on less harmful accidents. Strong precedent systems are less likely to distinguish between various types of accidents. Overall then, a weak precedent system has an efficiency advantage over stronger forms of precedent when the filtering advantage (higher deterrence of very harmful behaviour) is higher than the filtering disadvantage (lower deterrence of less harmful behaviour).

2.2. The influence on risk costs

When one or both litigants are risk-averse, the fact that litigation is fraught with uncertainty creates risk costs. Risk-averse individuals generally prefer a certain, safe amount to an uncertain gamble that offers the same expected value. In other words, a risk-averse plaintiff does not value a trial at its expected value $E(x)$, but discounts the expected value of trial taking into account the accompanying risk. Technically, if $u(x)$ is the utility function of the plaintiff, risk averseness means that $E(u(x)) < u(E(x))$. From the point of view of the plaintiff, the expected value needs to be discounted with a risk premium Rp. This risk premium is the additional amount that he or she would need to receive to make him or her indifferent between an uncertain adjudication with the average result $E(x)$ and a certain sum $E(x)$. The plaintiff is indifferent between set-

21 A strict precedent system reduces the range of uncertainty by ascertaining the applicable determining rule, but of course does not affect the remaining uncertainty with regard to the application of the particular rule to the specific facts of the dispute.

22 Of course, we could add other duties of care, e.g. a high trustee standard.

tling for E(x) - Rp and the risky trial. Likewise, if the defendant is risk-averse, he or she will be indifferent between a risky adjudication and settling for $E(x) + Rd$, with Rd being the defendant's risk premium. Thus, the direct effect of introducing risk is a total cost of $Rp + Rd$.

2.3. The influence on the settlement frequency

Uncertainty can influence the settlement frequency in two, diametrically opposed ways. Due to risk costs (see above), the settlement range may increase and this may stimulate the settlement frequency. However, uncertainty may also make it more difficult (more costly) to find out one's true probability of victory (or the damages the court will grant in the case of victory), and this can drive a wedge between the viewpoints of the parties (divergent expectations).

2.4. The influence on the settlement duration

For the relationship between uncertainty and the duration of settlement, we would refer to the section on settlement duration.

2.5. The influence on the settlement outcome

Uncertainty can influence the settlement outcome in different ways. Due to risk costs, the plaintiff may be willing to settle for less, and the defendant may be willing to settle for more. If the bargaining power of the plaintiff is relatively modest, the settlement amount can be expected to drop. When uncertainty increases the plaintiff's optimism, this can increase the settlement amount, because the reservation value of the plaintiff increases; when it increases her pessimism, this can decrease the settlement amount, since the reservation value of the plaintiff decreases.

2.6. The influence on filing suit

Theoretically, uncertainty affects the incentive to sue in different ways. Due to the increase in risk costs, the incentive to sue may decrease. Also, if the settlement frequency and settlement amounts go down, the incentive to file will decrease further.

2.7. The special case of ambiguity

Only very recently has the law and economics literature started to focus on the effects of ambiguity on the parties' incentives to settle and the injurer's incentive to take care. Chappe and Giraud (2013) construct a model featuring ambiguity and optimism/pessimism. Ambiguity corresponds to an agent's lack of confidence in his belief about the probability of uncertain events (e.g. the probability of prevailing). Optimism and pessimism corresponds to an agent overweighting the best and the worst outcome respectively (e.g. to receive damages or not). The authors develop a model encompassing two sources of settlement failure: asymmetric information about damages and divergent expectations (because of ambiguity) about the probability of a plaintiff victory.

Chappe and Giraud find the following results. First, given a settlement offer from the defendant, (a) the probability of settlement decreases with the plaintiff's degree of optimism and (b) decreases with the plaintiff's perceived degree of ambiguity if and only if the plaintiff is optimistic (or equivalently, it increases with the plaintiff's level of confidence if and only if the plaintiff is optimistic). With respect to (a), intuitively, the more optimistic the victim is, the more likely he is to reject the offer, since he thinks he will win in court. With respect to the intuition of (b), an optimistic plaintiff is an ambiguity-loving plaintiff, and more ambiguity means that a high probability of victory is more likely for such a plaintiff. A pessimistic plaintiff is an ambiguity-averse individual, so a low probability of victory is more likely for such an individual. Second, the defendant's settlement offer (c) increases with the plaintiff's level of optimism and (d) increases with the plaintiff's level of confidence if and only if the plaintiff is pessimistic. Intuitively, when the plaintiff's subjective probability of victory increases, his incentive to settle is weaker and the defendant will raise his offer.[23] Finally, the level of care the defendant adopts (e) increases with the plaintiff's level of optimism and (f) increases with the plaintiff's level of confidence if and only if the plaintiff is pessimistic.

23 Note that the authors show that all these results (a,b,c and d) only hold if the elasticity of the marginal plaintiff in equilibrium to a rise in the settlement offer is larger than one (in other words, plaintiffs are highly sensitive to rises in the settlement offer).

3. Fundamental issues related to theoretical modelling

3.1. Important existing models

3.1.1. Bias, uncertainty and deterrence

The unitary framework developed by Dari-Mattiacci (2005) discussed above considers accidents occurring between two risk-neutral parties, a victim and an injurer. Only the injurer can take precautionary measures to reduce the probability of an accident. The following symbols are used:

x = the injurer's level of care, $x \geq 0$;
z = the due level of care, $z \geq 0$;
$p(x)$ = the probability of an accident, $0<p<1$, $p'<0$, $p''>0$;
H = victim's harm, $H>0$;
D = the damage award, $D>0$.

The social cost of an accident is equal to the sum of the victim's expected harm and the injurer's cost of care. The socially optimal level of precaution, x^*, is the level that minimizes this sum:

$$\min_x \left[p(x)H + x \right]$$

Under strict liability, when damages are biased, the injurer's goal is to:

$$\min_x \left[p(x)D + x \right]$$

Under rules, when negligence with full damages is used, if due care is optimal but damages are biased, the injurer's minimization problem is:

$$\min_x \begin{cases} p(x)D + x & \text{if } x < x^* \\ x & \text{if } x \geq x^* \end{cases}$$

Under rules, when negligence with incremental damages is used, if due care is optimal but damages are biased, the injurer's minimization problem becomes:

$$\min_x \begin{cases} (p(x) - p(x^*))D + x & \text{if } x < x^* \\ x & \text{if } x \geq x^* \end{cases}$$

Under rules and negligence with full damages, if damages are equal to the harm, but due care is biased, the injurer's problem is:

$$\min_x \begin{cases} p(x)H + x & \text{if } x < x^* \\ x & \text{if } x \geq x^* \end{cases}$$

Under rules and negligence with incremental damages, if damages are equal to the harm, but due care is biased, the injurer's problem is:

$$\min_x \begin{cases} (p(x) - p(z))H + x & \text{if } x < x^* \\ x & \text{if } x \geq x^* \end{cases}$$

Under standards and negligence with full damages, if damages are biased, due care will also be biased, and the injurer minimizes:

$$\min_x \begin{cases} p(x)D + x & \text{if } x < x^* \\ x & \text{if } x \geq x^* \end{cases}$$

Under standards and negligence with incremental damages, if damages are biased, due care will also be biased, and the injurer minimizes:

$$\min_x \begin{cases} (p(x) - p(z))D + x & \text{if } x < x^* \\ x & \text{if } x \geq x^* \end{cases}$$

The case of uncertainty is similar, but some additional notations need to be introduced. For a given z, D varies between 0 and ∞ according to a distribution function g(D) with g(D)>0 over the relevant region and g(D)=0 elsewhere, and a cumulative distribution G(D). For a given D, z varies between 0 and ∞ according to a distribution function f(z) and a cumulative distribution F(z) with the same characteristic as g. The two distributions are independent of each other. To distinguish the effects of uncertainty from those of bias, it is assumed that awards and due care are correct on average: $\int_0^\infty DdG(D) = H$ and $\int_0^\infty zdF(z) = x^*$.

This formalization by Dari Mattiacci can be helpful when a cradle-to-grave model wishes to introduce aspects of bias or uncertainty.

3.1.2. How to model ambiguity?

We have previously described the results of Chappe and Giraud (2014) who study a litigation model under ambiguity. Here we will focus on the more formal aspects of their theory. The question is how to model a situation in which the decision maker does not possess all the information needed to come up with an objective probability distribution. In such a case, one cannot model the behaviour of the decision maker as if he were behaving under uncertainty. The most well-known models which have dealt with this issue are the Multiple Prior Expected Utility Model (Gilboa and Schmeidler, 1989) and the Choquet Expected Utility Model (Schmeidler, 1989). Chappe and Giraud use a variant of the last model: the NEO-additive model (Chateauneuf, Eichberger, and Grant, 2007). The advantage of this model is that it offers a parametric representation of perceived ambiguity (or alternatively, the degree of confidence in one's probabilistic estimation) and of attitude towards ambiguity (ambiguity-loving or ambiguity-aversion). More specifically, the plaintiff's expected recovery equals:

$$V = \alpha\pi L + (1-\alpha)\gamma L, \alpha \in (0,1), \gamma \in (0,1)$$

with:
α : a parameter indicating the plaintiff's confidence about his probability of victory (so $1\text{-}\alpha$ represents the degree of ambiguity);
π : the plaintiff's probability of victory;
L : the damages awarded in court in the case of a plaintiff victory; the plaintiff knows L, the defendant only knows the distribution of L;
γ : a parameter representing the plaintiff's optimism.

Note that if the decision maker is fully confident in the prior he uses ($\alpha=1$), then he would behave as an expected utility maximizer.

The objective probability of winning, π, can now be adjusted – taking into account the confidence and optimism parameters – into a subjective probability of winning, $\pi°$:

$$\pi° = \alpha\pi + (1\text{-}\alpha)\gamma$$

The difference between π° and π shows whether the plaintiff is optimistic or pessimistic. The plaintiff is optimistic if $\pi^\circ > \pi$, which is equivalent to $\gamma > \pi$. The plaintiff is pessimistic if $\pi^\circ < \pi$, which is equivalent to $\gamma < \pi$. Optimistic plaintiffs can thus be considered as ambiguity-loving, while pessimistic plaintiffs can be viewed as ambiguity-averse.

One can easily see that the subjective probability increases with the level of optimism: $\frac{\partial \pi^\circ}{\partial \gamma} = 1 - \alpha > 0$. The influence of a change in ambiguity is not as straightforward:

$\frac{\partial \pi^\circ}{\partial \alpha} = \pi - \gamma > 0$ *if and only if* $\pi > \gamma$ Consequently, when ambiguity increases, an

optimistic plaintiff becomes more optimistic and a pessimistic plaintiff becomes more pessimistic.

Chappe and Giraud (2014) incorporate this ambiguity in an asymmetric information model. The plaintiff has private information about the loss, and the uninformed defendant makes a take-it-or-leave-it settlement offer. The offer "screens" the plaintiffs in two groups. The major difference with the classic screening model by Bebchuk (1984) is that the plaintiff and the defendant consider a different probability of a plaintiff victory (π° and π respectively).

3.2. Possible extensions

3.2.1. How to vary uncertainty and accuracy in models with endogenous litigation expenditures?

We have seen above that many papers which introduce endogenous litigation expenditures model the plaintiff's probability of victory as follows (we use one of the simplest forms here):

$P(X,Y) = \frac{XF}{XF + Y(1-F)}$, with

F = index of the factual fault of the defendant, normalized so that $F \in [0,1]$;
X = plaintiff's litigation effort;
Y = defendant's litigation effort.

Obviously, the trial outcome in these models is always uncertain, no matter how much the parties spend. An important question is how the uncertainty of the outcome can be varied. To answer this, we need to look at how this formula can be derived.

We can use Bayes' rule to compute the conditional probability that the defendant was negligent, denoted P_1, as follows:

$$\frac{P_0XF}{P_0XF + (1 - P_0)Y(1 - F)}$$

where

P_0 = court's prior probability that the defendant is negligent, $P_0 \in [0,1]$.

In this formulation, P_0 can be interpreted as a measure of the court's "bias". Specifically, if $P_0 = 1/2$ the court is unbiased, whereas if $P_0 > 1/2$, it can be said to have a pro-plaintiff bias, and if $P_0 < 1/2$, a pro-defendant bias. Alternatively, one could interpret $P0$ as reflecting noise in the court's assessment of the evidence presented at trial. Under such an interpretation, if $P_0 = 1/2$, the court attaches equal weight to each party's evidence, whereas if $P_0 > 1/2$, the court gives more weight to the evidence produced by the plaintiff, and if $P_0 < 1/2$, it gives more weight to the defendant's evidence. This could, for example, be the case when judges think that on average, evidence produced by the plaintiff or defendant respectively is more reliable. Finally, XF can be interpreted as the "evidence against" the defendant, which depends positively on both the factual evidence against him and the plaintiff's litigation effort. Likewise, $Y(1–F)$ is the "evidence for" the defendant. The probability of plaintiff victory with a preponderance-of-the-evidence standard is given by:

$$P(X,Y) = \text{prob}\,(P_1 > 1/2) = \text{prob}\,(P_0 > \frac{Y(1-F)}{Y(1-F)+XF})$$

For a uniform distribution of the court's prior,[24] this becomes:

$$P(X,Y) = 1 - \frac{Y(1-F)}{Y(1-F)+XF} = \frac{XF}{XF+Y(1-F)}$$

24 Note that if there were no randomness in the court's behaviour – that is, if P_0 were known with certainty – then the only equilibrium would involve mixed strategies because the outcome of a trial would be determinate.

However, by varying the average value and the variance of the court's prior, we can effectively affect accuracy and uncertainty.

3.2.2. How to model investments by the parties which may reduce uncertainty?

Uncertainty is not necessarily a static concept. During litigation, uncertainty may be partially resolved after the parties have made some investments in the litigation. In the discussion of stage II, we have discussed such theories and models (cost divisibility and gradual information flow; sequential trials) so we can refer to those discussions.

4. Empirical issues

4.1. Uncertainty/Ambiguity

4.1.1. Measuring uncertainty: existing research

Measuring uncertainty is generally regarded as a daunting task.[25] Some possibilities have been advanced in the literature:

1. There are a lot of experimental studies that use questionnaires to find out how judges would adjudicate simulated cases (Austin and Williams 1977; Clancy *et al.* 1981; Guthrie, Rachlinski, and Wistrich 2007; Kapardis and Farrington 1981; Partridge and Eldrige 1974; Van Koppen and Ten Kate 1984; Wistrich, Guthrie, and Rachlinski 2005). Such studies can measure inconsistency[26] since one need only compare the survey responses of two judges and see how often they diverge. Such experimental studies have often been criticized because they lack authenticity (Anderson, Kling, and Stith 1999; Conley and O'Barr 1988; Sisk, Heise and Morriss 1998; Stith and Cabranes 1998, pp. 109–10) due to the use of highly simplified scenarios in written questionnaires, the fact that decision makers in questionnaires are not exposed to advocacy from both sides, are not required to write opinions justifying their decisions, and do not need to consider the consequences of their judgments on actual parties. Furthermore, surveys are usually conducted on students rather than actual judges.
2. One may undertake observational studies of multi-member courts that decide cases *en banc*, such as many Supreme Courts (Martin and Quinn 2002; Bailey 2007; Alarie and Green 2007). Obviously, there are no problems of authenticity here. Furthermore, the requirement of observability is satisfied in these studies as

25 See, for instance, Dari-Mattiacci and Deffains (2007, p. 648).

26 Note that inconsistency can be linked to both uncertainty and ambiguity, depending on whether litigants can form a probability distribution of the various points of view.

long as each judge takes a public position in every case. However, since such courts decide cases through deliberation, the condition of autonomy is not satisfied. Studies show that deliberation will lead to greater conformity among the votes of judges than would occur under autonomous decision making (Fischman 2011a, 2011b). Observed disagreement will only give an idea of the lower bound for inconsistency. This lower bound may not provide much information, in particular if decisions are announced under a strong norm of consensus. When interim decisions are available however, stronger bounds on inconsistency may be possible to provide if data on interim decisions are available (e.g. conference votes of Supreme Court justices; see Brenner (1980), Epstein, Segal, and Spaeth (2001) and Post (2001)).

3. One could look for divergence of judicial decisions on the same issue or a set of similar issues, because this is an indicator of how difficult it is for the parties to predict the outcome of litigation (Dari-Mattiacci and Deffains, 2007). For example, how many judges hold a shopkeeper liable for not preventing clients from slipping at the entrance to the shop when the weather is bad?
4. Osborne (1999) examines the predictability of court awards by looking at the relationship between court awards and the pre-trial expectations of litigants and their attorneys. If variance in expectations explains much of the variance in awards, then court awards are highly predictable.

The court award, A, can be estimated through its expected value $E(A)=f(X)$, and the award itself is given by $A = f(X)+u$, with X being a set of exogenous variables and u being a random shock distributed according to a known density function $g(uI0,\delta^2)$. X includes the facts of the case, the relevant law, the skills of the attorneys on both sides, and the history of the particular court with respect to sympathy for either side or with respect to procedural issues. Osborne tests the randomness hypothesis by measuring the relationship between court awards and *ex ante* expectations of those awards. These expectations have the advantage that they should capture all relevant information that can be obtained at a reasonable cost. In other words, using expectations as the exogenous variable has the benefit of incorporating all the factors that litigants need to make their decisions when pursuing litigation. Osborne performs an ordinary least-squares regression of award on stakes and on a series of dummy variables for (i) the type of case (e.g. torts, contracts, family, property etc.), (ii) the type of lawyer (sole practitioner or not), (iii) whether the case was heard by an arbitration panel, and (iv) whether the estimate of the stakes was made by the plaintiff or defend-

ant attorney. Osborne finds that stakes are highly significant ($p < 0.0001$) and an adjusted R^2 of 0.41, which leads him to conclude that the variance in expectations explains much of the variance in awards. Interestingly, none of the case-type dummy variables are statistically significant, which can be explained by the fact that information such as case type would be incorporated in the stakes estimation. Because the results may indicate an ability to broadly predict wins and losses rather than the award itself, the author runs the same regression on the same sample but minus the observations in which the plaintiff received nothing. The results are not significantly different (an adjusted R^2 of 0.46). Thus, defendant victories are not so contrary to expectations that they significantly increase the unpredictability of litigation.

5. Fischman (2013) shows how data on the decision rates of individual judges can shed light on the feasible rates of inconsistency, indeterminacy and error. We discuss this important article in detail in the next subsection (on error).
6. In the area of medical malpractice, insurers often hire physicians in the relevant specialties to determine whether the claimants' injuries were in fact caused by medical error. Often, the physicians are asked to rate a claim as "defensible", "indefensible" or "unclear". The fraction of times they rate a claim as unclear can be seen as an (imperfect) measure of legal uncertainty. For example, Taragin *et al.* (1991) found physician care was considered defensible in 62 percent of the cases and indefensible in 25 percent of the cases. In the remaining 13 percent of cases, it was unclear whether physician care was defensible.
7. The Priest-Klein model could offer some scope for measuring legal uncertainty for some categories of legal disputes. In the Priest-Klein model, the parties estimate the degree of defendant fault, or alternatively the merits of the case, (Y') with error. The plaintiff wins if the merits are larger than the decision standard (Y^*), thus if $Y'>Y^*$; otherwise the defendant wins. The plaintiff and the defendant make errors while estimating Y', but these estimates are not biased (on average, they are correct). If the parties settle, their joint settlement costs equal S. If they go to trial, their joint trial costs equal C. If the plaintiff wins in court, he will be awarded J. Given a value for $(C-S)/J$, a certain standard deviation of the prediction error, and a given decision standard, Priest and Klein can predict the trial rate and the plaintiff win rate, based on two further assumptions: (1) potential disputes (Y') are distributed normally, with mean zero and standard deviation 1, and (2) the prediction errors follow a normal distribution.

Priest and Klein and others have then used simulations to obtain numbers for the trial rate and the plaintiff win rate. We demonstrate this with a table provided by Lee and Klerman (2015):[27]

Table 1. Effect of changing Decision Standard on Plaintiff Trial Win Rate

Litigation Costs $\left(\frac{C-S}{l}\right)$	Prediction Errors ($\sigma_{\pi,\delta}$)	Percent Litigated	Decision Standard (Y*)				
			-0.1	-0.5	-0.0	0.5	1.0
0.80	1.5	1.4	78.7	65.6	50.2	34.5	21.4
	0.5	0.7	65.4	57.9	49.9	42.1	34.6
	0.1	0.1	53.9	51.1	50.2	48.2	46.9
0.67	1.5	4.0	79.0	65.8	50.0	34.3	20.9
	0.5	1.9	66.1	58.1	49.8	41.8	33.8
	0.1	0.4	53.8	51.7	49.9	47.9	46.3
0.33	1.5	18.2	80.6	66.6	50.0	33.3	19.4
	0.5	9.2	68.6	59.6	50.0	40.5	31.4
	0.1	2.0	54.6	52.1	50.0	47.9	45.8
Plaintiff trial win rate in population (i.e., if all cases went to trial)			84.1	69.1	50.0	30.9	15.9

Note. For all results, 10 million simulated observations (disputes). The results of additional simulations can be found in Online Appendix F.

This could be applied in reverse. If one knows the value of $(C\text{-}S)/J$ for a certain category of dispute, and the fraction of litigated cases and the plaintiff win rate within these cases, one can deduct the most likely combinations of prediction error and decision standard. For example, if one establishes after surveying a particular type of cases that $(C\text{-}S)/J$ is equal to 0.67, that the rate of litigation is about 2 percent and that the plaintiff win rate is around 42 percent, then one may conclude that the standard deviation of the prediction error is 0.5 and the decision standard is in favour of the defendants ($Y^*=0.5$).

27 Lee, Yoon-Ho Alex and Daniel Klerman, Updating Priest and Klein, working paper, 2015.

Note that this type of empirical research faces several difficulties, such as:

1. One needs to check whether the parties have symmetric stakes for the specific category of disputes;
2. One needs to be able to gather information of the value of $(C\text{-}S)/J$;
3. One needs information of the distribution of disputes (is Y' really normally distributed with mean zero and standard deviation 1 ?).

4.1.2. Influence of uncertainty on the settlement frequency

Fournier and Zuehlke (1989) use data from U.S. federal civil cases filed between 1979 and 1981. They find that a higher variance of trial awards increases the probability that the litigants will reach a settlement. A higher variance of trial awards increases the likelihood that the parties disagree on their predictions of trial awards, which increases the probability of trial. However, it also increases the incentives of risk averse parties to settle. The empirical evidence provided by Fournier and Zuehlke thus suggests that the latter effect outweighs the former effect.

