Efficient Incentives from Obligation Law and the Compensation Principle
Urs Schweizer,*†University of Bonn

September 14, 2015

Abstract

The compensation principle provides a link between the legal requirement to compensate for deviations from an efficient reference profile and the economic desideratum of welfare maximizing incentives. Quantifying damages in line with the difference hypothesis, even relative to an inefficient obligation profile, would ensure the compensating goal being achieved as required for the compensation principle. The paper applies this insight to various settings from tort and contract law, leading to new results but also to an unifying perspective on findings from the existing literature.

JEL classification: K12, K13

Keywords: legal obligation, difference hypothesis, compensation goal, compensation principle, efficient incentives, efficient breach, negligence rule, liability in sequential choice, hold-up situation

*address: Urs Schweizer, Department of Economics, University of Bonn, Kaiserstr. 1 53113 Bonn, Germany, schweizer@uni-bonn.de
†Support by the DFG through SFB/TR 15 is gratefully acknowledged.
1 Introduction

Legal rules are called efficient if they generate incentives for strategically acting parties to take decisions leading to a welfare maximizing outcome.

Obligation law provides general rules for contractual and tort relationships. If a debtor deviates from a (contractual or legal) obligation, the law offers remedies to creditors who suffer from harm caused by the debtor’s deviation. These remedies aim at compensating creditors.

The compensation principle, finally, refers to a link between the legal requirement of compensation and the economic concept of efficient rules. If each party is compensated for unilateral deviations from an efficient reference profile by the other party then all Nash equilibria of the underlying game are efficient and payoff equivalent.

The compensation goal is, in particular, achieved if creditors are awarded damages in line with the difference hypothesis. In Germany, this hypothesis is attributed to Friedrich Mommsen, a legal scholar from the nineteenth century. Accordingly, damages should account for that part of the harm that was caused by the deviation from the obligation and should be calculated as the difference of the hypothetical (hence counterfactual) value of the creditor’s assets if the debtor had met her obligation and their actual value, given that actually she has not.

Notice, in common law countries, damages in line with the difference hypothesis would just be called expectation damages, without giving credit to any legal scholar and without referring to a difference hypothesis explicitly. In the present paper, I will refer to this hypothesis nonetheless.

Traditional economic analysis of tort law has shown that, except for strict liability, most damages rules including the negligence rule with and without a defense of contributory negligence as well as many other related schemes provide efficient incentives for all parties if courts specify due care standards at their efficient level. The economic analysis of contract law, in contrast, has uncovered inefficiencies in the form of overreliance by the creditor. At first glance, such a discrepancy in terms of efficiency seems puzzling, in particular under the legal regime of an obligation law that governs contractual and tort relationships by common principles.

The puzzle is easily resolved. While, in tort cases, courts are expected to
specify obligations at their efficient level, parties to a contract define obligations themselves. To economize on transaction costs, they may deliberately abstain from taking all conceivable contingencies into account. The literature, in fact, has mainly focused on completely non-contingent obligations which parties stipulate even in the presence of uncertainty. Parties rather rely on remedies for breach of duties as offered by contract law whenever corresponding contingencies arise. In any case, non-contingent contracts define obligations that may fail to be efficient (for some contingencies at least) and are regularly claimed to generate, under expectation damages, excessive incentives for reliance investments.

The present paper, in contrast, takes a closer look at the difference hypothesis. As it turns out, even for inefficient obligations, a damages regime still in line with the difference hypothesis allows fully restoring efficiency. To maintain efficient incentives, a party who breaches her obligation should owe damages to the other party in line with the difference hypothesis based, not on the actual performance of the other party, but of that of a reasonable person. Put differently, if the other party has not behaved like a reasonable person, a cap should be put on her claims in order to restore efficient incentives. In this way, the suitable version of the difference hypothesis will generate incentives for efficient breach, given that the obligation profile fails to be efficient.

In fact, the appropriate reasonable person standard will ensure that the compensation goal relative to an efficient reference profile is still achieved. The efficiency claim of the general difference principle then immediately follows from the compensation principle.

As an important application, the proposed damages regime will be spelled out for a general hold-up situation involving non-contingent contracts and two-sided investments of selfish, cooperative or even hybrid nature. To implement the proposed damages regime, courts must be able to detect deviations of relationship-specific investments from their efficient level. In the traditional hold-up literature, investments are usually assumed to be hidden actions and, from that perspective, the present paper does not solve the hold-up problem.

In fact, the informational setting examined by the present paper would allow for many other efficient mechanisms. Yet, the additional merit of the
proposed scheme stems from being exclusively based on the logic behind the
difference hypothesis, a widely accepted legal concept.

Göller and Hewer (2014) have examined compensation rules for tak-
ings that provide two-sided efficient incentives for purely selfish investments.
Their main result can easily be reframed as a special case of the difference
principle in the sense of the present paper. The same holds true for Schweizer
(2006), who has proposed a bilateral damages regime that ensures an efficient
outcome in a setting of one-sided investments of the cooperative type.

The present paper generalizes these earlier findings, allowing for two-sided
investments even of hybrid type. Moreover, while these earlier contributions
have assumed courts to know counterfactuals, the present paper introduces
a damages regime in line with the difference hypothesis which courts could
implement even if counterfactuals remain unknown to them.

The difference principle as proposed by the present paper requires express-
ing the strategic interaction among parties in normal form even if sequential
choice is at stake. As a consequence, the efficient reference profile will only
form a Nash equilibrium of the underlying normal form game but need not
correspond to a subgame perfect equilibrium in its extensive form. In fact,
off the equilibrium path, the subgame perfect continuation may fail to be
ex post efficient, which paves the way for voluntary renegotiations. Yet, as
both parties must agree, renegotiations will reinforce the compensation goal
being achieved. It then follows from the compensation principle that not
even (anticipated) renegotiations would distort investment incentives.

Moreover, for unidirectional externalities, damages in line with the differ-
ence hypothesis prove flexible enough to provide efficient incentives also off
the equilibrium path. Based on the work of Rea (1987), Grady (1988) and
Kornhauser and Revesz (1991), the textbook by Miceli (2008) nicely summa-
rizes earlier findings on liability under sequential moves. Most of the rules
he examines, however, fail to generate efficient incentives off the equilibrium
path with the exception of marginal cost liability as pioneered by Wittman
(1981). In spite of their nice properties, however, (as Miceli argues) courts
do not seem to follow marginal cost liability in practice.

The justification proposed by the present paper, in contrast, rests on
damages in line with the difference hypothesis. As this concept is widely
recognized by law and courts, the proposed damages regime may be of prac-
tical relevance nonetheless.

On top of truly new results, the paper also takes a fresh look at well-known findings from the existing literature on the efficiency of a whole variety of damages regimes rule for the setting of the accident model. As all these regimes satisfy the conditions needed for the compensation principle, a single proof turns out sufficient to establish all these efficiency results at once. As a fringe benefit of the approach, much weaker assumptions are needed compared with the earlier literature, which predominantly has relied on concavity and calculus.

For sake of completeness, let me also mention Schweizer (2005) on the economic analysis of obligation law. While, in that paper, I had focussed on the mathematical saddle point property, the present paper takes a more legal perspective by referring to the compensation goal and the difference hypothesis instead.

The paper is organized as follows. Section 2 establishes the compensation principle for strategic interaction expressed in normal form. If each party is compensated for unilateral deviations from the efficient reference profile by the other party then this reference profile forms a Nash equilibrium and all Nash equilibria (if more than one exists) are payoff equivalent. Section 3 introduces the notion of damages in line with the difference hypothesis relative to a possibly inefficient obligation profile. To ensure efficient incentives nonetheless, a cap on claims is needed, based on a reasonable person standard. This general efficiency result is referred to as difference principle.

In section 4, the traditional model of sequential care choice serves as a first application of the difference principle. If the externality is of unidirectional nature (as in the accident model), the ex post efficient reference profile can be implemented, not only as a Nash equilibrium of the normal form, but even as a subgame perfect equilibrium in the extensive form of the game.