4.2. Measuring inaccuracy

4.2.1. Bad proxies

Some authors have argued that the appeal rate and the reversal rate on appeal can be used as proxies for the accuracy of court verdicts (Dimitrova-Grajzl *et al.*, 2016). However, we would stress that there are several reasons why these are not good proxies. First, whether or not litigants appeal depends not only on whether the lower court made a mistake, but also on other elements such as the level of appeal costs. Second, appeal courts may replace an inaccurate decision with an accurate decision, but the opposite sometimes happens too. Third, cases are filtered in the period between the first instance court handing down its decision and the appeal stage starting. If the lower court made a large mistake, often both parties will recognize this. This facilitates settlement after the first instance stage, because both parties will often agree that the court of appeal will overturn the decision. Settlement will be less likely when the lower court made relatively small mistakes.

4.2.2. Estimating error based on indeterminacy?

Fischman (2013) examines the relationship between judicial voting rates and inconsistency, indeterminacy and error. "Inconsistency" measures how often the outcome of a case will depend on the identity of the judge selected to decide it. Formally, it is the probability that two judges would differ in their disposition of a randomly selected case. "Indeterminacy" stands for the proportion of cases in which the law fails to require a unique result. "Error" denotes the proportion of cases in which the judge's decision conflicts with the results required by law. Fischman shows that it is possible: (a) to derive upper and lower bounds on inconsistency by exploiting information about different judges' rates of reaching positive decisions, and (b) to estimate a minimum rate of error as a function of the proportion of cases that are assumed to be indeterminate. The estimates can be used to construct an "indeterminacy-error curve" that demarcates a boundary between feasible and infeasible combinations of indeterminacy and error rates. Indeterminacy and error cannot however be measured in isolation. We will now go deeper into these results.

a. Looking at disparities among judges' decision rates in randomly assigned cases cannot shed light on the exact magnitude of inconsistency. For example, think of two judges who both decide for the plaintiff in exactly 50 percent of cases. If these judges side with the plaintiff in the same 50 percent of cases, their decisions would be perfectly consistent. However, if the second judge decides every case contrary to how the first judge decides, their decisions may be perfectly inconsistent, despite the identical decision rate. The fact that there is no disparity between the decision rates does not allow us to distinguish between the two extremes. Note further that the judges could always be correct, always incorrect, or anywhere in between.

   However, it is possible to derive bounds on inconsistency relying on the information about the judges' rates of reaching positive decisions. A simple example can provide the basic intuition. There are two judges, A and B. We (only) know that judge A reaches a positive decision in 30 percent of cases, and judge B in 60 percent of cases. What can be inferred from this about their rate of disagreement? First, we can say something about the lower bound of disagreement. They must disagree at least 60 percent - 30 percent = 30 percent of the time. This would occur if all of A's positive votes coincided with judge B's positive votes. We can also say something about the upper bound of disagreement. This would occur if each positive vote of one judge coincided with a negative vote of the other judge.

In our example, this means an upper bound of 30 percent + 60 percent = 90 percent.

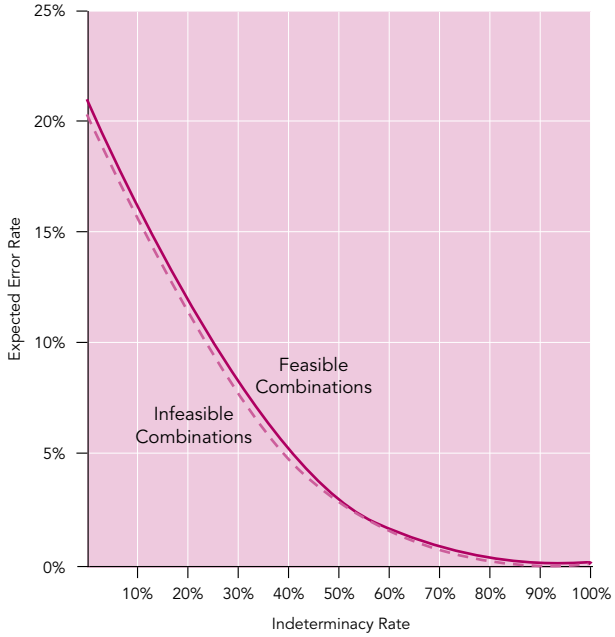
Without any additional information, we can only deduct that the rate of disagreement lies somewhere between 30 percent and 90 percent. This is obviously a wide interval. Whether the true degree of inconsistency will be closer to the lower or the upper bound depends on whether the judges are concordant or discordant. Two judges are concordant if the first judge reaches positive decisions more frequently in cases in which the other judge reaches positive decisions, and discordant if the first reaches positive decisions more frequently in cases in which the other reaches negative decisions. If judges are perfectly concordant, then inconsistency achieves the lower bound, and if they are perfectly discordant, inconsistency achieves the upper bound. Unfortunately, unless we have more detailed information from a representative sample of cases in which decisions are jointly observable, claims about concordance or discordance between judges are not empirically testable. Although intuition about judicial behaviour may narrow the range of inconsistency, this is far from a perfect guide.

b. Although the rates of indeterminacy and error cannot be measured in isolation, it is possible to derive joint bounds on these rates. It is possible to construct an "indeterminacy-error curve", which depicts how inter-judge disparity can be decomposed into indeterminacy and error. A simplified illustration may be helpful here. Suppose there are three judges (A, B and C). The true rate of indeterminacy is I. The proportion of cases in which the only correct outcome is a positive decision is x. Suppose judge A is never in the wrong. Then his (positive) decision rate D_A must satisfy $x < D_A < x + I$. Now suppose the decision rates of judges B and C are D_B and D_C, with $D_B < x < D_A < x + I < D_C$. Then judge B would be in the wrong in at least $x - D_B$ of the cases, and judge C in at least $D_C - x - I$ of the cases. Note that because judge A's decision rate is within the permissible range, it is possible that he is never in the wrong. In reality, the true rate of indeterminacy and the proportion of cases in which the only correct outcome is a positive decision are not known. But by keeping I fixed and allowing x to vary within the possible range of values, it is feasible to derive the lower bound for the expected error rate as a function of the indeterminacy rate.

Fischman applies this to observational data involving asylum adjudication in the New York Immigration Court. In this court, there are large inter-judge disparities in the resolution of claims. With a few exceptions, cases are randomly assigned. The figure

below shows the indeterminacy-error curve. If one makes the assumption that each case has only one correct result (I=0), then we can expect that at least 20 percent of the cases are wrongly decided. If however we assume that half of the cases are indeterminate, we should expect that at least about 2.5 percent of decisions are incorrect.

Indeterminacy-Error Curve: Affirmative Asylum Cases Involving Chinese Aliens, New York Immigration Court, 2003



Note: Solid line is based on sample decision rates. Dashed line is a 99% confidence curve.

The article by Fischman is an interesting one. However, we need to stress several limitations:

i. The decision rates concern the universe of litigated cases, not the universe of potential disputes. As Priest and Klein (1984) and many others have argued, litigated disputes are not representative of the whole set of potential disputes. For example, the rate of disagreement among judges may be very low for the great

majority of disputes which do not end up in court, and may at the same time be large for litigated cases.

ii. The theory is based on the assumption that cases are assigned randomly to judges. Thus each case is equally likely to be assigned to each judge. When this is true, the sample decision rate (the proportion of positive decisions) can be used. In reality however, cases are not always assigned randomly. When cases are not assigned randomly, regression models need to be incorporated which explain the likelihood of assignment (e.g. time periods, districts, etc).

iii. The methodology can help to shed light on which combinations of error and indeterminacy are possible, but cannot illuminate whether any combination is more likely than any other. In order to be able to provide more detailed results, we need to either gather qualitative observations or make subjective assumptions. For example, lawyers familiar with a certain type of cases may propose bounds on the correct rate of positive decisions, which could lead to more precise statements about the level of indeterminacy. Or in fields of law in which there is virtual consensus that the law is determinate, the lower bound of error could be determined more precisely.

iv. In the theory of Fischman, case outcomes can be coded dichotomously. The correct result is either a positive or a negative decision. In reality of course, things can be much more complex. For example, if a court provides relief to a plaintiff, this may be correct, but if the court still needs to make a choice about the type of remedy, it could also err here.

4.2.3. Using stock market reactions

In an unpublished paper, Marco (2006) uses information from stock market reactions to patent litigation decisions and to patent grants to structurally estimate the probability of validity, and the probability of Type I and Type II errors by courts in patent litigation. The estimation quantifies beliefs about patent validity and court errors in a Bayesian context by relying on observable win rates and stock market reactions. The author estimates that the underlying beliefs about validity average from 0.55 to 0.70 for litigated patents. For a number of different specifications, he finds that Type I errors (finding a valid patent invalid) occur with an estimated probability 20 to 25 percent. The range for Type II errors (finding an invalid patent valid) varies more broadly: from almost zero probability to 40 percent. So while the estimates of Type I errors (false negatives) are quite stable, the estimates of Type II errors (false positives) vary widely. This could potentially reflect more uncertainty in the market's beliefs surround-

ing court error for invalid patents. These results are of importance to anyone interested in legal reform, including policy makers interested in tort reform.

This methodology for estimating court error is very interesting, but we need to stress two limitations:

i. Courts may have a different rate of error for the cases they handle relative to the whole population of patents. In other words, there may be a selection bias.
ii. The methodology, if correct, can only be used in a limited number of case types, due to the fact that it relies on stock market reactions.

4.2.4. Expert evaluations

(a) Shavell (1995) mentions the possibility of having experts review samples of cases (which either resulted in appeal, settled after appeal, were not appealed or for which the threat of appeal and settlement did not occur) to examine the error rate.

(b) As explained before, insurance companies often ask one or more physicians to evaluate a claim shortly after its receipt. The great majority of (U.S.) studies relying on these assessments conclude that the outcomes of malpractice claims bear a good correlation with the quality of care provided to the patient (at least as judged by other physicians). However, some remarks are in order. First, the correlation is high for cases in which there is weak evidence of medical negligence. Physicians win around 80 percent to 90 percent of these cases (Taragin *et al.*, 1991; Studdert *et al.*, 2006). This can be seen as evidence of a relatively low level of Type I errors. This 10 to 20 percent Type I error rate may even be an exaggeration for several reasons: (1) the determination about physician care is often made very early after a claim was generated and may have become inaccurate as more information became available; (2) the review process may be biased toward an assessment in the physician's favour, since it is physician-based. Physicians are reluctant to label the conduct of another physician as negligent (Weiler *et al.*, 1993). Now we turn to the other cases. For the borderline cases and the cases in which there is strong evidence of medical negligence, things are different however. Physicians win approximately 70 percent of borderline cases, and 50 percent of cases with strong evidence of medical negligence (Peters, Philip, 2009). This may lead one to conclude that there is quite a high frequency of Type II errors in this field.

The level of accuracy in the courts trickles down to the settlement stage. Most empirical studies focusing on medical malpractice find that the likelihood of a settlement payment and the size of any payment is correlated with the strength of the evidence alleging negligence (see Peters, Philip, 2009). For example, Farber and White (1991) found a mean settlement of $14,109 for a case with "good" care, $146,160 for "ambiguous care", and $203,209 for "bad" care. More recently, Studdert *et al.* (2006) analyzed 1452 (randomly selected) claims files from five large malpractice insurers in the United States. The study used independent reviewers who had access to the entire closed-claims file. Claims were divided into six categories to measure the level of confidence of the reviewer for a determination of fault (from "little or no evidence" to "virtually certain evidence" of error). Payments were made in 19 percent of the claims with "little or no evidence of error", 32 percent of the claims with "slight to modest evidence", 52 percent of claims considered to be "a close call but less than 50-50", 61 percent of those rated as "close call but greater than 50-50", 72 percent of claims with "moderate-to-strong evidence," and 84 percent of the claims with "virtually certain evidence".

One study found no significant relationship between the settlement rate and case quality (Brennan *et al.*, 1996). The study concludes that "the determination of negligence may be an expensive sideshow". This study is often used by advocates of medical malpractice reform. However, several aspects of the study require scrutiny: (1) the sample size is very small (52 lawsuits); (2) in those cases in which reviewers had a different opinion regarding the issue of negligence, the case was systematically classified as a "good care" case. Of course, this artificially inflates the settlement rate for these cases; (3) the sample size may not have been representative of the underlying population because the plaintiffs with strong claims had considerably lower success obtaining a settlement than claimants with good claims in other studies.

4.2.5. Judicial awards and factors of the economic model

Osborne (1999) examines the extent to which judicial awards reflect the factors which need to be taken into account according to the economic model of litigation. He measures court awards against measurable valuations of damages (e.g. medical expenses already incurred, anticipated future medical expenses, cost of repairing property, the cost to replace destroyed property) by using a litigation set compiled by the Rand Corporation (data gathered in 1989 and 1990 from state-court tort lawsuits in

Fairfax County, Virginia). The set contains estimates by the litigants of medical costs and monetary damages suffered from lost or damaged property, and data on court awards for cases that ended with a verdict. Observations with awards of zero were deleted. Osborne finds that the measurable damages to property and body explain a significant amount of the variance in awards (adjusted R^2 of 0.21).

We would stress that the data used by Osborne are not ideal. First, cases in which damages were larger than $35,000 were excluded from the survey. Second, information was lacking for many economic relevant factors in damage calculation (e.g. loss of pay). Third, the number of observations is small (46).

4.2.6. Comparing assessed value with market value

Doerner and Ihlanfeldt (2014) evaluate the property tax appeals process for a large Florida County (Miami-Dade County). One of the main goals of their study is to ascertain whether the appeals process rectifies assessment errors. They measure error by comparing the assessed value with the market value of the property. To do this, the authors look at the price of the properties sold in the 24 months after the month of assessment (correcting for general inflation). Of course, using post-assessment sales to obtain market values is not flawless. Because not all properties are sold at market value, some sales were excluded (e.g. foreclosure-related sales). Because a property may undergo a physical transformation after it is assessed but before it is sold, properties with a recorded change in square footage were eliminated. Because only 7 percent of the properties on the tax roll are sold in a given year, there is potential for sample selection bias. The authors test and correct for bias (post-assessment sales periods were varied from 24 months to 12.6 and 3 months; market value is also assessed via a hedonic model). Turning to the results, the authors find that a majority of overassessed homeowners receive a reduction in their assessed value, but the assessment is often not adjusted downward enough so that the value would fall in a "correct range". Furthermore, a sizeable percentage of reductions are given to homeowners who already are underassessed prior to appeal.

C. TRIAL DURATION/DELAY

1. Description of trial duration/delay

1.1. What is trial duration, what is court delay?

For claims that do not settle (perhaps after a number of rounds of failed negotiations), we can divide *trial duration* into several distinct periods:

1. The procedural time to prepare the case;
2. The waiting time between the end of period (1) and the beginning of the trial itself;
3. The time between the first day and the last day of the trial;
4. The time between the last day of the trial and the day of the verdict.

Note that periods 1 and 3 are to a large extent within the influence of the parties.[28] The length of period 2 will be the result of the interplay between the demand for and the supply of legal services.

1.2. Determinants

Dividing trial duration into several periods is important because the determinants that influence these periods can be quite different. The procedural time to prepare the case will mainly be determined by the complexity of procedural law and by the incentives of the parties (and their lawyers) to stall litigation.[29] The time between the first and the last day of the trial will be determined to a great extent by the same factors which determinethe expenditures of the parties. For example, the time between the first and the last day of the trial will often be longer when the parties produce more evidence. This will also influence the time the court needs to reach its decision.

Now we turn to court delay. It is generally recognized in the law and economics literature that court delay is simultaneously determined by the supply of legal services (i.e. court output) and the demand for legal services.

28 Note however that when trials take place in a piecemeal fashion, delay may influence the time periods between the various "active" stages of a trial.

29 An empirical study by Castro and Guccio (2014) for example finds that in Italian courts opportunistic behaviour by lawyers and litigants is one of the major determinants of the length of court cases.

Some of the main factors that may – at least theoretically – determine the *supply side* are:

a. he number of judges per capita;
b. The age, education, and experience of judges (e.g. were they previously lawyers?);
c. The incentive schemes for judges (e.g. career possibilities and remuneration structure);
d. The number and quality of supporting staff;
e. The budget of the judiciary;
f. Whether the judiciary has to perform non-judicial tasks, and how many (e.g. land or firm registry);
g. The number of court layers and the number of personnel within these layers and sections of these layers ("court size");
h. The available technology (e.g. computers, adequate software, access to decision making collections etc.);
i. The complexity of the legal system;
j. The percentage of vacancies in relation to (a) – what matters is the number of judges who actually work;
k. The complexity of filed cases;
l. The control exercised by a judicial council;
m. The degree of specialization within courts;
n. The existence of training courses for judges.

Some of the main factors that may – at least theoretically – determine the *demand side* are:

a. The degree of optimism of the parties, which is affected by legal certainty, which is in turn influenced by the quality of precedent and the quality of legislation. This degree of optimism influences the settlement rate; .
b. The types and level of damages for which compensation can be awarded (e.g. are pure economic losses compensated? What is the level of compensation for pain and suffering? Are there caps on certain types of losses?)
c. The number of lawyers and the way they are incentivized (e.g. payment schemes);
d. The availability of close substitutes (e.g. ADR);

e. The direct costs of access to the courts and the indirect costs (e.g. foregone wages, psychological costs);
f. The general propensity to litigate/people's customs;
g. The attitudes of litigants towards risk;
h. The state of the economy;
i. Court delay itself.

The fact that demand could by influenced by court delay and that court delay could be influenced by demand means that we are possibly dealing with a circular relationship. Consequently, the interaction between supply and demand needs to be explicitly taken into account. If a reduction in court delay increases the value of litigation, this means that an outward shift in the supply curve would immediately be followed by an upward shift in the demand curve. Consequently, the new equilibrium may not be far from the old equilibrium. Increasing the capacity of the courts results in a reduction in delays, but this prompts potential litigants to file more cases because their costs will be lower. Thus, court caseloads increase and threaten the productivity benefits. Priest (1989) referred to this as the "congestion equilibrium". In other words, according to this view a certain degree of court delay is inevitable. Interestingly, and less well-known, the reverse may also be true. If the demand curve shifts to the right, then the supply curve may shift upwards. The idea behind this is that when the demand for legal services increases, judges will speed up their production.[30] This finds some confirmation in the empirical literature (see below). Note that all this does not mean that efforts to reduce delay are doomed to fail, but only that several effects working in different directions are present, and that the effects of measures to reduce delay may be smaller than an intuitive analysis would predict.

30 Of course, this mechanism cannot work *ad infinitum*.

2. Consequences of trial duration/delay

2.1. The influence on the settlement range, the incentive to sue and *ex ante* behaviour

Chappe (2012) examines the interaction between court congestion (court delay), the incentives of potential plaintiffs to sue, the incentives of litigants to settle, and the incentives of potential defendants to engage in actions that harm potential plaintiffs. In her model, waiting times affect plaintiffs directly through a waiting cost which may reflect all negative aspects of delay (e.g. net value of a judgment falls with a discount factor; the deterioration in the quality of evidence). When congestion costs decrease (e.g. due to an increase in court capacity), this encourages more plaintiffs to sue, which in turn encourages fewer potential defendants to commit illegal acts. A reduction in court congestion may lead to the counterintuitive result of decreased demand for trials if the impact on the behaviour of the defendant is greater than the impact on the plaintiff. Chappe also shows that an increase in the level of awards may either lower or raise the volume of trials because countervailing effects are at work. Holding everything else constant, increasing awards increases the incentives for plaintiffs to file suit. However, this will cause an increase in the level of deterrence, lowering the volume of suits, and will also raise the level of court congestion, which will discourage plaintiffs to file suit. For similar reasons, an increase in filing fees may also either raise or lower the volume of trials. Note that these results are in accordance with the court congestion hypothesis of Priest (1989): the effects of measures to reduce court delay may be offset by a resulting increase in the demand for litigation.

In the model by Chappe (2012), congestion increases the settlement range because plaintiffs are willing to make lower settlement offers in order to avoid congestion costs. Note however that the possibility of pre-judgment interest may complicate things. For example, Posner (1992, pp. 558-59) argues that pre-judgment interest discourages settlement by increasing the stakes. Gilson and Mnookin (1994) argue that the parties are less likely to act cooperatively as the difference between the market interest rate and the pre-judgment interest rate increases. Miller (1994) argues that pre-judgment interest both discourages and encourages settlement, by respectively increasing the stakes and by reducing the defendant's incentive to delay.

2.2. The influence on the quality of filed cases

Holding everything else constant, if delay makes it more costly for the plaintiff to pursue a claim, this may particularly affect claims with a low expected value, including low quality claims.

2.3. The influence on evidence production and the consequences for trial expenditures, the quality of trial outcomes and settlement outcomes

The further away in time (from the contract breach, from the accident, from the intrusion on property rights etc.) that evidence needs to be produced, the more difficult – expensive – it can become to do so. When it becomes relatively more difficult for the party with a meritorious case compared with a party with a non-meritorious case to produce evidence, the accuracy of the legal system will decline. In terms of the endogenous expenditures model, as more time passes, the unit costs of the parties may increase. This will have a particularly profound impact on the outcome of the trial if the passing of time has an asymmetric effect on the parties. For example, if the passing of time increases the unit costs of the party with a meritorious claim more than the unit costs of a party with an unmeritorious claim, then the probability of victory will decrease for meritorious claims; if the passing of time increases the unit costs of the party with a meritorious claim less than the unit costs of a party with an unmeritorious claim, the probability of victory for unmeritorious claims will increase. These effects may of course also affect the accuracy of settlement outcomes.

2.4. Emotional costs

Uncertainty regarding judicial outcomes may generate stress and anxiety among litigants. This is especially true for certain types of cases. Felsö *et al*. (2007) provide the examples of family law cases and purchase cases with a company involved where the owner and the director are one and the same (here the cases are closely connected with the private lives of those concerned). In other cases, emotional costs may be small or non-existent, such as Disablement Benefits Act cases and purchase cases with big companies involved, where a paid employee is handling the case (Felsö *et al*., 2007).

2.5. Opportunity costs

Furthermore, delays create an inefficiency, since the rationing mechanism generated by an insufficient supply of legal services creates an opportunity cost for litigants without generating any gains for the courts (Vereeck and Mühl, 2001). However, note that the opportunity cost of waiting for the parties is not the value of the time waited, since the parties wait *in absentia*. Rather, the cost is the depreciation of the good they are waiting for.

2.6. Costs due to additional work

Legal delay can cause additional work for several reasons: it may take longer to prepare for the case (e.g. memory fades after a long waiting time); there may be a need for additional communications due to and regarding the delay; a longer delay increases the probability of changes in personnel (e.g. a retiring judge, a lawyer leaving a law firm) and thus the need to transfer the case from one person to another, which means duplicative work.

2.7. Lost investment income

Litigants do not know the outcome of their dispute until it is resolved, and until then, they operate in the presence of uncertainty. This uncertainty makes businesses less prone to invest in and expand their operations. Resources under dispute between litigants are removed from circulation until disputes are resolved. When the duration of litigation increases, the total amount at issue becomes unavailable to any of the litigants for a lengthy period of time. For example, consider a dispute between a supplier and purchaser. The supplier believes the purchaser owes $100,000, but both parties are uncertain as to which party will retain that sum after the verdict. The purchaser cannot invest the $100,000 in new equipment since it may have to pay the supplier. The supplier cannot hire new employees with the $100,000 because it does not have the money in hand and may never actually receive it. Pearsall *et al.* (2012) develop a methodology to estimate these types of losses (see also subsection 4).

2.8. Influence on lawmaking

Court decisions are collective goods, in the sense that the information that they contain is non-rivalrous and public disclosure of this information makes citizens better off by clarifying the law. Unambiguous (or at least less ambiguous) rules lead to more settlement and thus the saving of trial costs. Delays in court stall the creation and circulation of information about the application and interpretation of the law (Vereeck and Mühl, 2000).

2.9. The influence on settlement delay

One may argue that in those areas of the law in which the vast majority of cases settle, trial delay should not be of great concern. This is incorrect not only for the reasons discussed above (trial delay could lead to less accurate outcomes, including settlement outcomes), but also because trial delay can cause settlement delay. The reason is the existence of a deadline effect in bargaining (see also the discussion of the settlement stage). Many cases settle just before the start of a trial. When the start of the trial is delayed, settlement negotiations will take longer. The deadline effect is found both in divergent expectations models and in asymmetric information models. In divergent expectations models, the parties can be so optimistic about their bargaining power at the deadline that both of them are willing to wait and reach an agreement just before the deadline. In the asymmetric information model, the party who can make an offer may be willing to make his offer on the courthouse steps so he can optimally extract rents from the other party. Even if the party making the offer can suffer losses due to the delay himself, the extracted rent may be larger. Miceli (1999) offers a second explanation. Defendants may use delay strategically. For example, plaintiffs may differ in their valuation of time. Suppose that Plaintiff 1 discounts future income at a rate of 5 percent, and Plaintiff 2 at 10 percent. A case with a value of 10,000 euros at trial in one year thus has a present value of 10,000/1.05 (=9524) euros for Plaintiff 1 and a value of 10,000/1.1 (=9091) for Plaintiff 2. The defendant may be uncertain which type of plaintiff he is facing, but he may have a rough idea of the distribution of plaintiff types. Suppose the defendant knows that each type is equally likely. He could make an offer of 9524 euros, which will be accepted by both types of plaintiffs. Here, a pooling equilibrium materializes and settlement occurs without delay. Of course, it is possible that the defendant could do better by offering 9091 euros. With such an offer, plaintiffs with a high discount rate will accept the offer

immediately, but plaintiffs with a lower discount rate will reject it and prefer to wait for trial. The defendant will settle with them just before the start of the trial ("on the courthouse steps") for 10,000 euros. For the defendant, the expected cost of this separating equilibrium is equal to (in present value terms) 0.5x9091 + 0.5x10,000/(1+r), with r being the discount rate of the defendant. The defendant will prefer delay if the cost of this is less than 9.524 euros, which is the case as long as r is larger than 0.43. So when the discount rate of defendants is sufficiently high, they will use settlement delay as a way of lowering their expected costs.

3. Fundamental issues related to modelling

3.1. The interaction between court congestion, the incentive to sue and *ex ante* incentives

Chappe (2012) models this as follows. She provides a model with perfect information where court congestion represents an additional cost for plaintiffs to file suit. In more concrete terms, the disutility of congestion equals θY, where θ denotes the waiting cost and Y the congestion level. Y is given by the function D/K, with D the total demand for litigation and K the capacity of the judicial system. A victim who has suffered harm has a probability of prevailing in court p (between 0 and 1), which is drawn from a commonly known cumulative probability distribution $F(p)$. Each defendant has a benefit from wrongdoing, b, drawn from a cumulative function $G(b)$. If the plaintiff wins, he gets an award a. Filing fees are represented by h, trial costs of the plaintiff by Cp and trial costs of the defendant by Cd. The parties are assumed to be risk-neutral.

A plaintiff will sue if the expected benefits of the trial exceed the expected costs, thus if:

$pa > h + C_p + \theta \frac{D}{K}$, thus if:

$$p > \frac{h + C_p}{a} + \frac{\theta}{a}\frac{D}{K}$$

The probability of suit is $S = 1\text{-}F(\dot{p})$, with $\dot{p} = \frac{h + C_p}{a} + \frac{\theta}{a}\frac{D}{K}$ denoting the marginal plaintiff.

Since the defendant has full knowledge, he can compute 1-F(ṗ) and integrate it into his utility function when deciding whether or not to commit an illegal act. His utility function is:

$$U_d(c_d, a, p) = b - \left[1 - F(\dot{p})\right]\left[pa + C_d\right]$$

The defendant commits the legal act if and only if $Ud > 0$, thus if:

$$b > (1\text{-}F(\dot{p}))\,(pa + C_d)$$

Consequently, the probability of civil wrongs is A = 1-G(ḃ), with $\dot{b} = (1\text{-}F(\dot{p}))\,(pa + C_d)$ denoting the marginal defendant.

The aggregate demand for litigation is the product of the individuals who commit civil wrongs and the proportion of plaintiffs who sue, thus:

$$D = S \times A = (1\text{-}G(\dot{b}))\,(1\text{-}F(\dot{p}))$$

By taking the derivate of D with respect to K, a and h, one can then examine the influence of an increase in the capacity of the judiciary, damages awarded and filing fees.

3.2. The influence on trial costs and on the outcome of a trial

As described above, we can examine the fact that delay makes the production of evidence more difficult from the perspective of endogenous expenditure models. These models can shed light on changes in equilibrium expenditures as well as in changes in the outcome of a trial. These elements can then be used to predict changes in settlement outcomes. Technically, when evidence becomes more difficult to produce, this means that the unit costs *Cp* and *Cd* may increase. The exact consequences for the expenditures and the trial outcome will depend *inter alia* on how $\frac{C_p}{C_d}$ changes. When this ratio is larger than one, the plaintiff is more adversely affected by the delay; when this ratio is smaller than one, the defendant is more negatively affected.

4. Empirical issues

4.1. Does a shift in the demand curve induce a shift in the supply curve?

Several authors have empirically established that an increase in newly filed cases leads to more cases being resolved by the courts. Beenstock and Haitovsky (2004) study Israeli courts and find this result using panel data and fixed effects. Dimitrova-Grajzl (2012) use the instrumental variables technique to establish the same result.

We can also mention a Dutch study which focuses on the developments and differences in productivity of the 19 district courts and five appeal courts in the Netherlands in the period 2002-2005, and analyzes the extent to which the developments and differences can be explained.[31] The study finds that the pool of cases has a demonstrable influence on output at both district courts and appeal courts. A relatively high number of cases can lead to a higher productivity (but will cause the throughput times to increase).

4.2. Does an increase in the number of judges increase court output?

The presumption that additional judicial staffing increases court output has underpinned a variety of court reform efforts worldwide (Buscaglia and Dakolias, 1999). Some empirical research however sheds some doubt on this assumption. Beenstock and Haitovky (2004) and Dimitrova-Grajzl *et al.* (2012) do not find a statistically significant effect of additional judicial staffing on the output of courts (in Israel and Slovenia respectively). A possible reason is that judicial output is endogenous: judges reduce their productivity when additional judges are appointed (Beenstock and Haitovsky, 2004, p. 366). In other words, the supply curve may not shift when additional judges are appointed due to effects working in opposite directions. Dimitrova-Grajzl *et al.* (2015), focusing on Bulgarian courts and addressing endogeneity concerns with a fixed effects framework and instrumental variables techniques, find that the number of judges is statistically significantly associated with court output in small district courts but not in large district courts. In the small district courts, output responds positively but inelastically to the number of serving judges. This variation in the effect of judicial staffing on court output in small versus large courts may reflect differences in the marginal utility of leisure between judges in small and large district courts. While the average caseload per judge in large district courts was 499 in Bulgaria, it was "only" 338 in small district courts. If the judges in the larger courts feel overburdened, in light of

31 *Rechtspraak: productiviteit in perspectief*, SCP/Rvdr, 2007, available at https://www.rechtspraak.nl/SiteCollectionDocuments/Rechtspraak-productiviteit-in-perspectief.pdf.

new appointments, incumbent judges may prefer that new judges take over a portion of their caseload rather than resolve any new cases. The result is a decrease in case dispositions by incumbent judges, and very few or no additional disposed cases at the court level. At the small district courts however, the marginal utility from leisure is relatively lower because judicial workload is smaller on average. When new judges are appointed, the incumbent judges reduce their effort to a lesser extent than their colleagues in large district courts. Here, an increase in total court output is more likely.

The Dutch study mentioned in 4.1. also sheds some light on the link between the number of judges and court outputs. Overall labour productivity in the 19 district courts increased by 5 percent between 2002 and 2005. This was the result of a productivity increase of 16 percent and an increase in personnel strength of 11 percent. At the appeal courts, labour productivity increased by 4 percent in the same period, which was the result of a productivity increase of 25 percent in combination with a staffing increase of 21 percent.

4.3. Measuring the capacity of the legal system

Referring to the article by Chappe (2012), empirically, it may not be easy to evaluate the capacity of a judicial system. It depends not only on the number of judges (and judicial staff) and on the budget of the justice ministry, but also on the productivity of judges. One proxy could be the number of cases adjudicated (Chappe, 2012).

4.4. Is there a trade-off between quantity and quality?

A larger number of resolved cases per judge could lower the quality of court decisions. Rosales-Lopez (2008) examined whether the verdicts of more productive judges in Spanish courts are more often reversed on appeal. She found no difference in reversal rates between courts performing above and courts performing below average. Dimitrova-Grajzl *et al*. (2015) study case disposition in Bulgaria and do not find evidence of a higher appeal rate for decisions by more productive courts.

We need to stress however that, as we have explained in the section on accuracy, appeal rates and reversal rates may not be good proxies for the quality of judicial decisions.

4.5. The influence of court delay on the incentive to sue

Landoni uses data from 90 of the 94 U.S. district courts spanning 14 years. The main analytical technique used is multivariate regression on panel data, with district fixed effects, which makes it possible to control for specific time-invariant characteristics of each district. The author finds that expected time of litigation, represented by the median time to civil disposition in a given year, has a clear and strong negative association with the number of cases filed. A 10 percent increase in duration is associated with a 2.6 percent decrease in filings. Landoni further stresses that the negative association between delay and number of filings does not by itself show the direction of causality. It could be that people are discouraged from filing a lawsuit when lawsuits take longer, but also that judges with larger or smaller caseloads dispose of cases faster or more slowly respectively.

4.6. The influence of court delay on settlement delay

In line with the theory discussed above, Kessler (1996) finds that the time taken to settle automobile bodily injury insurance claims increases with delay in trial courts.

4.7. Efforts to estimate the costs of judicial delay

4.7.1. Surveys on the willingness to pay to reduce delay

Felsö *et al*. (2007) undertake 12 case studies (four in the field of the Disablement Benefits Act, three related to purchase contracts where both parties are companies and five in the area of parental access arrangements). Cases were selected in such a way that the sample for each case type includes cases with a normal duration (i.e. around the median) as well as a long duration (i.e. just under the ninth decile). Respondents were asked "willingness to pay" questions in the form of "How much would you be prepared to pay to speed up the case by x number of months?". With respect to the Disablement Benefits Act cases, although all respondents experienced the duration of the procedure to a certain extent as unpleasant, none would have paid more money to shorten the procedure by six months. With respect to purchase cases, two defendants state that they would have been willing to pay an amount between 2,000 and 3,000 euros to halve the duration of the court case. Regarding the parental access arrangement cases, it was concluded that the question of willingness to pay cannot really be applied in that context.

We need to stress that these results must be interpreted with caution for several reasons, some of them obvious and some rather more subtle. First, the specific survey examined only a very limited number of cases. Second, surveys in general do not look at the actual actions of litigants. Third, the answers of the respondents may not only reflect their willingness to pay, but also their ability to pay. So the true costs of delay may have been underestimated. Fourth, the fact that it may be in the private interest of a litigant to reduce delay to a certain extent does not necessarily mean that it would be socially beneficial to do so..

4.7.2. Lost investment income due to delay

Pearsall *et al.* (2012) develop a methodology to estimate the lost investment income due to delay in litigation. To quantify this loss, Pearsall *et al.* (2012) look at the median trial award in a given year. Then they look at the number of unresolved cases in that year (the number of filings minus the number of dispositions). By multiplying the median trial award by the number of unresolved cases, they obtain the total monetary amount of pending civil litigation for the given year. The economic loss reflects the difference between the market-rate return and the risk-free return on the funds. If the funds were reserved in risk-free accounts due to the uncertainty of the litigation outcome, the return on the funds can be estimated using the one-year Treasury Bill rate. If however litigants did not have to hold the funds in reserve, they would generate a return equal to the weighted average cost of capital. Since funds reserved for litigation cannot be invested optimally, the loss can be estimated as the difference between the likely return of optimal investments and the return from risk-free investments. Based on this methodology, the authors estimate the loss due to reduced judiciary funding for U.S. state civil litigation to be $52.2 billion for the period 2009-2013.

A similar methodology was used in a Dutch study which provides an analysis of the gains to society that could be achieved by a reorganization of the judiciary.[32] In that study, using estimates of the total value of the cases, the costs incurred during the court case were approximated by the average return on investment minus the statutory interest rate which was used to calculate the compensation the losing party had to pay the winning party for the delay, for the average time taken by the courts to adjudicate the cases. The benefits of a feasible reduction of court delay for civil[33] and

32 Ministerie van Justitie, Rechtspraak en rechtshandhaving; maatschappelijke effecten van verbetering, Werkgroep Effecten Rechtspraak, research group: F. van Dijk, J. van Dijk and R. Teijl. Ministerie van Justitie, 1998.

33 Eight months for large claims (>5,000 euros) and two months for small claims (<5,000 euros).

administrative[34] cases at the first instance courts were estimated at 430 million euros per year in 1998. Van Dijk (2014), using better data for the value at stake, finds results that are slightly larger than the original estimate.

These studies are highly interesting, but certain criticisms must be noted:

1. Together with van Dijk (2014), we would stress the importance of not deducting the statutory interest rate. Deducting the statutory interest rate may lead to an underestimation of the impact of court delay. The statutory interest rate is only a matter of redistribution between the parties. It can be much larger than the risk-free return on the funds.
2. The methodology is difficult to apply to non-monetary claims. Many civil cases – and in particular many administrative cases – are not about direct financial claims (e.g. a denial of a permit to operate an industrial plant).
3. The figures may underestimate the true costs of delay, and may thus serve only as a lower bound, for several reasons:
   - delay may lead a victim to decide not to file a claim at all, which can have a negative influence on e.g. a contractual party's *ex ante* incentives;
   - some effects are not taken into account, e.g. the effect of the deterioration of evidence.
4. The figures may also overestimate the costs of delay for some types of disputes. Corrections in the methodology may be warranted, e.g. in the form of multipliers smaller than one. As discussed above, the study by Felso (2007) shows that the negative effects of delay are not large in some types of cases. Negative consequences are however substantial in trade cases. This may particularly apply to small and medium-sized companies, which may face more severe liquidity constraints than large companies.

34 Six months.

5

Stage V.
The appellate stage

1. Introduction

Parties involved in judicial disputes often have the right to appeal against a verdict by trial courts and bring the dispute to a higher court for review. The appeal procedure heavily influences the functioning of the judicial system. For example, it may substantially affect the length of proceedings and thus the time parties have to wait for a final judicial decision. With respect to Europe, the processing time may be more than twice as long in appealed cases (CEPEJ, 2005). Furthermore, appeals strongly influence the public expenditures related to litigation. An excessive number of appeals could reflect inefficient performance by the lower courts, and so more resources than necessary could be required by the judicial system. For many countries, the appeal rate (annual number of appealed cases divided by cases decided in trial courts) ranges between 15 and 20 percent (e.g. U.S., Germany, Italy and Spain, see CEPEJ, 2005 and DOJ, 2006). In the Netherlands, for family court cases, the total appeal rate was 4 percent in the period between 2009 and 2013. The total appeal rate for commercial law cases at the subdistrict courts was 8 to 9 percent in the period between 2009 and 2013 (default proceedings are excluded here). For commercial law cases at the civil-law section of the district courts, the appeal percentage increased from 18 percent (three-year average 2009-2011) to 30% (three-year average 2011- 2013) (Van Os and Smit, 2016).

The hierarchical structure of the judiciary has been examined from various points of view. In "team models", it is assumed that judges in a hierarchy share the same values, and only care about achieving the correct outcome in a given case. The main focus of study is how to best organize the hierarchy to minimize the number of incorrect outcomes. In this approach, the main problem of adjudication is informational. Although all judges agree on the correct outcome, due to resource constraints or

informational advantages of the litigants, they may lack understanding of the relevant facts. Kornhauser (1995) sums up the various advantages of a hierarchical structure compared with a flat organization for achieving the goal of error correction: it allows for a specialization of labour, fewer resources are wasted while scanning cases for precedential value, and judicial resources can more easily be shifted towards important precedential cases. Shavell (1995) argued that when the probability of reversal of an incorrect trial court verdict is larger than the probability of reversal of a correct verdict, there may be a separation of disappointed litigants by selecting the right fee or subsidy: those who are the victim of error find it worthwhile to bring an appeal, and those who have not been the victim of error do not find it worthwhile to appeal. Note that this works independently of the absolute levels of reversal. Obviously, selecting the appropriate fee or subsidy requires substantial information. Shavell also showed that relying on giving litigants the choice to appeal is superior to a random review of lower court decisions. Cameron and Kornhauser (2006) however argue that the equilibrium in which only incorrect decisions are appealed is delicate in a setting with asymmetric information. In their model, there is asymmetric information with respect to the liability issue of the defendant before the trial court. After the trial however, both parties become perfectly informed about this issue, but the appeal court judges only have a probabilistic understanding of this. The authors show that in a two-tiered hierarchy (one appellate court), when the level of accuracy of the trial courts is sufficiently large only incorrect decisions are appealed. The trial courts' level of accuracy decreases with their caseload. When the lower courts are overburdened relative to their resources, many losing litigants who lost deservedly will file an appeal.

In contrast with this, "agency models" view judges as political actors with differing preferences over legal rules and case outcomes. The main focus here is how to minimize the rate of non-compliance by the lower courts with higher court rules. These models apply the classic principal-agent framework, with the higher courts as the principal and the lower courts as agents. Unlike in "team models" in which errors are mistakes caused by imperfect information, errors in "agency models" are deliberate decisions by lower court judges to apply their favoured rule rather than the rule preferred by a higher court. Although high court justices do not have the classic tools superiors in firms can deploy to incentivize their subordinates (e.g. a higher salary or dismissal), studies generally find widespread compliance by lower courts (Songer, Segal and Cameron 1994; Benesh and Reddick, 2002). Agency models have explained this by relying heavily on the assumption that lower court judges suffer a

utility loss from having their decisions reversed by a higher court (Cameron, 1993; McNollgast, 1995). Note that the assumption that judges fear reversal is debated, especially for courts whose decisions are rarely reviewed by higher courts (Klein and Hume, 2003; Cross, 2005; Choi, Gulati and Posner, 2012).

Obviously, the determinants of the key elements of the appeal stage will resemble the determinants we have discussed in the trial stage. For example, the expenditures of the parties in the appeal stage will also depend on the inherent merits of the case, the amount at stake, the unit costs of the parties and the variability of the decision maker. Of course, the *values* of these determinants can change going from the trial stage to the appellate stage. For example, a plaintiff who initially thinks that the inherent merit of his case is extremely high may lower his expectations on appeal after the trial court has given a verdict to his detriment. Similarly, the values of the amount at stake, the unit costs and the variability of the decision maker may change, but they remain key determinants.

While the appellate courts and the appeals process have been examined by economic scholars from various points of view, it is only quite recently that the various decisions related to first instance proceedings (to file, settle, go to trial, appeal the verdict or not) have been modelled taking into account the possibility of appeal. We will now give an overview of these recent advances in the literature.

2. Consequences of introducing the possibility to appeal

2.1. A divergent expectations model

Kamphorst and van Velthoven (2009) look at the connection between the courts in the first and second instance from the perspective of the litigants. Their model shows how complicated the analysis becomes when the appellate stage is added. They rely on a numerical approach to derive meaningful results. We will return to this when looking at modelling issues related to adding an appeal stage. For now, we will discuss their main findings. The authors show that the consequences of adding an appeal stage depend on whether the parties learn and adjust their initial beliefs at the start of the appellate stage in light of the decision of the first instance court. They distinguish three situations. First, parties may be naive. In that case, they stick with their initial beliefs, notwithstanding any additional information that becomes available. Second, par-

ties may be myopic. While these parties are not aware that the decision of the first court may affect their beliefs, once they enter the second stage they become aware of the additional information which follows from the court verdict and adjust their beliefs. Finally, the parties may be far-sighted. Here, the litigants are fully aware how learning will affect their future beliefs, right from the start.

When parties are naive, adding an appellate stage has the following consequences. First, all trials in first instance are followed by an appeal. This follows from the facts/assumptions that (a) by definition these parties stick with their initial probability estimates of winning, (b) the joint money claim is similar in both stages, and (c) legal costs in an appeal are substantially smaller than in the trial in first instance, because the case has already been prepared and defended. Consequently, it is certain that the losing party will appeal. Second, introducing an appeals court leads to more settlement (relative to trial) in the first instance. This follows from the first consequence. If an appeal is certain, then litigants know at the first stage that going to court will bring them the combined costs of the two stages, and not just the cost of the first stage. And given that the joint money claim is similar with and without appeal, the litigants have a stronger incentive to settle. Third, whether the plaintiff in the first stage will give up (drop the suit) more or less often depends on the actual distribution of the parties' probability estimates of winning in court. When the (first stage) defendant rates his chances in court as relatively low, the plaintiff is less likely to give up. One element of the intuition is that even if the plaintiff loses the first stage, he may still get a relatively favourable settlement after that stage. But as the defendant's estimate of his chances in court increase, any settlement in the second stage becomes less favourable.

When parties are myopic, some things change, while others remain the same. First, given that players learn from the verdict *ex post*, they may now act differently and not necessarily file an appeal. The party who won the trial in first instance gains some faith in the merits of the case and the party who lost loses some of his faith. The appeal rate varies with the level of confidence in the courts. For example, when the confidence in the court of first instance is relatively low, litigants will be less inclined to take its verdict at face value. Second, as with naive parties, the introduction of the court of second instance leads to more settlement (relative to trial) in first instance. This follows from the fact that myopic litigants do not take into account *ex ante* that they may learn from the verdict of the first instance court. Thus, as far as decision

making in the first stage is concerned, the analysis remains the same. Note further that parties are less likely to end up in a trial when confidence in the courts is lower. This is because any joint optimism of the parties will decrease as confidence in the court declines.