Section 5 considers non-contingent contractual obligations in a relationship with two-sided reliance investments under uncertainty and a (binary) performance decision to be taken ex post. If counterfactuals were observed by courts then a damages regime could be implemented that remains in line with the difference hypothesis even from the ex post perspective. If counterfactuals remain unobservable, however, damages could still be quantified in line with the difference hypothesis, but only from the ex ante perspec-
tive (which is sufficient to generate efficient incentives). This result will be established in section 6. Section 7 concludes.

2 Compensation principle

I consider the following contractual or legal relationship between two parties A and B. Party A (you may think of a buyer in a contractual relationship or a victim in a tort relationship) takes a decision $a$ from the set $A$ of her alternatives. Party B (think of a seller/producer or an injurer/tortfeasor) takes a decision $b$ from the set $B$ at his disposal. The net values of party A’s and B’s assets are assumed to be functions $m(a, b)$ and $n(a, b)$ of the decision profile $(a, b)$ as chosen from the Cartesian product $A \times B$ by the parties. Welfare $w(a, b) = m(a, b) + n(a, b)$ amounts to the sum of the value functions.

The function $w(a, b)$ captures welfare free of possible frictions from the underlying institutional arrangement. Payoff functions reflecting this arrangement are denoted as $\phi(a, b)$ for party A and $\psi(a, b)$ for party B. Given the institutional arrangement, rational parties play a Nash equilibrium of the game in normal form with strategy sets $A$ and $B$ and payoff functions $\phi(a, b)$ and $\psi(a, b)$.

As no formal structure is imposed on the strategy sets $A$ and $B$ (they may combine elements of discrete and continuous choice and they may be one- or multi-dimensional, including complete contingent plans under sequential choice) this model is quite flexible and allows for disparate applications in tort and contract law.

Throughout the paper, $(a^*, b^*)$ denotes an efficient reference profile which maximizes welfare, i.e.

$$(a^*, b^*) \in \arg \max_{(a, b) \in A \times B} w(a, b).$$

The institutional setting may add frictions but, in any case, maximum welfare remains an upper bound for the sum of payoffs such that the inequality

$$\phi(a, b) + \psi(a, b) \leq w(a^*, b^*)$$

will be fulfilled for any strategy profile $(a, b)$. At the efficient strategy profile,
however, frictions are ruled out by assumption, i.e.

\[
\phi(a^*, b^*) + \psi(a^*, b^*) = w(a^*, b^*) \tag{3}
\]

is assumed to hold as equality.

In an obvious sense, the underlying institutional arrangement is said to achieve the compensation goal for unilateral deviations from the efficient reference profile \((a^*, b^*)\) if the inequalities

\[
\phi(a^*, b^*) \leq \phi(a^*, b) \text{ and } \psi(a^*, b^*) \leq \psi(a, b^*) \tag{4}
\]

are met for all deviations \(b\) and \(a\) by the other party. The following efficiency result refers to the game with the payoff functions \(\phi\) and \(\psi\) and will be referred to as compensation principle.

**Proposition 1** (compensation principle) Suppose conditions (1) – (4) are met. Then the efficient reference profile \((a^*, b^*)\) forms a Nash equilibrium and all Nash equilibria (if more than one exists) are payoff equivalent.

**Proof.** By (2), (1), (4) and (3) it follows that

\[
\phi(a^*, b^*) \leq w(a^*, b^*) - \psi(a, b^*) \leq w(a^*, b^*) - \psi(a^*, b^*) = \phi(a^*, b^*)
\]

must hold and, hence, \(a^*\) is a best response by party A to \(b^*\). For similar reasons, \(b^*\) is also a best response by party B to \(a^*\) and, hence, the reference profile must be a Nash equilibrium indeed.

Suppose \((a^N, b^N)\) is any other Nash equilibrium consisting of mutually best responses. It then follows from (4) that

\[
\phi(a^N, b^N) \geq \phi(a^*, b^N) \geq \phi(a^*, b^*)
\]

as well as

\[
\psi(a^N, b^N) \geq \psi(a^N, b^*) \geq \psi(a^*, b^*)
\]

and hence, from (2) and (3), that

\[
w(a^*, b^*) \geq \phi(a^N, b^N) + \phi(a^N, b^N) \geq \phi(a^*, b^*) + \psi(a^*, b^*) = w(a^*, b^*)
\]

must hold. Yet, as the reference profile maximizes welfare (see (1)), none of the above inequalities can hold in the strict sense. In particular,

\[
\phi(a^N, b^N) = \phi(a^*, b^*) \text{ and } \psi(a^N, b^N) = \psi(a^*, b^*)
\]
must be true such that payoff equivalence is established as well. ■

Notice, as the Nash equilibrium will be efficient, in equilibrium, there is no
scope for voluntary renegotiations. Off the equilibrium, however, inefficien-
cies may lead to voluntary renegotiations. Yet, even such renegotiations (if
anticipated) would neither distort incentives nor affect equilibrium payoffs.

In fact, let \( \phi^r(a, b) \) and \( \psi^r(a, b) \) denote post renegotiation payoffs. As the
consent of both parties is needed, the participation constraints

\[
\phi(a, b) \leq \phi^r(a, b) \quad \text{and} \quad \psi(a, b) \leq \psi^r(a, b)
\]

have to be met for any strategy profile \( (a, b) \). At the efficient strategy profile,
there is no scope for renegotiations and, hence, payoffs \( \phi(a^*, b^*) = \phi^r(a^*, b^*) \)
remain the same such that condition (3) is valid for post renegotiation payoff
functions as well. Moreover, for any unilateral deviation \( b \neq b^* \) by party B,

\[
\phi^r(a^*, b^*) = \phi(a^*, b^*) \leq \phi(a^*, b) \leq \phi^r(a^*, b)
\]

must hold such that the compensation goal for unilateral deviations by party
B is achieved with respect to party A’s post renegotiation payoff function.
For symmetry reasons, the same holds true for unilateral deviations by party
A and, hence, condition (4) is met for post renegotiation payoff functions as
well.

As maximum welfare remains, of course, an upper bound for the sum of
payoffs even post renegotiations, condition (2) is also met and, hence, the
compensation principle applies equally well to the game with payoff functions
\( \phi^r \) and \( \psi^r \). Therefore, the efficient profile remains to be a Nash equilibrium
of the game with payoff functions \( \phi^r(a, b) \) and \( \psi^r(a, b) \) and all Nash equilibria
of this game (if more than one exists) must be payoff equivalent.

3 Damages in line with the difference hypothesis

The assumptions needed for the compensation principle will, in particular,
be fulfilled if damages are granted in line with the difference hypothesis in
the following sense. Differences are taken relative to an obligation profile
\( (a^o, b^o) \). Depending on the actually chosen strategy profile \( (a, b) \), party A
receives net damages \( d(a, b) \) from party B (if \( d(a, b) < 0 \), party B receives damages \(-d(a, b)\) from party A), leading to payoff functions

\[
\phi_d(a, b) = m(a, b) + d(a, b) \text{ and } \psi_d(a, b) = n(a, b) - d(a, b)
\]

of party A and B, respectively (subscript \( d \) refers to difference hypothesis).

To begin with, suppose both parties stick to the efficient reference profile \((a^*, b^*)\) (i.e., both commit efficient breach). Compared with B’s obligation \( b^o \), party A suffers a loss of \( m(a^o, b^o) - m(a^*, b^*) \) whereas, compared with A’s obligation \( a^o \), party B suffers a loss of \( n(a^o, b^*) - n(a^*, b^*) \). Therefore, to be in line with the difference hypothesis at the efficient profile, (net) damages

\[
d(a^*, b^*) = [m(a^*, b^o) - m(a^*, b^*)] - [n(a^o, b^*) - n(a^*, b^*)] \tag{5}
\]

are awarded to A if both parties breach efficiently.