When parties are far-sighted, the consequences are very similar to the case in which parties are myopic. Once again, depending on the level of confidence in the courts, the appeal rate varies between zero and one. Further, the introduction of an appeal court leads to more settlement (less trial) in first instance. And given the parties' faith in the merits of their case, the parties are more likely to end up in trial when confidence in the courts is higher. The appeal rate decreases when confidence in the courts is higher. However, far-sighted players are more likely to go to trial instead of settle their claim in the first stage. Far-sighted players have different expectations in the first stage about the outcome of the second stage, because they know they will learn from the first instance verdict. Unlike myopic players, they may be aware that players in the second stage may settle or give up. This lowers the expected trial costs, which decreases the probability of settlement.

2.2. An asymmetric information model

Wohlschlegel (2014) analyzes how the availability of appeal influences the decisions of litigants at earlier stages of the legal process when the defendant has private information. The plaintiff may make a take-it-or-leave it settlement demand. In the event that the defendant rejects, the case goes to trial. If an appeal is filed, the plaintiff can make another take-it-or-leave it demand. If the defendant rejects it, the case will come before the appellate court, the final stage of the dispute. Wohlschlegel identifies two effects which influence the litigants' decisions. First, there is a "strategic effect". When choosing the pre-trial settlement demand, the plaintiff anticipates that her post-trial pay-off will depend on the private information that a defendant who has rejected the pre-trial settlement demand may have. The plaintiff will take into account which private information will cause the defendant to reject or accept the pre-trial demand, and how this will affect post-trial pay-offs. Thus, the plaintiff chooses the pre-trial settlement demand so as to optimize the strategic environment in which post-trial decisions happen. The second effect is an "information effect". Litigants take into account at the pre-trial stage how the information revealed by the verdict of the court of first instance will affect post-trial pay-offs. By making a tougher pre-trial settlement

demand, the plaintiff will improve the average case that goes to trial and thus increase his probability of winning in trial. The plaintiff will prefer to make a higher pre-trial settlement demand if the difference in pay-offs after winning and losing in trial court is higher. The precise effects of an appellate stage will depend on which effect dominates. If the information effect dominates, some counter-intuitive results follow. First, higher appellate legal costs will increase the *ex ante* probability that a case goes to trial court. Second, more accurate trial courts may attract fewer cases because the defendant expects to earn a lower information rent post-trial and will thus accept higher pre-trial settlement demands.

2.3. The lawmaking function of appellate courts: the influence of the judicial workload

When appellate caseloads increase, one possibility for appellate courts to deal with this is to make changes to their personnel, structure, and procedure (Marvell, 1989). Recently however, De Mot, Faure and Klick (2015) have shown that appellate courts may also respond to increasing caseloads by changing substantive legal rules. More precisely, they show that the switch from contributory to comparative negligence in the United States was not exclusively motivated out of a concern for justice. Language in state supreme court decisions suggested that some judges thought the switch would reduce appeal rates. De Mot, Faure and Klick hypothesize that courts were more likely to make the switch when their appellate caseloads were relatively high. They estimate hazard models, showing that states with appellate courts where caseloads grew relatively faster made the switch more quickly, and show that the effect was more pronounced for the switch to the pure, as opposed to the modified, form of comparative negligence. In the remainder of this section, we discuss the theoretical underpinnings for the fact that increased caseloads can lead to changes in judge-made law. In the section on empirical research, we discuss some problems empirical research faces in proving this.

From a theoretical point of view, the conjecture advanced above fits into a line of research that argues that judges are rational utility maximizers with relatively weak performance incentives and constraints on their decision making, at least at the highest levels. This issue has been stressed by Cooter (1983) and Posner (1993). Both authors assume that judges seek to minimize effort subject to various institutional constraints. Cooter assumes that judges providing private services have a financial

incentive to increase their caseload to the extent that it increases their income. In Posner's approach, focusing on federal judges, income is fixed and therefore cannot be increased by more effort. For Posner judicial utility is a function of income, status, and leisure. Since the income of judges is largely fixed, maximizing leisure becomes especially important, conditional on maintaining status levels (Posner 1993; Stras 2006). Posner further predicts that judges who have reached a high income level (e.g. Supreme Court justices) will prefer to maximize leisure.[1] Furthermore, "the opportunities for a leisured judicial life, especially at the appellate level, are abundant" (Posner, 2008, p. 61). Hence one can expect judges to try to reduce their workload. In theory, judges confronted with an increased workload could maximize leisure by simply deciding fewer cases. This would unavoidably lead to increased court congestion and a backlog of cases. This could harm the reputation of judges and will likely be avoided (Helland and Klick, 2007; Beenstock and Haitovsky, 2004). Judges could also lobby the legislator for more judges to deal with increasing workloads. However, this could reduce the prestige of the judges as more people attain the position. The judiciary will then look for alternative ways to reduce its workload according to Posner's model of judicial behaviour. Besides the research of De Mot, Faure and Klick (2015) to be discussed below, there is some other empirical evidence supporting this. For example, Helland and Klick (2007) show that judges in class action cases have an incentive to easily grant the attorney's fee request in order to terminate cases rapidly, thus avoiding court congestion. Research from Israel also shows that judges, for reputational reasons, will avoid a large case backlog and hence will dispose of more cases when the caseload increases (Beenstock and Haitovsky, 2004).

2.4. Appeals and the reduction of error

The appeal system is considered to be a tool for error correction (Shavell, 1995) because it harvests private information from litigants and allows better skilled or better informed judges to review the case.

Moreover, the appeals system may also reduce the incidence of error in an indirect way, by altering the incentives of judges in the lower courts. These judges are often assumed to dislike reversals due to reputation concerns (Shavell, 2006), and thus exert more effort in order to avoid reversal. Recently however, Chopard *et al*. (2014) have argued that this effort-increasing effect of reputational concerns may be perfectly balanced by a reduction in effort due to free-riding on the appeal court's error correction. However, this counterbalancing effect found by Chopard *et al*. depends on two very

1 In Posner's words: "I therefore predict that a higher judicial salary is likely to reduce the amount of work done by existing judges" (Posner, 1993, p. 33).

strong assumptions. First, they assume that opportunistic appeals (i.e. appeals against a correct result) are never filed. In other words, wrongful reversals are assumed not to occur. Second, they assume that the accuracy of the appeal court is exogenous. In other words, it is not affected by the effort of the judge's effort.

3. Modelling issues

3.1. Divergent expectations framework

Taking the possibility to appeal into account when modelling the litigation process introduces several complexities. Obviously, the introduction of the option to appeal considerably enlarges the strategy set of the parties. Solving the model by backwards induction becomes significantly more complex.

1. A first issue is that assumptions need to be made about whether the initial beliefs of the parties may change after the first verdict, and more importantly, whether the parties foresee at the first stage that their initial beliefs may change at the second stage. Kamphorst and Van Velthoven (2009) tackle this by examining three situations (parties are either naive, myopic or far-sighted, see the discussion above).
   We would stress that things may be even more complex than this. In Kamphorst and Van Velthoven (2009), both parties are either naive, myopic or far-sighted. In reality, parties may differ in the way their decisions at the various stages are related. For example, the plaintiff could be myopic, and the defendant could be far-sighted.
2. If the parties are assumed to be able to learn after the first stage, this process needs to be modelled. Kamphorst and Van Velthoven model learning as follows:
   a. The probability with which a player (i) thinks he will win a trial (i.e. "the first stage") can be dissected into two elements: his or her belief about the intrinsic merits of the case ($\rho_1^i, 0 \leq \rho_1^i \leq 1$), which is an estimate of the probability that he is right, and his or her level of confidence in the court ($r_1, 0 \leq r_1 \leq 1$).). At the first stage, player i expects that his chances of winning are equal to: $p_1^i = \rho_1^i r_1 + (1 - \rho_1^i)(1 - r_1)$
      Consequently, there are two reasons why a player's estimate of winning at the first stage and the second stage may differ: the confidence in the court of appeal may differ from the confidence in the court of first instance, and a party may alter his belief about the intrinsic merits after the verdict of the lower court.

b. When *r1* differs from 1/2, the verdict of the lower court is a signal of which player is right. The players interpret this signal consistently with their initial belief and the confidence they have in the court's verdict. They update their beliefs using Bayes' rule.
   If player i won the first stage, he updates his beliefs about the intrinsic merits of the case as follows: $\rho^i_{2+} = \frac{\rho^i_1 r_1}{\rho^i_1 r_1 + (1-\rho^i_1)(1-r_1)}$

   The numerator is equal to the probability estimate by player i that he is right and wins the first stage. The denominator gives the probability estimate of winning the first stage. The ratio thus gives that party's probability estimate that he is right given that he won the first stage.
   If player i lost the first stage, he updates his beliefs about the intrinsic merits of the case as follows: $\rho^i_{2-} = \frac{\rho^i_1 (1-r_1)}{\rho^i_1 (1-r_1) + (1-\rho^i_1) r_1}$

   The numerator is equal to the probability estimate by player i that he is right and loses the first stage. The denominator gives the probability estimate of losing the first stage. The ratio thus gives that party's probability estimate that he is right given that he lost the first stage.
c. Analogously to the equation in (a), the parties can then calculate their chances of winning the appeal after having won/lost the first stage, by combining the new beliefs in the intrinsic merits of the case with their level of confidence in the appeals court ($r^i_2, 0 \le r^i_2 \le 1$):

$$p^i_{2+} = \rho^i_{2+} r_2 + (1-\rho^i_{2+})(1-r_2)$$
$$p^i_{2-} = \rho^i_{2-} r_2 + (1-\rho^i_{2-})(1-r_2)$$

We would stress that the model by Kamphorst and Van Velthoven (2009) could be further modified so as to incorporate some potential complexities. First, they assume that the level of confidence is the same for both parties. This is not necessarily the case. For example, due to learning effects, repeat players may either have more or less confidence in the courts than one-shot players. Second, the level of confidence is represented by the probability with which the court is expected to judge correctly, without making a distinction between having a good case or a bad case. In reality, the level of confidence may vary with the beliefs about the strength of the case.

3.2. Asymmetric information framework

Here it will suffice to mention that the two-stage model by Wohlschlegel (2014) as described above is an extension of the one-stage model by Bebchuk (1984) which we have described extensively in the discussion of the settlement stage.

4. Empirical issues

4.1. General difficulties empirical research faces with respect to the appellate stage

With regard to analyzing the impact of adding an appellate stage, empirical research faces several difficulties. First, it does not often happen that an appellate stage is added or eliminated. However, it may happen more often that the financial boundaries which a claim must satisfy are strengthened (or weakened). While such reforms may provide opportunities for empirical research, one must take into account the possibility of plaintiffs being encouraged to misrepresent the value of their claim in the initial filing, which may pose a serious challenge. Second, even when a court of appeal is introduced (or abolished), the fact that other changes occur simultaneously complicates matters substantially. For example, the article by Kamphorst and Van Velthoven (2009) was inspired by the introduction of a court of appeal in Dutch tax litigation. However, testing their theory in the domain of tax law could be quite difficult, since substantive tax laws change very rapidly.

4.2. Empirical research on the determinants of the decision to appeal

Santolino (2010) examines the determinants of the decision to appeal against motor bodily injury judgements made by Spanish trial courts. Motor cases represent 10 percent of the total litigation dealt with by Spanish courts. The author implements a method to estimate the influence not only of several claim characteristics but also of the trial court's judgment on the probability that the claimant or the insurer will appeal against the verdict. More precisely, he applies a multinomial logit model (McFadden, 1974), because the dependent variable is nominal and the categories are unordered. The method makes it possible to make a distinction between the decision of the claimant and the decision of the insurer to appeal. The dependent variable covers the

three possible outcomes regarding the decision to appeal (no appeal, claimant appeals, insurer appeals).

The main empirical results are as follows. First, claims in which the insurer is doubtful about the fault of the insured driver are more likely to be appealed (+ 12.7 percent) by the insurance company but less likely to be appealed by the claimant (-6.6 percent). This makes sense. The insurer's expected marginal compensation increases when the full responsibility of the insured is not accepted, whereas the claimant's expected marginal compensation decreases. Second, an increase in the length of time until a judicial resolution was reached in the first instance decreases the claimant's incentive to appeal, but does not affect the insurer's decision. Given that the Spanish judicial system has mechanisms to update the financial amount so as to take into account the time value of money, the author concludes that this result may reflect the underlying wealth constraints of the parties. Wealthier litigants may be able to face a longer period of uncertainty. The result is in accordance with the hypothesis that wealth asymmetries leads to risk-averse behaviour by the victim, but not by the insurer. Third, the claimant's incentive to appeal depends inversely on the size of the compensation awarded. However, the size of the award does not seem to influence the insurer's decision.

We would stress that the database used by Santolino is rather small. It consists of 202 automobile bodily injury claim records provided by one Spanish insurance company. In 57 cases, the trial court's verdict was appealed against.

Seabury (2009) examines the importance of award size and financial constraints on the side of the defendant for the decisions of litigants to settle or appeal a case after a trial. The author uses data from a survey of 562 civil cases that were resolved at trial in California and New York from 2001-2004 with a positive verdict for the plaintiff. The survey questions both plaintiff and defence trial attorneys. From a theoretical perspective, the author remarks that defendants with unlimited financial resources should be more willing to appeal a marginal case (a case in which their position is more uncertain) as the size of the damage award increases. However, when the amount defendants are able to pay is less than the full amount of the verdict, the importance of a reduction by the appellate courts is decreased. Financially-constrained defendants would have more incentive to settle marginal cases. This leads to the empirical prediction that the defendant win rate in appellate courts should increase as the size

of the verdict award increases. The empirical results confirm the theory. First, both settlement and appeal are more frequent in cases with larger jury verdicts. Second, the likelihood of a defendant winning on appeal is higher in high-stakes cases. Third, the positive relationship between post-trial settlement and verdict size is driven to a considerable extent by defendant financial constraints. Consequently, the types of cases that end up in an appeals court seem to be heavily influenced by the ability of the defendant to pay a verdict in full.

6

Stage VI. The enforcement stage

1. Introduction

Civil money judgments are not self-executing. When a plaintiff obtains a judgment, it merely contains an authoritative declaration that the defendant owes the plaintiff a debt in the amount of the judgment. If the defendant does not comply with the judgment voluntarily, the plaintiff must be prepared to take post-judgment action to collect the amount awarded. There are two issues here which directly and indirectly influence private and social costs and benefits.

First, the enforcement procedure entails costs. In the Netherlands, when a debtor fails to pay its creditor, the creditor can resort to a judicial officer (*gerechtsdeurwaarder*) to pursue the official collection of the payment. The judicial officer performs certain acts for which some fees, with an upper limit, are charged to the debtor ("Btag-fees"[1]). The judicial officer makes a separate (financial) agreement with the creditor. It is the creditor who initially covers the costs, but he can recover them from the debtor, at least if the latter has sufficient funds (otherwise the costs continue to be borne by the creditor). In practice, the Btag (maximum) tariffs are applied both in the relationship with debtors and with occasional creditors (Ter Voert and Van Ewijk, 2006).[2]

Second, not every defendant has (collectible) assets. When the amount a court may award to a plaintiff is larger than what the plaintiff is able to collect from the defendant, a defendant is said to be "judgment-proof" (Shavell, 1986).[3] This can be caused

1 These are based on cost assessment studies from 1998 and time assessment studies from 2000. After that, the fees were only adjusted through indexing. See "Tarieven ambtshandelingen gerechtsdeurwaarders", 2016, WODC.

2 Bulk purchasers can negotiate over a lower price.

3 Note that "judgment-proofness" needs to be distinguished from "litigation-proofness". A defendant is litigation-proof when he causes losses which are too small for the plaintiff to file a claim. This is the issue of negative expected value (NEV) suits which we have discussed in stage 2 (the filing stage).

either by the lack of assets on the defendant's side, or by laws which entitle the defendant to be judgment-proof.[4] With respect to the United States, Gilles (2006, p. 607) writes: "...contemporary America, one of the most affluent societies in human history, is simultaneously – and largely by operation of law – a judgment-proof society". In the Netherlands, winning litigants often do not obtain the full (or even a fraction of the) award they obtained in court, even years after the judgment (see below, section 4). One of the most common reasons is that the debtor simply lacks funds.

In the context of this project, we are interested in the decisions of litigants at the enforcement stage, and in the consequences of the enforcement stage on the various other stages of a dispute. It needs to be said that the economic analysis of law has not focused as much attention on this stage as on the other stages. For example, in the famous overview article of Cooter and Rubinfeld (1989), the enforcement stage is not discussed. Also in later standard works like the Encyclopedia of Law and Economics, the enforcement stage itself receives no explicit attention.

With respect to the issue of judgment-proofness, the focus has traditionally been on the *ex ante* stage. With respect to this stage, most of the literature studies the effect of judgment-proofness on the incentives of tortfeasors to take adequate care. However, the judgment-proof problem is not peculiar to tort law. Contract creditors have often bemoaned the existence of exemptions from collection. The judgment-proof problem certainly undermines the efficacy of breach-of-contract remedies in some contexts. Bar-Gill and Ben-Shahar (2005, 724) write: "The judgment-proof problem is a key factor affecting the credibility of a threat to breach a contract as well as the credibility of any other threat to inflict an illegal outcome". This being said, on average it may be harder for tort claimants to succeed as creditors. Tort victims cannot usually investigate the financial status of potential tortfeasors prior to being harmed by them. In contrast, contractual creditors are sometimes in a position to assess the creditworthiness of those with whom they deal, and they may protect themselves by demanding a security interest in property belonging to the debtor.

Recently, several advances have been made with respect to the other stages such as the settlement stage, the trial stage (expenditures) and the appellate stage.

We would stress that models with asset constraints on the side of the defendant will not be relevant for all civil litigation. For example, if the defendant is a large corporation and the amount at stake is relatively small compared with the corporation's

4 E.g. rules exempting various types of assets and income from collection; rules allowing some claims to be discharged in bankruptcy; rules giving secured creditors priority over tort claimants.

assets, these models will not apply. However, when individuals are involved, these models may be highly relevant, especially when the individuals are uninsured or when the policy limits are low. Note also that large corporations are often thought to be able to avoid exposure to liability through judgment-proofing strategies such as asset securitization, manipulation of subsidiaries, and bankruptcy (LoPucki, 1996, pp. 14-37). Finally, section 4 below focuses on empirical issues.

We could not find a model in the existing literature focusing on the decisions of the victorious plaintiff and the losing defendant during the enforcement stage. Here, we develop a simple model ourselves. We introduce the following symbols:

J = the award granted to the plaintiff by the court;
A = the assets of the defendant at the moment of enforcement;
Ep = the costs of enforcement which ultimately rest on the plaintiff (e.g. his cost of time);
Ed = the costs of enforcement which the defendant normally has to pay, but which have to be paid by the plaintiff if the defendant's assets are insufficient .

We need to distinguish between three situations:

1. $A > J + Ed$: In this case, the defendant is able to pay the full judgment to the plaintiff. The costs of enforcement do not diminish his capability to pay J to the plaintiff. Consequently, the plaintiff will enforce the judgment if $Ep < J$. His value of enforcing the judgment thus equals $J - Ep$. The total costs for the defendant equal $J + Ed$. He keeps a part of his assets ($A-J- Ed$). The parties may find it mutually beneficial to agree on an amount between $J - Ep$ and J.
2. $Ed < A < J + Ed$: In this case, the defendant is not able to pay the full judgment and the enforcement costs. If the enforcement costs have priority, the plaintiff will only obtain $A - Ed$ from the defendant. The plaintiff will enforce the judgment if $Ep< A - Ed$. His value of enforcing the judgment equals $A - Ep - Ed$. The defendant loses all his assets (A). The parties may find it mutually beneficial to agree on an amount between $A - Ep - Ed$ and A.
3. $Ed \geq A$. If the plaintiff enforces the judgment, there will be no funds left to pay anything else after the defendant has paid the costs of enforcement. The plaintiff will not enforce the judgment. The defendant keeps his assets.

In cases (1) and (2), there is scope for an agreement (with voluntary payment) between the parties. If agreement is reached, the plaintiff does not have to rely on enforcement agents to obtain a payment. However, an agreement may fail for several reasons. First, the parties may not reach a consensus on how to divide the bargaining surplus. Second, we have assumed that the level of the defendant's assets is common knowledge. This is not necessarily the case. In contrast to the defendant, the plaintiff may be unaware of the exact financial situation of the defendant. Due to asymmetric information regarding the assets of the defendant, an agreement may fail. Some defendants who have sufficient assets may pretend not to have enough assets to pay the judgment. (3) A third potential reason is that the losing defendant does not grant the plaintiff the satisfaction of a voluntary payment, and that he prefers to lose all his assets (due to official enforcement) rather than pay slightly less through a voluntary payment.

2. Consequences of judgment-proofness

2.1. The *ex ante* stage

Summers (1983) and Shavell (1986) showed that judgment-proof injurers are inclined to exercise insufficient precaution. The reason is simple: judgment-proof injurers do not internalize all losses. However, this early literature made two simplifying assumptions: (1) precautions only reduce the probability but not the magnitude of accidental harm (the "probability model"), and (2) precaution costs do not reduce the total assets available for paying compensation. This is the "two-pocket model": it is as if the injurer has one "pocket" for paying precaution costs and another for paying compensation. Such models are realistic for situations in which there is a legal liability cap or when precaution is non-monetary, but not otherwise. Beard (1990) relaxed the second assumption and studied a one-pocket probability model, in which precaution and liability expenses are paid out of the same pocket, and taking care thus reduces the assets available for compensation. A surprising result is that under these circumstances judgment-proofness may lead to overprecaution. Intuitively, taking care reduces the assets available for paying damages, and it is as if a fraction of the cost of precaution taken by the injurer is externalized to the victim in the form of reduced compensation. In other words, there is an implicit precaution-subsidy effect.

Dari-Mattiacci and De Geest (2005, 2006) show that one needs to look at many other situations to get a full picture of the possible consequences of judgment-proofness on

the *ex ante* incentives of injurers. Apart from the "probability model" discussed above, they describe and discuss three other models. In the "magnitude model", care only influences the possible magnitude of the losses. In the "joint-probability-magnitude model", adopting a care measure simultaneously influences the accident probability and the magnitude of the losses (e.g. driving more slowly lowers the probability of an accident and the harm in event that an accident happens). In the "separate-probability-magnitude model", some care measures influence the accident probability (e.g. steering a boat more carefully) while others influence the magnitude of the losses (e.g. having life jackets on board) (Dari-Mattiacci and De Geest, 2005, pp. 42ff). The main conclusions from these additional models are: (1) overprecaution can only concern probability-precaution, and (2) overprecaution may result not only as a consequence of the precaution-subsidy effect, but also by a substitution effect between reducing the probability of an accident and reducing the magnitude of the harm.

The first result can be explained as follows. Judgment-proofness provides the injurer with two implicit subsidies: the implicit harm subsidy (a portion of the harm is externalized on the victim) and the implicit precaution subsidy (the more the injurer spends on precaution, the less remains available for compensation). These subsidies have opposite effects on the incentives to take care: the former reduces these incentives while the latter reinforces them. Most importantly, while the precaution subsidy has comparable effects under all models, the harm subsidy is weaker in probability models than in magnitude models: in a magnitude model, an insolvent injurer fully externalizes the incremental loss due to reduced precaution; in a probability model in contrast, an insolvent injurer only partially externalizes the effects of reduced precaution, which increases the probability to pay a (truncated) award. Thus, the harm subsidy works at a 1:1 ratio in magnitude models, but at a lower ratio in probability models. The precaution subsidy functions at a ratio lower than 1:1 in both models, since the cost of precaution is externalized on the victims only if an accident occurs (thus with a probability lower than 1). Consequently, the harm subsidy always prevails over the precaution subsidy in magnitude models. As a result, overprecaution never occurs.

The second result can only occur under the separate-probability-magnitude model (in which the injurer can control the probability and the magnitude of the harm through two different precautionary measures). In both a one-pocket and a two-pocket model, it may be optimal for an otherwise solvent injurer to decrease his investments in magnitude precaution to zero and to increase his investments in probability precautions. This choice may be privately optimal because the increased harm is externalized to

the degree that it exceeds the injurer's assets, while the benefit of increasing probability precaution is partially internalized by the injurer, since it reduces the probability of losing all his assets.

We summarize the main outcomes for the eight different possibilities in the table below.[5][6][7]

	Two-pocket models	One-pocket models
Probability model	Only underinvestment is possible	Both underinvestment and overinvestment are possible
Magnitude model	Either optimal care or no care at all[5]	Either optimal care or no care at all
Joint-probability-magnitude model	Either optimal care or less than optimal care[6]	Both underinvestment and overinvestment are possible
Separate-probability-magnitude model	Either optimal care (for both probability and magnitude) or no care for magnitude and some care for probability (which may be less or more than optimal care)[7]	Either optimal care (for both probability and magnitude) or no care for magnitude and some care for probability (which may be less or more than optimal care)

5 If care of the injurer influences the magnitude of the losses, the injurer can decide whether or not the harm will exceed his assets (or the threshold introduced by a legal rule). If he decides not to be judgment-proof, (he remains in the "solvent zone") he will take optimal precautions since he internalizes the full harm. However, if he decides to be in the "judgment-proof zone", he will choose no care at all. The intuition is that in this zone taking care does not reduce his liability but would only entail care costs.

6 In the joint-probability-magnitude model, if the injurer chooses to be in the solvent zone, he will take optimal care. He will do this if the threshold is sufficiently high. If he chooses a lower than optimal care level in order to enter the judgment-proof zone, however, he will still take some care measures to lower the accident probability. The lower the threshold, the lower the care level chosen.

7 In the separate-probability-magnitude model, if the injurer chooses optimal magnitude-reducing precautions, he is solvent and will consequently take optimal probability-reducing precautions. He will do this if the threshold is sufficiently high. If he were to choose to take no magnitude-reducing care on the other hand, all probability-reducing care levels are possible and they depend on the threshold. Lower than optimal care is possible due to the familiar judgment-proof argument. However, (more than) optimal care is also possible, because by taking no magnitude-reducing care, liability costs *if* an accident happens are higher than optimal. Taking excessive probability-reducing care can mitigate this by lowering the accident probability.

2.2. The filing and trial/settlement stage

A. Symmetric information and exogenous litigation costs

We did not find a simple model in the literature with symmetric information and exogenous litigation costs that would answer the question of what the effect is of an asset-constrained defendant on the settlement rate and the incentive to file suit. We develop such a model in section 3. In this section, we provide the intuition, which is quite simple. (1) When the defendant is asset-constrained, the settlement frequency increases. Intuitively, the stakes of the dispute decrease with an asset constraint. The effect of relative optimism is known to be more modest for small-stake disputes. (2) The plaintiff's incentive to sue always decreases. The asset constraint simply means that there is less for the plaintiff to collect. One may argue that the increased probability of settlement could increase the incentive to file suit. However, the most advantageous settlement offer the defendant is willing to give with the constraint is still lower than the plaintiff's expected value of trial without the constraint (for formal proof: see the model in section 3).

B. Asymmetric information

Farmer and Pecorino (2015) analyze the case in which some defendants are judgment-proof and the defendant's asset level is not observable to the plaintiff. This is an interesting extension of the literature. While many informational asymmetries may be partially or fully eliminated either through voluntary disclosure or through a discovery process, asymmetric information related to the defendant's assets may be persistent (just like some other asymmetries like AI of preferences, see e.g. Farmer and Pecorino (1994) on risk preferences). First, defendants have an incentive to claim poverty and it is difficult to prove the absence of assets via disclosure. Second, many courts have ruled that the defendant's assets are not subject to discovery.[8] Logically, an asset constraint can increase the probability of trial (decrease the likelihood of settlement) in an asymmetric information model (for a general explanation of why asymmetric information leads to more trials, see the section on settlement). Note that this result is in contrast with the divergent expectations model, in which an asset constraint can increase the settlement frequency.

C. Endogenous litigation expenditures

How does the fact that the defendant is asset-constrained influence the litigation expenditures at trial? And how does this in turn affect the plaintiff's probability of winning in court, and his/her incentive to file suit?

8 For the U.S., see for example *Doak v. Superior Court* 257 Cal. App. 2nd 285 (1968).

We begin with the influence of an asset constraint on the plaintiff's expenditures. These expenditures are always lower with the constraint (Friehe, 2009). Intuitively, the asset constraint lowers the contest prize in comparison with an unconstrained situation. There is simply less for the plaintiff to reap in the case of a victory. The influence on the defendant's behaviour is however more complex. Just like for the plaintiff, the contest prize is smaller, which lowers the marginal benefit from increasing effort, all else being equal. However, there is also a second, expenditure increasing effect for the defendant. The defendant only bears the costs of additional effort when he wins the case. Due to the asset constraint, the higher expenditure due to higher effort is irrelevant from the defendant's perspective when he loses the case. In that instance, he loses all his assets anyway. Put differently, the additional expenditures are externalized to the plaintiff to a certain extent, because the plaintiff will receive less due to the asset constraint. A simple example can illustrate this. Suppose the plaintiff and the defendant have a dispute over an amount of 100. If there were no asset constraint, it would be optimal for each of them to spend 25 at trial, and the probability of a plaintiff victory would be 50 percent. Now we introduce an asset constraint: the defendant only has financial means of 130. Will he be willing to spend an additional amount of 10 so that his probability of success increases from 50 percent to 59 percent? Note that in the absence of a constraint, he will not be willing to do so: the marginal cost of the extra effort is 10, but the expected benefit is only 9 (0.09x100). With a constraint however, the defendant is willing to make the investment. When he makes the additional investment of 10, he would normally lose 135 if the plaintiff were to succeed in court (100 + 25 + 10). But the defendant can lose a maximum of 130. So the difference of 5 (135-130) is subsidized by the plaintiff (with a 50-9= 41 percent probability): he will simply receive less compensation.

These findings have important consequences for the plaintiff's probability of victory at trial. Remember that the asset constraint only has an expenditure decreasing effect on the plaintiff, while it has an expenditure increasing and decreasing effect on the defendant. As a result, the winning probability of the plaintiff decreases with an asset constraint. The asset constraint introduces an asymmetry between the parties, and the defendant will spend relatively more than the plaintiff. The presence of the constraint changes the strategic interaction of the parties to the detriment of the plaintiff. Furthermore, the plaintiff's expected value of trial (and thus his incentive to sue) decreases: although his expenditure decreases (which affects his incentive to sue positively), his probability of success and the amount he will obtain in the case of a win decrease as well (which affects his incentive to sue negatively). The latter, negative effects outweigh the former, positive effect.

3. Modelling issues

3.1. *Ex ante* incentives

In this section, we provide the essence of economic models of the judgment-proofness problem. The table below describes the symbols and notations used in the various models, and the table after that gives an overview of the social cost function and the injurer's private cost function for the eight different models. Under that second table, we describe the algorithm solution to be used for solving the models. We start with a brief repetition of the content of the models. In one-pocket models, precaution and liability expenses are paid out of the same pocket. In two-pocket models, precaution costs do not reduce the total assets available for paying compensation. In probability models, precautions only reduce the probability but not the magnitude of accidental harm. In the magnitude model, care only influences the possible magnitude of the losses. In the joint-probability-magnitude model, adopting a care measure simultaneously influences the accident probability and the magnitude of the losses. In the separate-probability-magnitude model, some care measures influence the accident probability while others influence the magnitude of the losses.

	Symbols
Probability model	x = *the injurer's precaution cost, $x \geq 0$;* $p(x)$ = *probability of an accident, $0<p(x)<1, p'<0, p''>0$;* h = *magnitude of the harm, $h>0$;* t = *the injurer's assets, $t>0$*
Magnitude model	x = *the injurer's precaution cost, $x \geq 0$;* p = *probability of an accident, $0<p<1$;* $h(x)$ = *magnitude of the harm, $h(x)>0$, $h'<0$, $h''>0$;* t = *the injurer's assets, $t>0$*
Joint-probability-magnitude model	x = *the injurer's precaution cost, $x \geq 0$;* $p(x)$ = *probability of an accident, $0<p(x)<1, p'<0, p''>0$;* $h(x)$ = *magnitude of the harm, $h(x)>0$, $h'<0$, $h''>0$;* t = *the injurer's assets, $t>0$*
Separate-probability-magnitude model	s = *the injurer's precaution cost w.r.t. the probability of an accident, $s \geq 0$;* z = *the injurer's precaution cost w.r.t. the magnitude of the harm, $z \geq 0$;* $p(s)$ = *the probability of an accident, $0<p(s)<1, p'<0, p''>0$;* $h(z)$ = *the magnitude of the harm, $h(z)>0$, $h'<0$, $h''>0$;* t = *the injurer's assets, $t>0$*

	Two-pocket models	One-pocket models
Probability model	$S(x)=p(x)h+x$ $J(x)=\begin{cases} p(x)h+x \text{ if } h \le t \\ p(x)t+x \text{ if } h > t \end{cases}$	$S(x)=p(x)h+x$ $J(x)=\begin{cases} p(x)h+x \text{ if } h+x \le t \\ p(x)(t\text{-}x)+x \text{ if } h+x > t \end{cases}$
Magnitude model	$S(x)=ph(x)+x$ $J(x)=\begin{cases} ph(x)+x \text{ if } h(x) \le t \\ pt+x \text{ if } h(x) > t \end{cases}$	$S(x)=ph(x)+x$ $J(x)=\begin{cases} p(x)h+x \text{ if } h(x)+x \le t \\ p(t\text{-}x)+x \text{ if } h(x)+x > t \end{cases}$
Joint-probability-magnitude model	$S(x)=p(x)h(x)+x$ $J(x)=\begin{cases} p(x)h(x)+x \text{ if } h(x) \le t \\ p(x)t+x \text{ if } h(x) > t \end{cases}$	$S(x)=p(x)h(x)+x$ $J(x) = \begin{cases} p(x)h(x)+x \text{ if } h(x)+x \le t \\ p(x)(t\text{-}x)+x \text{ if } h(x)+x > t \end{cases}$
Separate-probability-magnitude model	$S(s,z)=p(s)h(z)+s+z$ $J(s,z)=\begin{cases} p(s)h(z)+s+z \text{ if } h(z) \le t \\ p(s)t+s+z \text{ if } h(z) > t \end{cases}$	$S(s,z)=p(s)h(z)+s+z$ $J(s,z) = \begin{cases} p(s)h(z)+s+z \text{ if } h(z)+s+z \le t \\ p(s)(t\text{-}s\text{-}z)+s+z \text{ if } h(z)+s+z > t \end{cases}$

These models can be solved with the following algorithm solution:

1. find the levels of precaution which minimize the total expenditures for a solvent and an insolvent injurer;
2. compare the total expenditures and check whether it is privately optimal to be solvent or insolvent;
3. verify the validity of the solution (that the injurer is actually solvent/insolvent at the chosen levels of precaution.

3.2. Suit and settlement: A simple divergent expectations framework

We could not find a simple divergent expectations framework in the literature which deals with two simple but important questions: (1) how does an asset constraint influ-

ence the parties' incentive to settle/litigate, and (2) how does it affect the plaintiff's incentive to sue. We develop a simple model below.

We use the following notations:
J: the judgment at trial;
A: the assets of the defendant;
p_p: the plaintiff's estimate of victory at trial;
p_d: the defendant's estimate of a plaintiff victory at trial;
C_p and C_d: the plaintiff's and the defendant's trial costs respectively.

We assume that when the defendant is asset-constrained, her lawyer is paid in full before any assets are assigned to the plaintiff. We further assume that the defendant's assets are not sufficient to pay the sum of judgment at trial and the litigation costs, but are sufficient to go to trial. More formally, we assume that $C_d < A < J + C_d$. We further assume that each party pays their own trial costs.

With the asset constraint, the plaintiff's expected value of trial equals:

$$p_p(A - C_d) - C_p$$

The defendant's expected loss equals:

$$p_d A + (1-p_d)C_d$$

The parties will settle the dispute if:

$$p_d A + (1-p_d)C_d > p_p(A - C_d) - C_p$$

which can be rewritten as:

$$C_p + C_d > (p_p - p_d)(A - C_d) \quad (1)$$

We can now compare this condition with the condition for settlement in the absence of an asset constraint:

$$C_p + C_d > (p_p - p_d)J \quad (2)$$

Clearly, where there is a constraint, settlement is more likely given that the right-hand side of condition (1) is smaller than right-hand side of condition (2). The intuition is simple: when there is a constraint, the stakes of a dispute decrease.

So for cases in which $(p_p - p_d)J > C_p + C_d > (p_p - p_d)(A - C_d)$, cases settle with the constraint but not without it. Note that increasing trial costs of the defendant have two positive effects on the settlement range when there is a constraint: (1) the classic, trial cost increasing effect, and (2) a reduction in what the plaintiff can recover if he wins the case, because the defendant first has to pay his lawyer and does not have enough financial means to pay the full award and the trial costs. Unlike the situation where there is no constraint, when there is such a constraint the settlement frequency should increase more when the defendant's trial costs increase by a certain amount than when the plaintiff's trial costs increase by that amount.

An interesting question is whether the asset constraint of the defendant could ever be positive for the plaintiff? In more concrete terms, could the fact that the parties settle more often when there is a constraint than when there is no constraint have a positive effect on the plaintiff's incentive to sue? The answer is no. Although the parties will more often avoid the costs of a trial, even the most advantageous settlement offer the defendant is willing to give when there is a constraint is still lower than the plaintiff's expected value of trial without such a constraint. This can be proven as follows:

We want to prove that the most advantageous settlement amount when there is a constraint is still lower than the plaintiff's expected value at trial without such a constraint, thus that: $p_pJ - C_p > p_dA + (1-p_d)C_d$. We can rewrite this as: $p_pJ - P_dA > C_p + (1-p_d)C_d$.

Given that $A < J + C_d$, it follows that:

$p_pJ - P_dA > p_pJ - P_d(J + C_d)$, which can be rewritten as:

$p_pJ - P_dA > (p_p - p_d)J - P_dC_d$

From the fact that $C_p + C_d > (p_p - p_d)J$, it follows that:

$p_pJ - P_dA > C_p + C_d - P_dC_d$, thus that:

$p_pJ - P_dA > C_p + (1 - p_d)C_d$

3.3. Introducing asymmetric information

For an asymmetric information model examining a situation in which some defendants are judgment-proof and the defendant's asset level is not observable to the plaintiff, we would refer to Farmer and Pecorino (2015).

3.4. Introducing endogenous litigation expenditures

In this section we provide a simple model to analyze the consequences of an asset constraint on the litigation expenditures of the parties, the probability of a plaintiff victory, and the parties' pay-offs at trial.[9] Let X and Y represent the litigation effort of the plaintiff and the defendant respectively. The parties have a dispute over an amount J. The unit costs of litigation are Cp for the plaintiff and Cd for the defendant. The defendant has assets A. The amount at stake is slightly larger than the defendant's assets. The probability of a plaintiff victory depends on the expenditures of the parties as follows: $p(X,Y) = \frac{X}{X+Y}$.

The plaintiff's wishes to maximize his/her expected value:

$$EVpl = p\,(X,Y)(A - Y) - CpX$$

The defendant wishes to minimize his/her expected loss:

$$ELd = p\,(X,Y)A + (1 - P(X,Y))CdY$$

The model can be easily solved by taking the first order conditions.

9 We expand a model by Friehe (2009) by incorporating a general amount at stake (which is set equal to 1 in Friehe's model, and by introducing unit costs of litigation, which are absent in the Friehe's model).

4. Empirical issues

1. Empirical research in the Netherlands (Eshuis, 2009) finds that the number of cases in which the defendant complies with what has been decided depends strongly on whether the case ended with a court verdict or a settlement. When the case ended with settlement, 85 percent of defendants fully complied with the settlement terms within three years. When the case ended with a (non default) court verdict, this figure decreases to 74 percent. The fraction of cases which ended with a settlement and in which the defendant did absolutely nothing during three years was 5 percent, while this figure was 10 percent for non default verdicts.
   It is dangerous to assume a causal link between settlement and a higher compliance rate for at least two reasons. (1) There could be a selection effect. The elements which may lead a case to settle, may also influence the chance that the settlement is respected. One example concerns the amount at stake. Economic analysis shows that cases with relatively low amounts at stake are more often settled than higher-stake cases. At the same time, it is more likely that cases dealing with low amounts will more often be respected, because the defendant is able to pay the low amount. (2) Specific aspects of procedural law may influence the finding. In the Netherlands, during non default procedures, *conditional* settlements are often agreed. The settlement only becomes a "formal settlement" if the agreed action has been fulfilled. If this does not happen, the case will still end with a court verdict. So the higher "success rate" for settlements is partly due to the fact than an agreement becomes a settlement only if the agreement has been respected.
   For default court cases, only 31 percent of defendants complied fully with the verdict within three years. The fraction of cases which ended with a default verdict and in which the defendant did absolutely nothing during three years was 42 percent, Once again, there is a selection effect at play here. Parties who know that they will not be able to pay a verdict will often not defend their case in court.
2. The empirical study by Eshuis (2009) finds that the major reason why court verdicts and settlement agreements are not respected is a lack of funds. Logically, the author finds that as the amount the defendant needs to pay increases, the probability that the verdict or agreement will be respected declines. However, this finding is only statistically significant for amounts of 15,000 euros and higher. The reason why the relationship is unclear for smaller amounts is *inter alia* caused by the presence of default cases (the probability of default is larger for smaller amounts, and default verdicts are complied with less frequently).

3. Eshuis (2009) finds no correlation between the duration of the trial and the probability of compliance with the verdict (both for non default and default verdicts).
4. When the defendant may not be able to pay the (full) judgment, we have seen that this reduces the plaintiff's incentive to sue. Logically, this will affect the weakest claims first (these may become negative expected value suits rather quickly). So conversely, we should expect that lower average quality suits are brought more often when defendants are richer. Gross and Syverud (1991) find that in cases with high potential damages and rich defendants, the plaintiff success rate at trial decreases. This may be evidence that lower quality suits are brought.

7

Stage VII. Integrated model

1. Introduction

Here we illustrate the potential for examining the effects of policy parameters on key elements of civil litigation. We use a divergent expectations model where analysis of the litigation (should it take place) involves examining whether there is a bargaining window in which settlement can take place and assuming that the parties act accordingly if such exists. While the model overlooks a number of important features of litigation (such as asymmetric information and dynamics), it focuses in a tractable way on the central question of what conditions are *necessary* for settlement to occur. Its tractability also allows us to examine efficiency by endogenizing the initial care decision of one of the parties (the prospective defendant), and therefore illustrate how the model can be used for welfare analysis.[1]

During the analysis, we perform a sample analytical comparative static (looking at the effects of changes in the plaintiff's litigation costs on care and welfare), and illustrate this by programming the model in MATLAB.

2. The model

We model the following situation:

- Stage 1: The prospective defendant chooses a level of care. More care lowers the probability of an accident (at a decreasing rate).

1 Our use of "care" and "accidents" (in Stage 2 below) implies a personal injury context but, in fact, we could think of this "care" as being the effort taken by a contracting party to limit the possibilities of breach of contract with the subsequent litigation relating to a contractual dispute. This is similar to the approach taken in Cooter, R. and Ulen, T., *Law and Economics*, 6th edition, 2011, Pearson International Publishers.

- Stage 2: If an accident occurs, the harmed party decides whether to file a claim. This decision is based on his/her view of the claim's strengths, the likely damages and the costs associated with obtaining these damages.
- Stage 3: If a case is filed, litigation ensues and the parties seek to settle or take the case to trial. This decision is taken on the basis of both parties' views of the claim's strengths, the likely damages and the costs associated with obtaining these damages, which reflect their views of the likely court outcome.

We make the following assumptions about the parameters of the litigation:

- Defendant (D) and plaintiff (P);
- Each has beliefs about P's chances at trial, $0 < q_i < 1$, $i = p,d$;
- They also have views about the likely damages at trial ($x_i, i = p,d$) and face trial costs ($k_i, i = p,d$);
- D's initial care is c and this has a unit cost of $w > 0$;
- The probability of an accident, given care, is e^{-c};
- U.S. cost rules apply (i.e. each of the parties bears his/her own costs);
- For simplicity, the model assumes that q_d is fixed while $q_d \sim U[0,1]$.

The last two assumptions can be amended (with varying degrees of straightforwardness); we illustrate this below for the case of cost-shifting rules.

3. Solving the model

The model is solved by backwards induction, starting with Stage 3.