Moreover, in order to achieve the compensation goal for unilateral deviations from the efficient reference profile,

\[
d(a, b^*) \leq [m(a^*, b^o) - m(a^*, b^*)] - [n(a^o, b^*) - n(a, b^*)] \tag{6}
\]

for all \( a \neq a^* \) as well as

\[
[m(a^*, b^o) - m(a^*, b)] - [n(a^o, b^*) - n(a^*, b^*)] \leq d(a^*, b) \tag{7}
\]

for all \( b \neq b^* \) must hold. Notice, the term \( m(a^*, b^o) - m(a^*, b^*) \) in (6) captures claims of party A for B’s deviation from \( b^o \), not at the decision \( a \) as actually taken by A, but at the efficient decision \( a^* \) as would have been taken by a reasonable person in the position of A. Referring to a reasonable person standard captures the idea that A’s claims are reduced for contributory negligence if, in the absence of such a reduction, they would exceed the difference \( m(a^*, b^o) - m(a^*, b^*) \) based on the reasonable person standard. A similar interpretation justifies the term \( n(a^o, b^*) - n(a^*, b^*) \) in (7).

The next proposition deals with Nash equilibria of the game in normal form with A and B as the strategy sets and \( \phi = \phi_d \) and \( \psi = \psi_d \) as payoff functions and will be referred to as difference principle.

**Proposition 2** (difference principle) Suppose damages \( d(a, b) \) are granted in line with the difference hypothesis relative to the obligation profile \((a^o, b^o)\) and based on a reasonable person standard according to the efficient reference
profile \((a^*, b^*)\) (i.e. \((5) - (7)\) are met). Then the compensation goal relative to the efficient reference profile is achieved and, hence, the efficient reference profile forms a Nash equilibrium. Moreover, all Nash equilibria (if more than one exists) are payoff equivalent.

**Proof.** The proof is straightforward. Consider any deviation \(a \neq a^*\). It then follows from \((6)\) and \((5)\) that

\[
\psi_d(a, b^*) \\
\geq n(a, b^*) - [m(a^*, b^*) - m(a^*, b^*)] + [n(a^*, b^*) - n(a, b^*)] = \\
= n(a^*, b^*) - [m(a^*, b^*) - m(a^*, b^*)] = \\
= n(a^*, b^*) - [n(a^*, b^*) - n(a, b^*)] - d(a, b^*) = \psi_d(a, b^*)
\]

and, hence, party B is compensated for deviations from the efficient reference profile by party A.

For symmetry reasons, party A is also compensated for unilateral deviations by B and, hence, \((4)\) holds indeed. The remaining claims then immediately follow from the compensation principle. \(\blacksquare\)

The existing literature on the economic analysis of tort law mainly deals with the case where the efficient reference profile and the obligation profile coincide, i.e. \((a^*, b^*) = (a^o, b^o)\). Under such circumstances, damages in line with the difference hypothesis have to fulfill \(d(a^o, b^o) = 0\) as follows from \((5)\). Moreover, for \(a \neq a^o\),

\[
d(a, b^o) \leq - [n(a^o, b^o) - n(a, b^o)]
\]

and, for \(b \neq b^o\),

\[
[m(a^o, b^o) - m(a^o, b)] \leq d(a^o, b)
\]

must hold as follows from \((6)\) and \((7)\).

As a prominent example, consider the accident model where party A is the victim whose expected value of assets amounts to \(m(a, b) = -a - \varepsilon(a, b) \cdot L\) and where party B is the injurer with value of assets \(n(a, b) = -b\). In the accident model, \(a\) and \(b\) denote precaution expenditures of the two parties, \(\varepsilon(a, b)\) captures the probability of an accident and \(L\) is the loss of fixed size which party A suffers from if an accident has occurred.

In this example, the externality is of unidirectional nature (as B’s value function \(n = n(b)\) does not depend on A’s decision \(a\)) and, hence, the unilateral specification \(d(a, b^o) = 0\) for all \(a\) would be consistent with the difference
hypothesis. All traditional damages rules (except strict liability) such as the negligence rule and negligence with a defense of contributory negligence satisfy (8). It then follows from the difference principle that all these damages rules generate two-sided efficient incentives under the provision that the obligation profile is equal to the efficient profile. While these efficiency claims are well-known from the literature, the above proof offers a common principle behind all of them.