3.1. Stage 3: Settlement or trial?

The expected value of trial for each party is:[2]

$$EVT_p = q_p x_p - kp \qquad (1)$$

$$EVT_d = q_d x_d + k_d \qquad (2)$$

A necessary condition for trial is $EVT_p > EVT_d$ or, equivalently (since EVT_p is monotonic in q_p), $q_p > \hat{q}_p$, where the threshold is defined by $EVT_p(\hat{q}_p) \equiv EVT_d$:

2 We use the generic term "expected value" for both parties. For the plaintiff, this value can be positive or negative. For the defendant, the value is always negative. Therefore, it is often termed "expected loss" in the literature.

$$\hat{q}_p \equiv \frac{k_p + k_d}{x} + \frac{x_d}{x} q_d \quad (3)$$

As q_p is uniformly distributed on [0,1], the unconditional probability of trial is $1 - \hat{q}_p$. If settlement occurs (i.e. if $q_p < \hat{q}_p$), we assume that the terms involve a weighted average of the two parties' threat points; i.e. if $0 < \beta < 1$ is the weight on EVT_p we have:

$$\begin{aligned} S &= \beta EVT_p + (1-\beta) EVT_d \\ &= EVT_d - \beta x_p (\hat{q}_p - q_p) \end{aligned} \quad (4)$$

3.2. Stage 2: P's decision to file a case

Now consider the decision to file a case. P needs $EVT_p \geq 0$. This allows us to define $\tilde{q}_p$ from $EVT(\tilde{q}_p) \equiv 0$; i.e.:

$$\tilde{q}_p \equiv \frac{k_p}{x_p} \quad (5)$$

It is clear that $\hat{q}_p > \tilde{q}_p$ and the probability of filing is $1 - \tilde{q}_p$ (given the uniform distribution of q_p).

3.3. Stage 1: D's care decision

In order to endogenize D's initial care decision, we consider D's problem:

$$\min_c wc + e^{-c} H_d(q_p, q_d, x_p, x_d, k_p, k_d, \beta) \quad (6)$$

where $H_d(.)$ is the *ex ante* expected costs for D when she is choosing a level of care c and an accident happens. It reflects both settlement costs, trial costs and damages to be paid. Call the solution c^*. Using the short-hand $\mathcal{S}$ for "settle", $\mathcal{T}$ for "trial" and $\mathcal{F}$ for "file", we can evaluate this as follows:

$$H_d(\cdot) = (1 - \tilde{q}_p)[\Pr(\mathcal{S}|\mathcal{F})\, S + \Pr(\mathcal{T}|\mathcal{F})\, (q_d x_d + k_d)]$$

where we have assumed that D will expect to pay her assessment of damages (x_d) if the case makes it to trial.

For later comparative static purposes, however, it is helpful to takes things slightly further. We have:

$$\Pr(\mathcal{S}|\mathcal{F}) = \frac{\hat{q}_p - \tilde{q}_p}{1 - \tilde{q}_p} = \frac{EVT_d}{x_p - k_p} \tag{8}$$

$$\Pr(\mathcal{T}|\mathcal{F}) = \frac{1 - \hat{q}_p}{1 - \tilde{q}_p} = \frac{x_p - k_p - EVT_d}{x_p - k_p} = 1 - \Pr(\mathcal{S}|\mathcal{F}) \tag{9}$$

Since $1 - \tilde{q}_p = \frac{x_p - k_p}{x_p}$, we can substitute this, (8) and (9) into (2) to yield:

$$\begin{aligned} H_d(\cdot) &= \frac{x_p - k_p}{x_p}\left\{\left(\frac{EVT_d}{x_p - k_p}\right)\left[\beta EVT_p + (1-\beta)EVT_d\right] \right. \\ &\qquad \left. + \left(\frac{x_p - k_p - EVT_d}{x_p - k_p}\right)EVT_d\right\} \\ &= \frac{EVTd}{x_p}\left[\beta\left(EVT_p - EVT_d\right) + EVT_d + x_p - k_p - EVT_d\right] \\ &= \frac{EVT_d}{x_p}\left[\beta\left(EVT_p - EVT_d\right) + x_p - k_p\right] \end{aligned} \tag{10}$$

Returning to D's care decision, it is readily shown that the solution to (6) is given by:

$$c^*(\cdot) = \begin{cases} \ln\left(\frac{H_d}{w}\right) & \text{if } H_d > w \\ 0 & \text{if } H_d < w \end{cases} \tag{11}$$

4. Welfare

Welfare analysis can be conducted using the social loss function:

$$L = wc^*(\cdot) + e^{-c^*(\cdot)}\left[\bar{x} + \left(1 - \tilde{q}_p\right)\Pr(\mathcal{T}|\mathcal{F})\left(k_p + k_d\right)\right]$$

$$= wc^*(\cdot) + e^{-c^*(\cdot)}\left[\bar{x} + (x_p - k_p - EVT_d)(k_p + k_d)\right]$$

where $\bar{x} \equiv \frac{x_p + x_d}{2}$.

Note that when part of the legal costs are publicly financed, these costs will count entirely in the social loss function, but will need to be reduced in the private decisions of the litigants.

5. A simple comparative static exercise: changing k_p

Using the above, we can perform comparative statics. To illustrate this, consider the effects of an increase in k_p on $c^*(\cdot)$. Assuming $H_d > w$, we know from (11) that:

$$\frac{\partial c^*}{\partial k_p} = \frac{w}{H_d}\frac{\partial H_d}{\partial k_p}\frac{1}{w} = \frac{1}{H_d}\frac{\partial H_d}{\partial k_p} \qquad (12)$$

Notice that $k_p \frac{\partial c^*}{\partial k_p}$ is the elasticity of D's expected costs with respect to P's costs; a result that could be useful in future empirical work.) From (10) and (1), we have:

$$\frac{\partial H_d}{\partial k_p} = -\frac{EVT_d}{x_p}(1+\beta) < 0 \qquad (13)$$

Hence, from (11), we have $\frac{\partial c^*}{\partial k_p} < 0$. The explanation is that an increase in k_p makes trial less likely ($1 - \hat{q}_p$ falls) and also lowers the settlement amount.

The effects of k_p are:

$$\frac{\partial L}{\partial k_p} = \underbrace{w\frac{\partial c^*}{\partial k_p}}_{(A)} - \underbrace{e^{-c^*(\cdot)}\frac{\partial c^*}{\partial k_p}\left[\bar{x} + \left(x_p - k_p - EVT_d\right)\left(k_p + k_d\right)\right]}_{(B)} + \underbrace{e^{-c^*(\cdot)}\left[x_p - 2k_p - k_d - EVT_d\right]}_{(C)} \qquad (14)$$

In order to sign (14), we know that $(A) < 0$ (from (12)), and that $(B) < 0$ (since the square bracket must be positive if $1 - \hat{q}_p > 0$), while $(C) \gtrless 0$. Hence $\frac{\partial L}{\partial k_p} \gtrless 0$:

raising P's costs may, or may not, increase social loss. The reason for this is that the higher trial costs reduce care costs (and hence reduce social losses – see (A)), but this makes an accident more likely (which raises expected losses by rendering trial costs more likely – see (B)). In addition, should an accident arise, the higher trial costs make trial more costly but they also make it less likely as settling is more attractive to P –

this is why $(C) \gtreqless 0$. It is apparent that $(C) > 0$ if x_p is large enough (so the square bracket in (C) is positive): i.e. if P thinks trial is sufficiently profitable that the rise in k_p has too small a deterrent effect on the chances of trial. In this case, the increased costs of trial dominate and this part of social loss rises with P's trial costs.

To illustrate these results, consider the following parameter values:

- $q_p = 0.8$
- $x_p = 1{,}000$
- $k_p \in \{100,101, \ldots, 200\}$
- $q_d = 0.7$
- $x_d = 800$
- $k_d = 100$
- $\beta = 0.5$
- $w = 1$

Using MATLAB to program the model produces the following figures. In Figure 1, we see the probabilities of P not filing the case (blue) and the probability of settlement (red). The results are consistent with our observation that $\hat{q}_p > \tilde{q}_p$. As P's trial costs increase from 100 to 200, her incentive not to file the case rises (from 0.1 to 0.2) and the probability of settlement rises too. In Figure 2, the settlement sum (S) drops from 580 to 530 as k_p rises. As predicted above, the combination of reduced chances of litigation and trial, along with a slower settlement amount, make D relatively unconcerned about litigation and care (c^*) falls (Figure 3), making an accident more likely. Finally, in Figure 4, the combination of more accidents and costlier trials increases the social loss from litigation; i.e. L rises. Thus, in this case (relative to the analytical result above), these cost effects outweigh the effects of filing being less likely – an effect which, as we have seen, is relatively small (the increase from 0.1 to 0.2 described above).

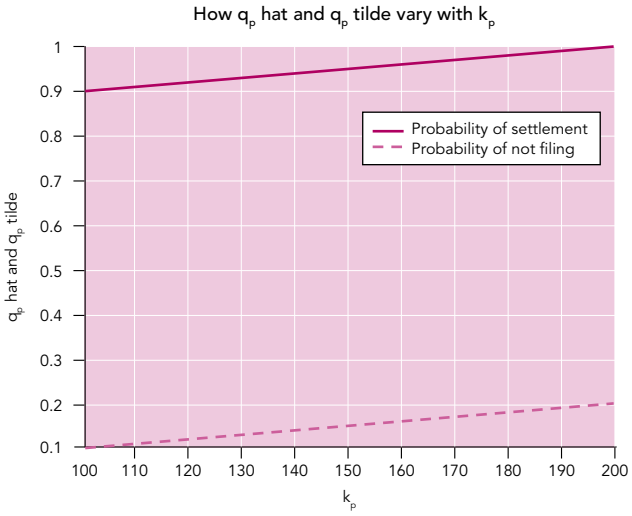


Figure 1

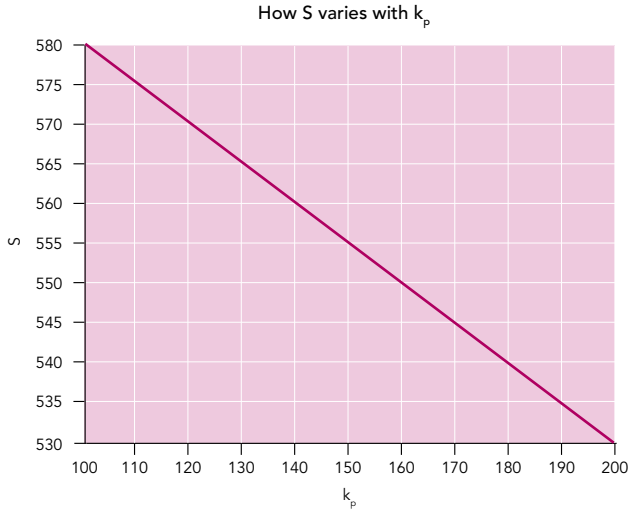


Figure 2

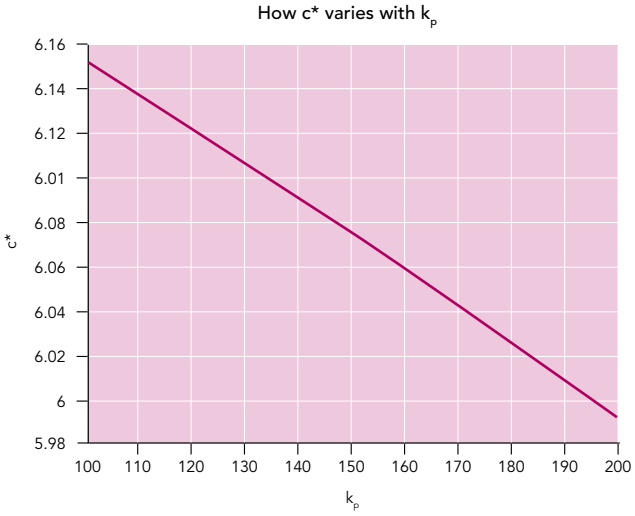


Figure 3

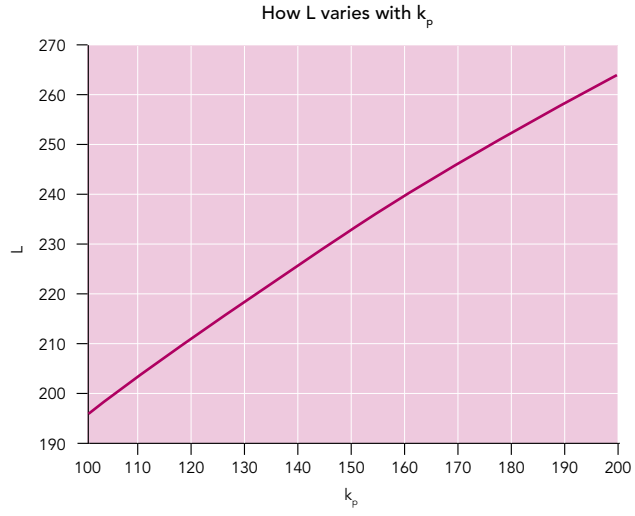


Figure 4

6. Developing the model

6.1. Moving to U.K. cost rules

A number of extensions/alterations to the current model are possible. Without providing equivalent detail to that above, we illustrate one possibility here: a move towards cost rules where the loser pays a fraction of the winner's costs (i.e. the "U.K. rule").

To model this, let $\tau_i = t[q_i k_p - (1 - q_i)k_d]$, $i = p,d$, be the net transfer expected by P from D at the end of the case (or, when $i = d$, by D from P) where $t \in [0,1]$ is the fraction of costs transferred according to the cost-shifting rule. For instance, $t = 0$ will return the foregoing ("U.S. rule") analysis, and $t = 1$ will be a full "U.K. rule", so we are effectively presenting a more general model than the one before. We can then re-write (1) and (2) as follows:

$$EVT_p^t = q_p x_p - k_p + \tau_p \qquad (15)$$

$$EVT_d^t = q_d x_d + k_d + \tau_d \qquad (16)$$

We can proceed as before, to find threshold values of q - p for settlement and filing:

$$\hat{q}_p^t = \frac{k_p + tk_d + q_d x_d + k_d + \tau_d}{x_p + k_p + tk_d}, \qquad \tilde{q}_p^t = \frac{k_p + tk_d}{x_p + t(k_p + k_d)} \qquad (17)$$

Again, notice that setting $t = 0$ returns (h), (i), (p) and (e). Analysis can follow the earlier procedure, with an additional comparative static on the "amount" of cost-shifting, t.

6.2. Time, filing costs and correlated beliefs

Litigation takes place over time (and trial can be long or short – with potential cost implications), it can be costly to file claims, and it seems likely that the parties' beliefs about case strength will be correlated. We now indicate how these variations can be added to the model.

- Filing costs: We now distinguish between the plaintiff's costs for filing/starting a case (k_p^F) and both parties' trial costs (k_p^T and k_d^T).[3]
- Correlated beliefs: A simple way to deal with these are to specify $q_d = \alpha q_p$. In principle, $\alpha \gtreqless 1$. When $\alpha<1$, the plaintiff is relatively optimistic. In the opposite case, the plaintiff is relatively pessimistic. Of course, there are bounds on α, in the sense that q_p and q_d can never be larger than 1.
- Time: We can handle time in three ways:
  - Discounting: The parties can discount the future, with outcomes further ahead in time receiving smaller weights. This may be a natural way to capture the effects of time since we (implicitly) are imagining a two-stage process with filing, then trial. Thus, we could assume that the parties discount the future with weight $\delta \in (0,1]$.
  - Trial costs: Longer trials are generally costlier and, as such, examining the effects of higher k_p^T and k_d^T can help to capture this. In fact, this has already been covered (see section 5).
  - Lost investment income: we could divide the settlement stage into several smaller stages, and let the settlement costs of a party reflect the lost investment income associated with the time that elapses when settlement is not reached during that substage.

Given the variety of amendments to be introduced above, we make one simplification: $x_p = x_d \equiv x$. This does not seem too dramatic since at least some of the damages at stake in a case can be verified from pay slips, invoices, medical bills, etc.

The equations in sections 3 and 4 are amended as follows as a result of the above changes:

$$EVT_p = q_p x - k_p^T \tag{18}$$

$$EVT_d = q_d x + k_d^T = \alpha q_p x - k_p^T \tag{19}$$

$$\hat{q}_p \equiv \frac{k_p^T + k_d^T}{(1-\alpha)x} \tag{20}$$

3 Note that the plaintiff's costs for filing/starting a case could not only represent court fees, but also the costs associated with an informal assertion of a claim.

$$S = \delta\left[EVT_d - \beta\left(\hat{q}_p - q_p\right)\right] \tag{21}$$

$$\tilde{q}_p \equiv \frac{k_p^F + \frac{k_p^T}{\delta}}{x} \tag{22}$$

$$\min_c wc + e^{-c} H_d(q_p, x, k_p^T, k_d^T, \beta, \delta, \alpha) \tag{23}$$

where:

$$H_d(\cdot) = \left(\hat{q}_p - \tilde{q}_p\right)S + \delta(1 - \hat{q}_p)(\alpha q_p x + k_d^T) \tag{24}$$

$$c^*(\cdot) = \begin{cases} \ln\left(\frac{H_d}{w}\right) & \text{if } H_d > w \\ 0 & \text{if } H_d < w \end{cases} \tag{25}$$

and:

$$L = wc^*(\cdot) + e^{-c^*(\cdot)}\left[\delta x + (1 - \tilde{q}_p)(k_p^F + \delta \Pr(\mathcal{T}|\mathcal{F})(k_p^T + k_d^T))\right] \tag{26}$$

Now consider several effects.

- The roles of k_p^T (proxying extended trials) and k_p^F are similar – but not entirely identical – in nature to the effects of k_p in section 5: i.e. the final welfare effects are ambiguous because matters that make filing or trial less attractive to P make care less attractive to D and, therefore, this may make trials *more* likely (through a higher accident volume). The higher fees reduce care costs (and hence reduce social losses), but this makes an accident more likely (which raises expected losses by rendering trial costs more likely). In addition, should an accident arise, the higher costs for the plaintiff make trial less likely as settling is more attractive to the plaintiff. One difference however with increasing trial costs is that increasing fees do not directly increase the social costs of trial, since the fees are mere transfers. Decreasing filing fees (or introducing subsidies) may be socially beneficial especially for areas where the following conditions are cumulatively met: (1) there is currently substantial underdeterrence, (2) potential injurers are likely to increase their care level in light of an increasing number of plaintiffs able to sue, and (3) the settlement frequency is (or would be) relatively large.

- Higher values of α mean that the parties' beliefs are more closely correlated. This makes trial harder to reach (which is favourable for welfare) but also runs the risk of increasing the settlement offer that D must pay. As a result, D's *ex ante* care level behaves ambiguously as α increases and welfare may (in principle) fall – indeed, as the care level falls, so filing becomes more likely (which is costly for welfare).
- A higher value of δ means that the outcome of the case is "nearer" (e.g. as a result of faster trials). This makes filing more likely and increases D's *ex ante* costs (and hence D's *ex ante* care). From a welfare perspective, increasing filing prospects is a negative outcome, but bringing the future "closer" also favours settlement over trial ($\hat{q}_p - \tilde{q}_p$ rises) and this is helpful because it makes k_p^T and k_d^T less likely to be expended; but the higher δ makes these larger in present value terms – which is damaging for welfare. Thus, the overall effect on welfare is ambiguous, though it could be argued that promoting settlement is often regarded as a valuable policy outcome.

6.3. Endogenous litigation costs

The model can easily be extended to incorporate litigation expenditures:

a. Given that empirical research finds that litigation expenditures increase with the amount at stake, and that theoretical models and experimental research establish that the parties roughly spend an equal amount, we could adapt the model by setting kp and kd equal to a fraction of the amount at stake (e.g. αJ, with α the fraction and J the stakes).
b. Note that this formula could be made more complex in several ways: (i) by introducing fixed costs as well (then k_p and k_d would equal some constant plus αJ); (ii) for those cases in which the parties typically spend a different amount: one could set kp equal to αpJ and kd equal to αdJ; (iii) by introducing the more complex formulas we have discussed in the text, in which the expenditures of the parties depend not only on the amount at stake, but also on the inherent merits of the case and the unit costs of the parties. The advantage of this approach would be that it could lead to more refined results. A disadvantage would be the greater complexity, and the need to explicitly link each possible care decision of the potential injurer with a certain inherent quality of the case.

6.4. Activity levels

As it is currently conceived, the model does not shed light on activity levels (e.g. the surplus of contractual transactions which did not take place; the costs associated with deterring efficient activities in a non-contractual setting). To include activity levels, the model can be extended as follows: (a) we need to introduce a distribution for the values of contractual transactions/extra-contractual activities (e.g. the value of a certain type of contract could be uniformly distributed between 0 and 100,000 euros); (b) by using backward induction, through which the parties compare their share of the value with the expected pay-offs from later stages (care, filing, settlement, trial), we can shed light on which transactions will take place and which transactions will not (formally, there could be a cut-off value in the distribution of values). In this way, the social losses of foregone transactions can be incorporated in the model.

VIII. Conclusion

The general analysis and the model we have developed show the importance of taking into account all the litigation stages in order to describe the decisions of the parties and the consequences for welfare. We focus here on (1) filing costs, (2) uncertainty, and (3) time. (1) Increasing or decreasing filing costs has ambiguous welfare effects because measures that make filing or trial less interesting for a plaintiff may make efficient *ex ante* behaviour (e.g. fulfilling contractual duties, taking care) less attractive to the defendant, which can increase the number of trials. (2) Our general framework has shown the complex consequences of uncertainty. At the *ex ante* level, both overprecaution and underprecaution are possible. Much depends on whether rules or standards are used, the type of legal rule deployed, the accuracy of compensation, the parties' risk aversion (and, thus, the payment contracts they have with their lawyers), etc. And uncertainty may also have some beneficial consequences due to filtering effects. At the filing level, uncertainty may decrease the number of filings due to risk costs and lower settlement amounts. At the settlement stage, uncertainty has two diametrically opposed effects. Uncertainty may increase the settlement range due to risk costs, but it may also decrease the settlement range due to an increase in the possibility that the expectations of the litigants will diverge. Furthermore, uncertainty may prolong settlement negotiations and has ambiguous effects on the settlement amount. When the bargaining power of the plaintiff is relatively small, uncertainty most likely leads to lower settlement amounts. In our model, less uncertainty means the parties' beliefs become more correlated. This makes trials less likely, which is favourable for welfare, but also creates the risk of increasing the settlement offer that the defendant must pay. As a consequence, the *ex ante* incentives of the defendant are ambiguous as uncertainty drops. Welfare may either increase or decrease. (3) Trial delay has many consequences. Legal delay threatens the effective operation of the judicial system, imposes additional burdens (monetary and psychological) on litigants, denies proper access to justice and increases costs for the courts. While the conven-

tional wisdom that delay in case court disposition time creates a set of problems for the judicial system is certainly well founded, delay also generates less obvious and intuitive – but equally relevant – benefits for the system. As a first consideration, in most cases delaying the decision on whether to alter the status quo (e.g. avoiding prison, deciding on children's custody, and the payment of a claimed debt) is beneficial for one of the parties involved in the trial. Moreover, delay guarantees zealous representation and rigorous evidence examination, if only due to the opportunity for in-depth investigation of the available evidence. Less intuitive beneficial effects of delay are connected to its rationing function (the so called "Gravelle's hypothesis"). According to some scholars, court delays are an efficient rationing system that reduce the total number of cases litigated, specifically deterring plaintiffs with weaker cases from filing suit. In a related way, delay may also improve the accuracy of trial outcomes, with consequent benefits for welfare (see Fenn and Rickman, 2013). In our model, the overall effect of faster trials on welfare is ambiguous.

More empirical research is absolutely necessary for all three domains:

(1) Filing costs

Our model shows that raising the plaintiff's costs through increased filing fees may, or may not, increase social losses. The reason for this is that the higher costs for the plaintiff reduce care costs (and hence reduce social losses), but this makes an accident more likely (which raises expected losses by rendering trial costs more likely). In addition, should an accident arise, the higher costs for the plaintiff make settling more attractive, which lowers social costs. In some instances, the positive effects may outweigh the negative effect, and in other cases the opposite may be true.

Evidence from other countries suggests that reforming the level of court fees can be a challenging task and can lead to unintended effects. For instance, the increase in court fees trialled in 2013 by the U.K. Ministry of Justice caused a reduction of 70 percent in the number of cases brought to court. The most common types of claim saw an unprecedented contraction: Working Time Directive, down 78 percent; unauthorized deductions from wages, down 56 percent; unfair dismissal, down 72 percent; equal pay, down 58 percent; breach of contract, down 75 percent, and sex discrimination, down 68 percent. Most worryingly, the Discrimination Law Association reported that reduced access to tribunals had fallen disproportionately on women and traditionally disadvantaged groups.

Empirical studies are necessary to provide sound evidence-based policy suggestions on how to balance access to justice and the level of court fees in different jurisdictions. There will be variation across different types of cases and across jurisdictions. Clever research design should be employed to estimate the effects on the level of care of potential offenders. Empirical estimates of fee elasticities of the number of court cases can help shed light on several elements of our model.

(2) Uncertainty

a. In order to stack future models with reliable estimates of uncertainty, the Priest-Klein model could offer some opportunities to measure legal uncertainty for some categories of legal disputes. More specifically, the Priest-Klein model can be used "in reverse" to shed some light on the most plausible combinations of the prediction errors of the parties and the applicable decision standard. Applied in the "classic way", the Priest-Klein model predicts the trial rate and the plaintiff win rate from the following information regarding a certain category of disputes: (a) the amount at stake, (b) the trial costs and settlement costs, (c) a decision standard (e.g. preponderance of the evidence) and (d) a standard deviation of the prediction error.

This can be applied in reverse. If one knows the value of the trial and settlement costs for some category of dispute, and the fraction of litigated cases and the plaintiff win rate within these cases, one can deduct the most likely combinations of prediction error and decision standard. We would refer to the section on uncertainty (and more precisely 4.1.1. (7) of that section) for a deeper analysis.

b. Future models may also benefit substantially by incorporating realistic risk costs. With respect to the United States, Viscusi (1988) found that risk aversion has a relatively modest effect on the decision to drop a lawsuit – though it is possible that attitudes to risk may have filtered cases at the filing stage. This result may also depend strongly on the specific context. In the United States, plaintiffs in tort cases (including product liability claims) pay their lawyer based on a contingency fee arrangement. This greatly limits the risk exposure of the plaintiff by transferring a portion of the risk to the lawyer, who can diversify risk by handling a portfolio of many claims. A Dutch study on the effect of risk aversion could provide different and valuable insights for future models – both to inform Dutch policy and

more generally to promote a better understanding of litigation models. If an appropriate Dutch database exists, the method employed by Viscusi could be used: for a certain area of tort cases (e.g. product liability), a division should be made between types of case that differ as to (a) the level of average damages, and (b) the variance of the damages. The risk premium can be modelled as a function of the variance.

(3) Time

Existing empirical research shows that the benefits of decreasing the duration of litigation may be quite large for some types of disputes. More research is necessary to refine the lost investment income method. For example, since the methodology is difficult to apply to non-monetary claims, one possibility would be to use surveys in which litigants are asked to reveal the monetary value of the interest at stake. However, such surveys may obviously not produce reliable results. One proxy that we suggest may be interesting are the parties' investments in the case (most importantly lawyer fees). Research shows that the parties' litigation costs increase, although not proportionally, with the financial value of the case. If for example one knows that for cases with a known monetary value of 100,000, the parties spend X, then one could impute a value of 100,000 to cases involving non-monetary claims in which parties spend X. Although this methodology is certainly not perfect, it may be an improvement compared with the survey method.

Furthermore, future research may combine the lost investment method and the willingness to pay method. The willingness to pay can shed some light on whether there are any costs at all associated with delay. The lost investment method, with modifications, can provide more guidance on the magnitude of the losses.

An important variable when analyzing litigation (especially over time – see Fenn and Rickman, 1999) is the parties' estimates of their case strength. This, in fact, would influence all three of the issues we have highlighted in this section – the likelihood of filing, the riskiness of the case and the time it will take to resolve (since case strength will generally be private information). A study enabling collection of these data would allow genuinely innovative analysis of Dutch policies in these areas, as well as the opportunity to produce studies at the forefront of empirical law and economics.

Bibliography

Administrative Office of the US Courts, (1980). Statistical codes for civil reports submitted by clerks of court, United States District Courts.

Alarie, B. R., & Green, A. (2007). The reasonable justice: an empirical analysis of Frank Iacobucci's career on the Supreme Court of Canada. *University of Toronto Law Journal*, *57*(2), 195-226.

Anderson, J. M., Kling, J. R., & Stith, K. (1999). Measuring interjudge sentencing disparity: before and after the federal sentencing guidelines. *The Journal of Law and Economics*, *42*(1), 271-308.

Ashenfelter, O., & Dahl, G. B. (2005). Strategic bargaining behavior, self-serving biases, and the role of expert agents: an empirical study of final-offer arbitration (National Bureau of Economic Research). No. 11189.

Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics*, *4*(6), 1236-1239.

Austin, W., & Williams III, T. A. (1977). A survey of judges' responses to simulated legal cases: Research note on sentencing disparity. *Journal of Criminal Law & Criminology*, *68*(2), 306-310.

Babcock, L., & Loewenstein, G. (1997). Explaining bargaining impasse: The role of self-serving biases. *The Journal of Economic Perspectives*, *11*(1), 109-126.

Banks, J. S., & Sobel, J. (1987). Equilibrium selection in signaling games. *Econometrica: Journal of the Econometric Society, 55*(3), 647-661.

Bailey, M. A. (2007). Comparable preference estimates across time and institutions for the court, congress, and presidency. *American Journal of Political Science*, *51(3),* 433-448.

Bar-Gill, O., & Ben-Shahar, O. (2005). Credible coercion. *Texas Law Review*, *83*, 717.

Beard, T. R. (1990). Bankruptcy and care choice. *The RAND Journal of Economics*, *21*(4), 626-634.

Bebchuk, L. A. (1984). Litigation and settlement under imperfect information. *The RAND Journal of Economics*, *15*(3), 404-415.

Bebchuk, L. A. (1988). Suing solely to extract a settlement offer. *The Journal of Legal Studies*, *17*(2), 437-450.

Bebchuk, L. A. (1996). A new theory concerning the credibility and success of threats to sue. *The Journal of Legal Studies*, *25*(1), 1-25.

Bebchuk, L. A., (1998). Suits with negative expected value. The New Palgrave Dictionary of Economics and the Law, 3, (Peter Newman, ed.) London. Palgrave MacMillan.

Beenstock, M., & Haitovsky, Y. (2004). Does the appointment of judges increase the output of the judiciary?. *International Review of Law and Economics*, *24*(3), 351-369.

Benesh, S. C., & Reddick, M. (2002). Overruled: An event history analysis of lower court reaction to Supreme Court alteration of precedent. *Journal of Politics*, *64(2),* 534-550.

Black, B., Silver, C., Hyman, D. A., & Sage, W. M. (2005). Stability, not crisis: medical malpractice claim outcomes in Texas, 1988–2002. *Journal of Empirical Legal Studies*, *2*(2), 207-259.

Bosnak, J. and Jonk, M., (2003). "The Netherlands", in International civil procedure (Vol. 5). (Grubbs, S. R., & North, P. M ed.) Kluwer law international. Alphen aan den Rijn.

Brennan, T. A., Sox, C. M., & Burstin, H. R. (1996). Relation between negligent adverse events and the outcomes of medical-malpractice litigation. *New England Journal of Medicine*, *335*(26), 1963-1967.

Brenner, S. (1980). Fluidity on the United States Supreme Court: A Reexamination. *American Journal of Political Science*, *24*(3), 526-535.

Buscaglia, E., & Dakolias, M. (1998). An analysis of the causes of corruption in the judiciary. *Law and Policy in International Business*, *30*, 95-116.

Calfee, J. E., & Craswell, R. (1984). Some effects of uncertainty on compliance with legal standards. *Virginia Law Review*, *70*(5), 965-1003.

Cameron, C. M. (1993). New avenues for modelling judicial politics (Paper presented at the Conference on the Political Economy of Public Law).

Cameron, C. M., & Kornhauser, L. A. (2006). Appeals mechanisms, litigant selection, and the structure of judicial hierarchies. In Institutional Games and the US Supreme Court, (Rogers J. R., Flemming R. B., Bond J. R., ed.). University of Virginia Press, Charlottesville and London.

Castro, M. F., & Guccio, C. (2014). Searching for the source of technical inefficiency in Italian judicial districts: an empirical investigation. *European Journal of Law and Economics*, *38*(3), 369-391.

CEPEJ (2006), European judicial systems: Facts and figures (data 2002). Strasbourg.
Chandra, A., Nundy, S., & Seabury, S. A. (2005). The growth of physician medical malpractice payments: evidence from the National Practitioner Data Bank. *Health Affairs*, *24*, 240-245.
Chappe, N. (2012). Demand for civil trials and court congestion. *European Journal of Law and Economics*, *33*(2), 343-357.
Chappe, N., & Giraud, R. (2013). Confidence, optimism, and litigation: A litigation model under ambiguity (Working Paper, University of Franche-Comté.) No. 05.
Chateauneuf, A., Eichberger, J., & Grant, S. (2007). Choice under uncertainty with the best and worst in mind: Neo-additive capacities. *Journal of Economic Theory*, *137*(1), 538-567.
Che, Y. K., & Yi, J. G. (1993). The role of precedents in repeated litigation. *Journal of Law, Economics, & Organization*, *9*(2), 399-424.
Chen, K. P., Chien, H. K., & Cyrus Chu, C. Y. (1997). Sequential versus unitary trials with asymmetric information. *The Journal of Legal Studies*, *26*(1), 239-258.
Choi, S. J., Gulati, M., & Posner, E. A. (2012). What do federal district judges want? An analysis of publications, citations, and reversals. *Journal of Law, Economics, and Organization*, *28*(3), 518-549.
Chopard, B., Marion, E., & Roussey, L. (2014). Does the appeals process lower the occurrence of legal errors? (Working Paper University of Paris West-Nanterre la Défense) No. 43.
Clancy, K., Bartolomeo, J., Richardson, D., & Wellford, C. (1981). Sentence decision-making: The logic of sentence decisions and the extent and sources of sentence disparity. *The Journal of Criminal Law and Criminology (1973-)*, *72*(2), 524-554.
Conley, J. M., & O'Barr, W. M. (1988). Hearing the hidden agenda: The ethnographic investigation of procedure. *Law and Contemporary Problems*, *51*(4), 181-197.
Cooter, R. D. (1983). The objectives of private and public judges. *Public Choice*, *41*(1), 107-132.
Cooter, R. D. (1985). Unity in tort, contract, and property: The model of precaution. *California Law Review*, *73*(1), 1-51.
Cooter, R. D., & Rubinfeld, D. L. (1989). Economic analysis of legal disputes and their resolution. *Journal of Economic Literature*, *27*(3), 1067-1097.
Cooter, R., & Ulen, T. (2004). Law and Economics. London. Pearson.
Cornell, B. (1990). The incentive to sue: An option-pricing approach. *The Journal of Legal Studies*, *19*(1), 173-187.

Craig, A. (1991). Product liability and safety in general aviation. In: The Liability Maze: The Impact of Liability Law on Safety and Innovation, (Peter W. Huber & Robert E. Litan eds.) Brookings Institution Press, Washington D.C.

Craswell, R., & Calfee, J. E. (1986). Deterrence and uncertain legal standards. *Journal of Law Economics & Organization*, *2*, 279-314.

Cremers, K., & Schliessler, P. (2015). Patent litigation settlement in Germany: why parties settle during trial. *European Journal of Law and Economics*, *40*(2), 185-208.

Cross, F. (2005). Appellate court adherence to precedent. *Journal of Empirical Legal Studies*, *2*(2), 369-405.

Cummins, J. D., Phillips, R. D., & Weiss, M. A. (2001). The incentive effects of no-fault automobile insurance. *The Journal of Law and Economics*, *44*(2), 427-464.

Dainow, J. (ed.) (1974) The role of judicial decisions and doctrine in civil law and in mixed jurisdictions. Baton Rouge: Louisiana State University Press. Baton Rouge.

Danzon, P. M., & Lillard, L. A. (1983). Settlement out of court: The disposition of medical malpractice claims. *The Journal of Legal Studies*, *12*(2), 345-377.

Dari-Mattiacci, G. (2005). Errors and the functioning of tort liability. *Supreme Court Economic Review*, 13, 165-187.

Dari-Mattiacci, G., & De Geest, G. (2005). Judgment proofness under four different precaution technologies. *Journal of Institutional and Theoretical Economics*, *161*(1), 38-56.

Dari-Mattiacci, G., & De Geest, G. (2006). When will judgment proof injurers take too much precaution?. *International Review of Law and Economics*, *26*(3), 336-354.

Dari-Mattiacci, G., & Deffains, B. (2007). Uncertainty of law and the legal process. *Journal of Institutional and Theoretical Economics*, *163*(4), 627-656.

Daughety, A. F., & Reinganum, J. F. (1993). Endogenous sequencing in models of settlement and litigation. *Journal of Law, Economics, & Organization*, *9*(2), 314-348.

Daughety, A. F., & Reinganum, J. F. (1994). Settlement negotiations with two-sided asymmetric information: Model duality, information distribution, and efficiency. *International Review of Law and Economics*, *14*(3), 283-298.

De Mot, J. (2013). Comparative versus contributory negligence: A comparison of the litigation expenditures. *International Review of Law and Economics*, *33*, 54-61.

De Mot, J., & Miceli, T. J. (2015). Comparing all-or-nothing and proportionate damages: A rent-seeking approach. *Review of Law & Economics*, *11*(1), 1-17.

De Mot, J., Faure, M., & Klick, J. (2015). Appellate caseload and the switch to comparative negligence. *International Review of Law and Economics*, *42*, 147-156.

Dennis, J. L. (1993). Interpretation and application of the civil code and the evaluation of judicial precedent. *Louisiana Law Review*, *54(1)*, 1-18.
Devlin, R. A. (1992). Liability versus no-fault automobile insurance regimes: An analysis of the experience in Quebec. In Contributions to Insurance Economics. (G. Dionne ed.) Haarlem. Springer Netherlands.
Dewees, D. N., Duff, D., & Trebilcock, M. J. (1996). Exploring the domain of accident law: Taking the facts seriously. Oxford. Oxford University Press.
Dimitrova-Grajzl, V., Grajzl, P., Slavov, A., & Zajc, K. (2016). Courts in a transition economy: Case disposition and the quantity–quality tradeoff in Bulgaria. *Economic Systems*, *40*(1), 18-38.
Dimitrova-Grajzl, V., Grajzl, P., Sustersic, J., & Zajc, K. (2012). Court output, judicial staffing, and the demand for court services: Evidence from Slovenian courts of first instance. *International Review of Law and Economics*, *32*(1), 19-29.
Doerner, W. M., & Ihlanfeldt, K. R. (2014). An empirical analysis of the property tax appeals process. *Journal of Property Tax Assessment & Administration*, *11*(4), 5-34.
Eastman W. & Viswanath P.V., (2003). Repeated Interaction, Deep Pockets and Litigation Spending, (Working Paper).
Eisenberg, T. (1990). Testing the selection effect: A new theoretical framework with empirical tests. *The Journal of Legal Studies*, *19*(2), 337-358.
Eisenberg, T., & Farber, H. S. (1997). The Litigious plaintiff hypothesis: Case selection and resolution, *RAND Journal of Economics*, 28(0), 92-103.
Epstein, L., Segal, J. A., & Spaeth, H. J. (2001). The norm of consensus on the US Supreme Court. *American Journal of Political Science*, *45*(2), 362-377.
Farber, H. S., & White, M. J. (1991). Medical malpractice: An empirical examination of the litigation process, RAND Journal of *Economics*, *22*(2), 199-215.
Farmer, A., & Pecorino, P. (1994). Pretrial negotiations with asymmetric information on risk preferences. *International Review of Law and Economics*, *14*(3), 273-281.
Farmer, A., & Pecorino, P. (1999). Legal expenditure as a rent-seeking game. *Public Choice*, *100*(3), 271-288.
Farmer, A., & Pecorino, P. (2002). Pretrial bargaining with self-serving bias and asymmetric information. *Journal of Economic Behavior & Organization*, *48*(2), 163-176.
Farmer, A. and Pecorino, P. (2002) Pretrial signaling with negative expected value suits, (University of Alabama, Economics, Finance and Legal Studies Working Paper) No. 02-07-01.

Farmer, A., Pecorino, P., & Stango, V. (2004). The causes of bargaining failure: Evidence from Major League Baseball. *The Journal of Law and Economics*, *47*(2), 543-568.

Farmer, A., & Pecorino, P. (2013). Discovery and disclosure with asymmetric information and endogenous expenditure at trial. The Journal of Legal Studies, *42*(1), 223-247.

Farmer, A., & Pecorino, P. (2016). Litigation with a variable cost of trial. *Review of Law & Economics*, *12*(2), 203-226.

Fehr, E. & Schmidt, K.M., (2000) Theories of fairness and reciprocality: evidence and economic applications, (CESinfo Working Paper Series), No. 413.

Felsö, F., Hollanders, D., van Honk, J., Laemers, M., de Nooij, M., & Rombouts, W. (2007). De tijd loopt door: de gevolgen van lange doorlooptijden in de rechtspraak. (SEO-rapport). No. 994.

Fenn, P. & Vlachonikolis, I., (1990) Bargaining behavior of defendant insurers: an economic model, *Geneva Papers on Risk and Insurance*, 14, 41-52.

Fenn, P., & Rickman, N. (1999). Delay and settlement in litigation. *The Economic Journal*, *109*(457), 476-491.

Fenn, P., & Rickman, N. (2001). Asymmetric information and the settlement of insurance claims. *Journal of Risk and Insurance*, *68*(4), 615-630.

Fenn, P., & Rickman, N. (2005) Legal liability and the timing of settlement in medical malpractice. (American Law & Economics Association Annual Meetings).

Fenn, P. & Rickman, N, (2013). Information and the disposition of medical malpractice claims: A competing risks analysis. *Journal of Law and Economics*, DOI: https://doi.org/10.1093/jleo/ewt002

Fischman, J. B. (2011). Estimating preferences of circuit judges: A model of consensus voting. *The Journal of Law and Economics*, *54*(4), 781-809.

Fischman, J. B. (2013). Measuring inconsistency, indeterminacy, and error in adjudication. *American law and economics review*, *16*(1), 40-85.

Fournier, G. M., & Zuehlke, T. W. (1989). Litigation and settlement: an empirical approach. *The Review of Economics and Statistics*, *71*(2), 189-195.

Fournier, G. M., & Zuehlke, T. W. (1996). The timing of out-of-court settlements. *The RAND Journal of Economics*, *27*(2), 310-321.

Freidman, E., & Wickelgren, A. L. (2008). No free lunch: How settlement can reduce the legal system's ability to induce efficient behavior. *SMU Law Review*, *61*(4), 1355-1376.

Froeb, L. (1993). The adverse selection of cases for trial. *International Review of Law and Economics*, *13*(3), 317-324.
Galanter, M. (1974). Why the "haves" come out ahead: Speculations on the limits of legal change. *Law & Society Review*, *9*(1), 95-160.
Genn, H., (1995) "Access to just settlement: The case of medical negligence", In Reform of Civil Procedure: Essays on Access to Justice (A.A.S. Zuckerman and R. Cranston ed.), Oxford, Oxford University Press.
Gennaioli, N., & Shleifer, A. (2007). The evolution of common law. *Journal of Political Economy*, *115*(1), 43-68.
Gennaioli, N., & Shleifer, A. (2007). Overruling and the Instability of Law. *Journal of Comparative Economics*, *35*(2), 309-328.
Gertner, R. H. (1998) Disclosure and Unravelling, in The New Palgrave Dictionary of Economics and the Law, (Newman, P. ed.), London, Macmillan.
Gilboa I. (1992) "Why the empty shells weren't fired: a semi-biographical note", (Northwestern University, Center for Mathematical Studies in Economics and Management Science), No. 987.
Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, *18*(2), 141-153.
Gilles, S. G. (2006). The judgment-proof society. *Washington & Lee Law Review*, *63*(2), 603-718.
Gilson, R. J., & Mnookin, R. H. (1994). Disputing through agents: Cooperation and conflict between lawyers in litigation. *Columbia Law Review*, *94*(2), 509-566.
Goriely, T., Moorhead, R. and Abrams, P. (2002) More civil justice? The impact of the Woolf reforms on pre-action behavior, (Law Society and Civil Justice Council), No. 43.
Gould, J. P. (1973). The economics of legal conflicts. *The Journal of Legal Studies*, *2*(2), 279-300.
Graham, J. D. (1991). Product liability and motor vehicle safety. In The liability maze: The impact of liability law on safety and innovation (Peter W. Huber & Robert E. Litan ed.) Brookings Institution Press, Washington D.C.
Gross, S. R., & Syverud, K. D. (1991). Getting to no: A study of settlement negotiations and the selection of cases for trial. *Michigan Law Review*, *90*(2), 319-393.
Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *The Journal of Law and Economics*, *24*(3), 461-483.

Grubbs, S. and DeCambra E., (2003) "United States", in International civil procedure (Vol. 5). (Grubbs, S. R., & North, P. M ed.) Kluwer law international. Alphen aan den Rijn.

Grundfest, J. A., & Huang, P. H. (2006). The unexpected value of litigation: A real options perspective. *Stanford Law Review*, *58*(5), 1267-1336.

Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2007). Blinking on the bench: How judges decide cases. *Cornell Law Review*, *93*(1), 1-44.