The same principle also allows to handle sequential choice. Sequential choice is usually modelled in extensive form. By looking at the associated normal form, however, the difference principle still applies. In the next section, a setting of an unidirectional externality is confronted with the difference principle. The remaining sections will then deal with bidirectional externalities as regularly studied in the economic analysis of contract law and with respect to obligation profiles that fail to be efficient.

4 Sequential choice

As a first application of the difference principle to a setting of sequential choice, the following game in extensive form is examined. Suppose it is party B who, at stage 1, takes a decision \( b \in B \) before party A, after having observed the decision \( b \), chooses an action \( x \) from \( X \) at stage 2. The values of A’s and B’s assets amount to \( M(x, b) \) and \( n(b) \), respectively. The externality is of unidirectional nature as party B’s value function \( n(b) \) does not depend on party A’s decision.

Expressed in normal form, party B still chooses \( b \in B \) but party A now chooses a complete contingent plan \( a \in A = X^B \). Such a plan consists of a function \( a : B \rightarrow X \) with the interpretation that, if party B has chosen \( b \), party A responds with the action \( x = a(b) \). Expressed in normal form, party A’s value function amounts to \( m(a, b) = M(a(b), b) \) whereas party B’s value function is still \( n(b) \) as the externality is of unidirectional nature and as B moves first.

In the extensive form, there exist many efficient reference profiles. Let me focus on the ex post efficient one that can be obtained by backward induction.
as follows. For any \( b \), let

\[
a^*(b) \in \arg \max_{x \in X} M(x, b) + n(b) = \arg \max_{x \in X} M(x, b)
\]

denote a socially best response. Due to the unidirectional nature of the externality, any socially best response maximizes party A’s value function \( M(x, b) \) at the same time.

Anticipating this socially best response, the efficient decision at the first stage solves

\[
b^* \in \arg \max_{b \in B} M(a^*(b), b) + n(b) = \arg \max_{b \in B} m(a^*, b) + n(b).
\]

In the following, the ex post efficient profile \((a^*, b^*)\) serves both as the efficient reference profile as well as the obligation profile. Moreover, for later reference, let \( x^* = a^*(b^*) \in X \) be the socially best response to party B’s efficient decision \( b^* \).

If at all, courts will have to quantify damages ex post when, by assumption, courts can observe the actual decisions \( x \) and \( b \) but not party A’s complete contingent plan \( a \). If \( D(x, b) \) denotes the damages claims of party A as a function of decisions \( x \) and \( b \), then damages expressed in the normal form (as needed for the difference principle) amount to \( d(a, b) = D(a(b), b) \).

To be in line with the difference hypothesis (see (5)),

\[
d(a^*, b^*) = D(x^*, b^*) = 0
\]

must hold as reference and obligation profile coincide in the present application. Moreover, for deviations \( a \neq a^* \) (see (6)),

\[
d(a, b^*) = D(a(b^*), b^*) \leq 0
\]

as the externality is of unidirectional nature (i.e. \( n(a^o, b^*) = n(a, b^*) \)) and, for \( b \neq b^* \) see (7)),

\[
M(a^*(b^*), b^*) - M(a^*(b), b) \leq d(a^*, b) = D(a^*(b), b)
\]

must hold to remain in line with the difference hypothesis relative to the ex post efficient obligation profile \((a^*, b^*)\).

There exist several quantifications of damages \( D(x, b) \) that fulfill the above conditions. The following one would be among them. If party B
has taken the efficient decision \( b^* \), he will never be liable, i.e. \( D(x, b^*) = 0 \) independent of party A’s action \( x \). This specification ensures that conditions (9) and (10) are already met, both in fact as equalities.

If, however, party B deviates by taking an inefficient decision \( b \neq b^* \) but party A responds with a socially best response \( x = a^*(b) \), damages

\[
D(x, b) = M(x^*, b^*) - M(x, b)
\]

ensure that condition (11) is met also as an equality.

Notice, if party A does not respond to B’s deviation in a socially best way, the difference principle does not restrict the specification of damages \( D(x, b) \) at all such that we are free to put the already defined \( D(a^*(b), b) \) as a cap

\[
D(x, b) \leq D(a^*(b), b)
\]
on them.