Hay, B. and Spier, K. (1998) Settlement of litigation, in (Newman, P. ed), The New Palgrave Dictionary of Economics and the Law, London, MacMillan.

Harris, L. (1989). Judges' opinions on procedural issues: A survey of state and federal judges who spend at least half their time on general civil cases. *Boston University Law Review*, *69*(3), 731-768.

Harris, W. T. (1992). A public choice analysis of the evolution of tort law. *American Journal of Economics and Sociology*, *51*(1), 101-108.

Harvard Medical Practice Study. (1990) Patients, doctors and lawyers: studies of medical injury, malpractice litigation, and patient compensation in New York.

Helland, E., & Klick, J. (2007). The effect of judicial expedience on attorney fees in class actions. *The Journal of Legal Studies*, *36*(1), 171-187.

Helland, E. A., & Tabarrok, A. (2012). Product liability and moral hazard: Evidence from general aviation. *The Journal of Law and Economics*, *55*(3), 593-630.

Hillman, A. L., & Katz, E. (1984). Risk-averse rent seekers and the social cost of monopoly power. *The Economic Journal*, *94*(373), 104-110.

Hirshleifer, J., & Osborne, E. (2001). Truth, effort, and the legal battle. *Public Choice*, *108*(1), 169-195.

Horwitz, J. R., & Mead, J. (2009). Letting good deeds go unpunished: Volunteer immunity laws and tort deterrence. *Journal of Empirical Legal Studies*, *6*(3), 585-635.

Hylton, K. N. (1991). Litigation costs and the economic theory of tort law. *University of Miami Law Review*, *46*(1), 111-148.

Hylton, K. N. (1993). Asymmetric information and the selection of disputes for litigation. *The Journal of Legal Studies*, *22*(1), 187-210.

Hylton, K. N. (2002). An asymmetric-information model of litigation. *International Review of Law and Economics*, *22*(2), 153-175.

Jia, H., Skaperdas, S., & Vaidya, S. (2013). Contest functions: Theoretical foundations and issues in estimation. *International Journal of Industrial Organization*, *31*(3), 211-222.

Kahan, M. (1989). Causation and incentives to take care under the negligence rule. *The Journal of Legal Studies*, *18*(2), 427-447.
Kakalik, J.S. and Robyn, A. (1982) Costs of the civil justice system: Court expenditures for processing tort cases, (RAND- ICJ), No. 2793.
Kakalik, J.S. (1983) Costs of the civil justice system: Court expenditures for various types of civil cases, (RAND - ICJ), No. 2985.
Kakalik, J. S., Ebener, P. A., Felstiner, W. L., Haggstrom, G. W., & Shanley, M. G. (1984). Variation in asbestos litigation compensation and expenses (RAND-ICJ) No. 3132.
Kakalik, James S., Elizabeth M. King, Michael Traynor, Patricia A. Ebener, and Larry Picus (1988). Costs and compensation paid in aviation accident litigation, (RAND-ICJ) No. 3421.
Kakalik, J. S., Hensler, D. R., McCaffrey, D., & Oshiro, M. (1998). Discovery Management: Further Analysis of the Civil Justice Reform Act Evaluation Data. *Boston College Law Review, 39*(3), 613-682.
Kamphorst, J. J., & Van Velthoven, B. C. (2009). The introduction of an appeals court in Dutch tax litigation. *International Review of Law and Economics*, *29*(1), 13-24.
Kapardis, A., & Farrington, D. P. (1981). An experimental study of sentencing by magistrates. *Law and Human Behavior*, *5*(2), 107.
Kaplow, L., & Shavell, S. (1994). Accuracy in the determination of liability. *The Journal of Law and Economics*, *37*(1), 1-15.
Kaplow, L., & Shavell, S. (1996). Accuracy in the assessment of damages. *The Journal of Law and Economics*, *39*(1), 191-210.
Katz, A. (1988). Judicial decisionmaking and litigation expenditure. *International Review of Law and Economics*, *8*(2), 127-143.
Keeling, B. C. (1993). Toward a balanced approach to frivolous litigation: A critical review of federal rule 11 and state sanctions provisions. *Pepperdine Law Review*, *21*(4), 1067-1160.
Keeney R. & Raiffa H. (1976), Decisions with multiple objectives: Preferences and value tradeoffs, Cambridge, Cambridge University Press.
Kennan, J., & Wilson, R. (1993). Bargaining with private information. *Journal of Economic Literature*, *31*(1), 45-104.
Kessler, D. (1996). Institutional causes of delay in the settlement of legal disputes. *Journal of Law, Economics, and Organization*, *12*(2), 432-460.
Klein, D. E., & Hume, R. J. (2003). Fear of reversal as an explanation of lower court compliance. *Law & Society Review*, *37*(3), 579-581.

Klijn, A. (1993) Verleidt de WROM tot lossigheid? Over (in)efficiënt procedeergedrag van advocaten, *Advocatenblad*, 73.

Kobayashi, B., (1996) Case selection, external effects, and the trial/settlement decision, in Dispute Resolution: Bridging the Settlement Gap, (Anderson, David A. ed.), Greenwood, JAI Press.

Konrad, K. A., & Schlesinger, H. (1997). Risk aversion in rent-seeking and rent-augmenting games. *The Economic Journal*, *107*(445), 1671-1683.

Koopmans, C. C., & Gerritsen, M. (2014). Evaluatie van het model 'In de schaduw van de rechter'. (SEO Economisch Onderzoek Report), No. 53.

Kornhauser Lewis, A. (1989). An economic perspective on stare decisis. *Chicago-Kent Law Review*, 65(1), 63-92.

Kritzer, H. M. (1985) The form of negotiation in ordinary litigation, Institute for Legal Studies, (Working Paper Series 7), No. 2.

Landes, W. M. (1971). An economic analysis of the courts. *The Journal of Law and Economics*, *14*(1), 61-107.

Landes, W. M., & Posner, R. A. (1976). Legal precedent: A theoretical and empirical analysis. *The Journal of Law and Economics*, *19*(2), 249-307.

Landes, E. M. (1982). Insurance, liability, and accidents: a theoretical and empirical investigation of the effect of no-fault accidents. *The Journal of Law and Economics*, *25*(1), 49-65.

Landes, W. M. (1993). Sequential versus unitary trials: An economic analysis. *The Journal of Legal Studies*, *22*(1), 99-134.

Landes, W. M., & Posner, R. A. (1979). Adjudication as a private good. *The Journal of Legal Studies*, *8*(2), 235-284.

LoPucki, L. M. (1996). The death of liability. *The Yale Law Journal*, *106*(1), 1-92.

Macey, J. R. (1989). The internal and external costs and benefits of stare decisis. *Chi.-Kent Law Review*, *65*(1), 93-114.

Macey, J. R. (1998). Winstar, bureaucracy and public choice. *Supreme Court Economic Review*, *6*, 173-200.

Martin, R. (1991). General aviation manufacturing: An industry under siege. In the Liability Maze, (Huber and Litan ed.), Brookings Institution Press, Washington.

Marshall, L. C., Kritzer, H. M., & Zemans, F. K. (1991). Use and impact of Rule 11. *Northwestern University Law Review*, *86*(4), 943.

Martin, A. D., & Quinn, K. M. (2002). Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999. *Political Analysis*, *10*(2), 134-153.

Marvell, T.B., (1989). State appellate court responses to caseload growth, *Judicature*, 72(5), 282-291.
McFadden, D. 1973. Conditional logit analysis of qualitative choice behavior. In Frontiers in Economics (P. Zarembka ed.). New York. Academic Press.
McNollgast (1995) Politics and the courts: A positive theory of judicial doctrine and the rule of law. *Southern California Law Review*, *68*(6), 1631-1684.
Miceli, T. J. (1999). Settlement delay as a sorting device. *International Review of Law and Economics*, *19*(2), 265-274.
Miceli, T. (2008). The social versus private incentive to sue (University of Connecticut, Department of Economics Working Paper) No. 12.
Miceli, T. J. (2009). Legal change: selective litigation, judicial bias, and precedent. *The Journal of Legal Studies*, *38*(1), 157-168.
Miceli, T. J. (2010). Legal change and the social value of lawsuits. *International Review of Law and Economics*, *30*(3), 203-208.
Ministerie van Justitie, Rechtspraak en rechtshandhaving; maatschappelijke effecten van verbetering, Werkgroep Effecten Rechtspraak, research group: F. van Dijk, J. van Dijk and R. Teijl. Ministerie van Justitie, 1998. Also available at www.rechtspraak.nl.
Molot, J. T. (2014). The feasibility of litigation markets. *Indiana Law Journal*, *89*, 171-194.
Muthoo, A. (1999) Bargaining theory with applications, Cambridge, Cambridge University Press.
Nalebuff, B. (1987). Credible pretrial negotiation. *The RAND Journal of Economics*, *18*(2), 198-210.
Nash, J. (1953). Two-person cooperative games. *Econometrica: Journal of the Econometric Society*, 21(1), 128-140.
New York Times, (2011, August 24) Addressing the justice gap. New York Times, 22.
Ordover, J. A. (1978). Costly litigation in the model of single activity accidents. *The Journal of Legal Studies*, *7*(2), 243-261.
Osborne, E. (1999). Who should be worried about asymmetric information in litigation?. *International Review of Law and Economics*, *19*(3), 399-409.
Parisi, F. (2002). Rent-seeking through litigation: adversarial and inquisitorial systems compared. *International Review of Law and Economics*, *22*(2), 193-216.
Partridge, A., & Eldridge, W. B. (1974). The Second Circuit sentencing study: A report to the judges of the Second Circuit. (Federal Judicial Center Working Paper). No. 015058.

Pecorino, P., & Van Boening, M. (2004). An empirical analysis of bargaining with voluntary transmission of private information. *The Journal of Legal Studies*, *33*(1), 131-156.

Pecorino, P., & Van Boening, M. (2010). Fairness in an embedded ultimatum game. *The Journal of Law and Economics*, *53*(2), 263-287.

Pecorino, P., & Van Boening, M. (2014). Bargaining with asymmetric dispute Costs. *Review of Law & Economics*, *10*(1), 31-58.

Peters, P. G. (2009). Twenty years of evidence on the outcomes of malpractice claims. *Clinical Orthopaedics and Related Research*, *467*(2), 352-357.

Polinsky, A. M., & Che, Y. K. (1991). Decoupling liability: optimal incentives for care and litigation. (National Bureau of Economic Research Working Paper). No. 3634.

Posner, R. A. (1973). An economic approach to legal procedure and judicial administration. *The Journal of Legal Studies*, *2*(2), 399-458.

Posner, R.A. (1992). *Economic Analysis of Law*. Boston. Little, Brown and Company.

Posner, R. A. (1993). What do judges and justices maximize? (The same thing everybody else does). *Supreme Court Economic Review*, *3*, 1-41.

Posner, R. A. (2008). How judges think. Harvard University Press.

Post, R. (2001). The Supreme Court opinion as institutional practice: Dissent, legal scholarship, and decisionmaking in the Taft court. *Minnesota Law Review*, *85*(5), 1267-1391.

Priest, G. L. (1977). The common law process and the selection of efficient rules. *The Journal of Legal Studies*, *6*(1), 65-82.

Priest, G. L. (1988). Understanding the liability crisis. *Proceedings of the Academy of Political Science*, *37*(1), 196-211.

Priest, G. L., & Klein, B. (1984). The selection of disputes for litigation. *The Journal of Legal Studies*, *13*(1), 1-55.

Priest, G. L. (1989). Private litigants and the court congestion problem. *Boston University Law Review*, *69*(3), 527.

Reinganum, J. F., & Wilde, L. L. (1986). Settlement, litigation, and the allocation of litigation costs. *The RAND Journal of Economics*, 17*(4)*, 557-566.

Rosales-López, V. (2008). Economics of court performance: an empirical analysis. *European Journal of Law and Economics*, *25*(3), 231-251.

Roth, A., (1995) Bargaining experiments, in The Handbook of Experimental Economics (J.H. Kagel and A.E. Roth ed.), Princeton. Princeton University Press.

Rubin, P. H. (1977). Why is the common law efficient?. *The Journal of Legal Studies*, *6*(1), 51-63.

Rubin, P. H., & Shepherd, J. M. (2008). The demographics of tort reform. *Review of Law and Economics*, *4*(2), 591-620.
Salop, S. C. & White, L. J., (1988) Private antitrust litigation: An introduction and framework, in Private Antitrust Litigation, (White, L. ed.) Cambridge. MIT Press.
Sanchirico, C. W. (2004). Evidence tampering. *Duke Law Journal*, 53(4), 1215-1336.
Santolino, M. (2010). Determinants of the decision to appeal against motor bodily injury judgements made by Spanish trial courts. *International Review of Law and Economics*, *30*(1), 37-45.
Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica: Journal of the Econometric Society*, 57(3), 571-587.
Schwartz, W. F. (1966). Severance-A means of minimizing the role of burden and expense in determining the outcome of litigation. *Vanderbilt Law Review*, *20*(6), 1197-1215.
Schwartz, W. F. (2003). Can suits with negative expected value really be profitable? defendants can play games too. *Legal Theory*, 9, 83–97.
Schwartz, W. F., & Wickelgren, A. L. (2009). Advantage defendant: why sinking litigation costs makes negative-expected-value defenses but not negative-expected-value suits credible. *The Journal of Legal Studies*, *38*(1), 235-253.
Schweizer, U. (1989). Litigation and settlement under two-sided incomplete information. *The Review of Economic Studies*, *56*(2), 163-177.
Seabury, S. A. (2009). Case selection after trial: A study of post-trial settlement and appeal. (RAND-ICJ) No. 638.
Seron, C., Frankel, M., Van Ryzin, G., & Kovath, J. (2001). The impact of legal counsel on outcomes for poor tenants in New York City's housing court: results of a randomized experiment. *Law and Society Review*, 35(2), 419-434.
Shapiro, M. (1972). Toward a theory of "stare decisis". *The Journal of Legal Studies*, *1*(1), 125-134.
Shavell, S. (1982). Suit, settlement, and trial: A theoretical analysis under alternative methods for the allocation of legal costs. *The Journal of Legal Studies*, *11*(1), 55-81.
Shavell, S. (1986). The judgment proof problem. *International Review of Law and Economics*, *6*(1), 45-58.
Shavell, S. (1989). Sharing of information prior to settlement or litigation. *The Rand Journal of Economics*, *20*(2), 183-195.
Shavell, S. (1995). The appeals process as a means of error correction. *The Journal of Legal Studies*, *24*(2), 379-426.

Shavell, S. (1999). The level of litigation: Private versus social optimality of suit and of settlement. *International Review of Law and Economics*, *19*(1), 99-115.

Shavell, S. (2004). Foundations of Economic Analysis of Law. Cambridge. Harvard University Press.

Shavell, S. (2006). The appeals process and adjudicator incentives. *The Journal of Legal Studies*, *35*(1), 1-29.

Shepherd, G. B. (1999). An empirical study of the economics of pretrial discovery. *International Review of Law and Economics*, *19*(2), 245-263.

Sieg, H. (2000). Estimating a bargaining model with asymmetric information: evidence from medical malpractice disputes. *Journal of Political Economy*, *108*(5), 1006-1021.

Siegelman, P., & Waldfogel, J. (1999). Toward a taxonomy of disputes: New evidence through the prism of the Priest/Klein model. *The Journal of Legal Studies*, *28*(1), 101-130.

Sisk, G. C., Heise, M., & Morriss, A. P. (1998). Charting the influences on the judicial mind: An empirical study of judicial reasoning. *New York University Law Review*, 73(5), 1377-1500.

Skaperdas, S., & Gan, L. (1995). Risk aversion in contests. *The Economic Journal*, 105(431), 951-962.

Sloan, F. A., Reilly, B. A., & Schenzler, C. (1995). Effects of tort liability and insurance on heavy drinking and drinking and driving. *The Journal of Law and Economics*, *38*(1), 49-77.

Snyder, E. A., & Hughes, J. W. (1990). The English rule for allocating legal costs: evidence confronts theory. *Journal of Law, Economics, & Organization*, *6*(2), 345-380.

Sobel, J. (1989). An analysis of discovery rules. *Law and Contemporary Problems*, *52*(1), 133-159.

Songer, D. R., Segal, J. A., & Cameron, C. M. (1994). The hierarchy of justice: Testing a principal-agent model of Supreme Court-circuit court interactions. *American Journal of Political Science*, 38(3), 673-696.

Spier, K. E. (1992). The dynamics of pretrial negotiation. *The Review of Economic Studies*, *59*(1), 93-108.

Stanley, L. R., & Coursey, D. L. (1990). Empirical evidence on the selection hypothesis and the decision to litigate or settle. *The Journal of Legal Studies*, *19*(1), 145-172.

Stein, P. & Rueb, A. (2003) Burgerlijk procesrecht. Deventer, Kluwer.

Stith, K., & Cabranes, J. A. (1998). Fear of judging: Sentencing guidelines in the federal courts. Chicago. University of Chicago Press.

Stone, M. P., & Miceli, T. J. (2013). The impact of frivolous lawsuits on deterrence: Do they have some redeeming value. *Journal of Law, Economics & Policy*, *10(2)*, 301-344.

Stras, D. R. (2006). The incentives approach to judicial retirement. *Minnesota Law Review*, *90*(5), 1417-1446.

Studdert, D. M., Mello, M. M., Gawande, A. A., Gandhi, T. K., Kachalia, A., Yoon, C., Puopolo A. L., & Brennan, T. A. (2006). Claims, errors, and compensation payments in medical malpractice litigation. *New England Journal of Medicine*, *354*(19), 2024-2033.

Summers, J. (1983). The case of the disappearing defendant: An economic analysis. *University of Pennsylvania Law Review*, *132*(1), 145-185.

Swanson, T., & Mason, R. (1998). Nonbargaining in the shadow of the Law. *International Review of Law and Economics*, *18*(2), 121-140.

Thaler, R. H. (1988). Anomalies: The ultimatum game. *The Journal of Economic Perspectives*, *2*(4), 195-206.

Treich, N. (2010). Risk-aversion and prudence in rent-seeking games. *Public Choice*, *145*(3), 339-349.

Trubek, D. M., Sarat, A., Felstiner, W. L., & Kritzer, H. M. (1983). The costs of ordinary litigation. *University of California Law Review*, *31*(1), 72.

Van Dijk, F. (2014). Improved performance of the Netherlands Judiciary: Assessment of the gains for society. *International Journal for Court Administration*, *6*(1), 1-17.

Van Koppen, P. J., & Kate, J. T. (1984). Individual differences in judicial behavior: Personal characteristics and private law decision-making. *Law and Society Review*, 18(2), 225-247.

Van Os R. & Smit M. (2016) Civielrechtelijke appelrechtspraak, een cijfermatige erkenning. (WODC, Cahier) No. 9.

Van Tulder, F. (2014). In de schaduw van de rechter: Individuele en maatschappelijke kosten en baten van de juridische infrastructuur. (Den Haag: Raad voor de rechtspraak. Research Memoranda) No. 4.

Van Tulder, F. (2016). In the shadow of the Dutch courts Individual and social costs and benefits of litigation. (Council for the Judiciary, Working Paper) No. 2.

Van Velthoven, B. Voert, M., (2003) Geschilbeslechtingsdelta. Over verloop en afloop van (potentieel) juridische problemen van burgers, (Ministerie van Justitie, WODC Onderzoek en Beleid) No. 219,

Van Wijck, P., & Van Velthoven, B. (2000). An economic analysis of the American and the continental rule for allocating legal costs. *European Journal of Law and Economics*, *9*(2), 115-125.

Vereeck, L., & Mühl, M. (2000). An economic theory of court delay. European Journal of Law and Economics, 10(3), 243-268.

Viscusi, W. K. (1988). Product liability litigation with risk aversion. *The Journal of Legal Studies*, *17*(1), 101-121.

Voert, M. T., & van Ewijk, M. D. (2006). Trendrapportage gerechtsdeurwaarders 2006. (Onderzoek en Beleid). No. 247.

Waldfogel, J. (1995). The selection hypothesis and the relationship between trial and plaintiff victory. *Journal of Political Economy*, *103*(2), 229-260.

Waldfogel, J. (1998) Selection of cases for trial, in (P. Newman ed.), The New Palgrave Dictionary of Economics and the Law, 3, London, MacMillan.

Waldfogel, J. (1998). Reconciling asymmetric information and divergent expectations theories of litigation. *The Journal of Law and Economics*, *41*(2), 451-476.

Wang, G. H., Kim, J. Y., & Yi, J. G. (1994). Litigation and pretrial negotiation under incomplete information. *The Journal of Law. Economics, & Organization*, *10*(1), 187-200.

Watanabe, Y. (2005). Learning and bargaining in dispute resolution: Theory and evidence from medical malpractice litigation. (Northwestern University Working Paper).

Weiler, P. C. (1993). The case for no-fault medical liability. *Maryland Law Review*, *52*(4), 908-950.

Wiggins E. e.a., (1991) Federal judicial CTR., Rule 11: final report to the advisory committee on civil rules of the judicial conference of the United States (Federal Judicial Center Report) No. 46.

Willging, T. E., Stienstra, D., Shapard, J., & Miletich, D. (1998). An empirical study of discovery and disclosure practice under the 1993 Federal Rule Amendments. *Boston College Law Review*, *39*(3), 525-596.

Williams, G. (1983) Legal negotiations and settlement. St. Paul. West Publishing Company.

Wilson, R., (1998) Private information and legal bargaining, in Newman, P. (ed.), The New Palgrave Dictionary of Economics and the Law, 3, London, Macmillan.

Wistrich, A. J., Guthrie, C., & Rachlinski, J. J. (2005). Can judges ignore inadmissible information? The difficulty of deliberately disregarding. *University of Pennsylvania Law Review*, 153(4), 1251-1345.

Wittman, D. (1985). Is the selection of cases for trial biased?. *The Journal of Legal Studies*, *14*(1), 185-214.
Wittman, D. (1988). Dispute resolution, bargaining, and the selection of cases for trial: A study of the generation of biased and unbiased data. *The Journal of Legal Studies*, *17*(2), 313-352.
Wohlschlegel, A. (2014). The appeals process and incentives to settle (University of Portsmouth).
Yildiz, M. (2003). Bargaining without a common prior—an immediate agreement theorem. *Econometrica*, *71*(3), 793-811.
Yildiz, M. (2004) Optimism, Deadline Effect, and Stochastic Deadlines, (MIT Department of Economics Working Paper) No. 36.
Zywicki, T. J. (2003) The Rise and Fall of Efficiency in the Common Law: A Supply-Side Analysis. *Northwestern University Law Review*, 97(4), 1551-1634.

List of Cases

Neitzke v. Williams, 490 U.S. 319, 325 (1989).
Doak v. Superior Court 257 Cal. App. 2nd 285 (1968).