With and without such a cap, it follows from the difference principle, that the ex post efficient reference profile \((a^*, b^*)\) forms a Nash equilibrium of the normal form game. The above cap, however, ensures that the ex post efficient reference profile must even be a subgame perfect equilibrium of the game in extensive form.

In fact, due to this cap, the socially best response \( a^*(b) \) maximizes party A’s damages claims \( D(x, b) \). Moreover, due to the unidirectional nature of the externality, the socially best response \( a^*(b) \) also maximizes her value function \( M(x, b) \) and, hence, the socially best response \( a^*(b) \) will be a best response of party A to any \( b \), even off the equilibrium path.

Since the ex post efficient reference profile forms a Nash equilibrium of the normal form game and since, due to the above cap on damages, the socially best response is a best response of party A to any choice \( b \) by party B, this reference profile must be a subgame perfect equilibrium of the game in extensive form indeed.

Let me point out that this cap allows for a neat legal interpretation. If party B has violated his obligation, in most legal regimes, it would be party A’s duty to minimize damages by maximizing the remaining value \( M(x, b) \) of her assets. If she fails to do so, damages could be reduced for contributory negligence.
The German civil code (§ 254 BGB), e.g., requires that where fault (including the fault of failing to reduce the damage) on the part of the injured person contributes to the occurrence of the damage, the extent of compensation to be paid depend on the circumstances, in particular to what extent the damage is caused mainly by one or the other party.

As a final comment, the above damages regime is also related to marginal cost liability LMC, introduced by Wittman (1981). Surprisingly enough, Wittman himself did not find it worthwhile to examine what, in his notation, would correspond to the negligence version of LMC explicitly. Yet, it is exactly this version of marginal cost liability that can be justified in terms of the difference hypothesis combined with the appropriate reduction of damages for contributory negligence.

5 Non-contingent contracts

As a second application of the difference principle, the following hold-up situation is examined. At stage 0, parties A and B agree on a simple (i.e. non-contingent) contract involving a binary performance decision to be taken by party B at stage 3. After having signed the contract, parties A and B decide, at stage 1, on investments \( a \in A \) and \( y \in Y \), respectively. Investment costs \( g(a) \) are borne by party A, investment costs \( h(y) \) by party B. At the investment stage (i.e. stage 1), benefits from investments are still uncertain as, not until stage 2, nature randomly selects party A’s utility \( v \) from and party B’s cost \( c \) of performance. At stage 3, party B takes the (binary) performance decision \( q \in \{0, 1\} \). The payoff functions net of investment costs but before damages amount to \( v \cdot q - g(a) - T \) and \( T - c \cdot q - h(y) \) for party A and B, respectively, where \( T \) is the price as stipulated in the contract.

To make use of the difference principle, the sequential interaction must be expressed in normal form. Party A still chooses investments \( a \in A \) whereas party B chooses \( b = (y, \eta) \in Y \times H \) where \( y \) denotes his investments and \( H \) denotes the set of all complete contingent performance plans \( \eta \) that are feasible. Feasibility depends on the informational setting.

To begin with, think of a state \( \omega \) that is randomly drawn by nature at
stage 2 such that utility and costs of performance are deterministic functions

\[ v = V(a, y, \omega) \quad \text{and} \quad c = C(a, y, \omega) \]

of investments \((a, y)\) and the state \(\omega\). Moreover, in the present section, investments and the state are assumed observable.

Given this informational setting, a complete contingent performance plan consists of a function \(\eta : A \times Y \times \Omega \rightarrow \{0, 1\}\) with the interpretation that \(\eta(a, y, \omega) \in \{0, 1\}\) denotes the performance decision taken by party B if confronted with investment profile \((a, y)\) and state \(\omega\). The value functions of party A and B in the normal form are

\[
\begin{align*}
    m(a, y, \eta) &= E[V(a, y, \omega) \cdot \eta(a, y, \omega)] - g(a) - T \\
n(a, y, \eta) &= T - E[C(a, y, \omega) \cdot \eta(a, y, \omega)] - h(y),
\end{align*}
\]

respectively, where expectations are taken with respect to the exogenously given distribution of the state \(\omega\).

Due to sequential choice, there exist many candidates that could serve as efficient reference profile (performance decisions off the equilibrium path do not affect the outcome). I focus on the following one. For given investments \((a, y)\) and state \(\omega\), let

\[
\eta^+(a, y, \omega) \in \arg \max_{q \in \{0, 1\}} [V(a, y, \omega) - C(a, y, \omega)] \cdot q
\]

denote the (if there exists more than one, an) ex post efficient performance decision. Anticipating the ex post efficient performance decision, efficient investments solve

\[
(a^*, y^*) \in \arg \max_{(a,y) \in A \times Y} m(a, y, \eta^+) + n(a, y, \eta^+).
\]

The ex post efficient profile \((a^*, y^*, \eta^+)\) would, of course, be a candidate for the reference profile.

Instead, in the present section, I consider the profile \((a^*, y^*, \eta^*)\) with the performance decision \(\eta^*(\omega) = \eta^+(a^*, y^*, \omega)\) that would be ex post efficient at efficient investments \((a^*, b^*)\) but may fail to be so at other investment choices. Since party B knows the state \(\omega\), the performance decision \(\eta^*(\omega)\) would be feasible for the informational setting of the present section.
Notice, if parties stick to efficient investments, the outcome will be the same as under the ex post efficient profile and, hence, \((a^*, y^*, \eta^*)\) is an efficient profile of the normal form game as well, its crucial property being that the performance plan \(\eta^*\) does not depend on party A’s actual investments \(a\). As it turns out, this reference profile \((a^*, y^*, \eta^*)\) allows aligning earlier findings with the difference principle.

Because parties have signed a non-contingent contract requiring B always to perform, the relevant obligation profile is \((a^*, y^*, \eta^o)\) with efficient investments and the unconditional duty \(\eta^o \equiv 1\) to perform. Since performance may turn out to be inefficient ex post, this obligation profile is inefficient with respect to party B’s performance obligation \(\eta^o\) whereas A’s obligation is efficient, i.e. \(a^o = a^*\).

In the present section, courts are assumed to know the state \(\omega\) such that they will be able to calculate counterfactuals. But they still cannot observe B’s complete contingent performance plan and, hence, damages \(D(a, y, \omega, q)\) can only depend on the performance decision \(q\) actually taken by party B but not on his complete contingent plan \(\eta\).

The difference principle refers to the associated normal form. Expected damages amount to

\[
d(a, y, \eta) = E [D(a, y, \omega, \eta(a, y, \omega))]
\]
as a function of the strategy profile \((a, y, \eta)\) in the normal form game.

To be in line with the difference hypothesis (i.e. to satisfy (5) – (7)), the following conditions must be met in expected terms (recall, that party A’s obligation \(a^o = a^*\) is efficient):

\[
E [D(a^*, y^*, \omega, \eta^*(\omega))] = E [V(a^*, y^*, \omega) \cdot (1 - \eta^*(\omega))],
\]

\[
E [D(a, y^*, \omega, \eta^*(\omega))]
\]

\[
\leq E [V(a^*, y^*, \omega) \cdot (1 - \eta^*(\omega))] + E [((C(a^*, y^*, \omega) - C(a, y^*, \omega)) \cdot \eta^*(\omega)]
\]
for all \(a \neq a^*\), and

\[
E [V(a^*, y^*, \omega)] - E [V(a^*, y, \omega) \cdot \eta(a^*, y, \omega)] \leq E [D(a^*, y, \omega, \eta(a^*, y, \omega))]
\]
for all \(b = (y, \eta) \neq b^* = (y^*, \eta^*)\). Constraint (13) follows from the fact that the performance decision \(\eta^*(\omega)\) of the reference profile remains independent
of party A’s actual investments $a$. Otherwise, this constraint would look more complicated.

Again, constraints (12) – (14) leave some degrees of freedom as far as the quantification of damages is concerned. For example, damages

$$D(a, y, \omega, q) = V(a^*, y^*, \omega) - [V(a^*, y, \omega) + C(a, y^*, \omega) - C(a^*, y^*, \omega)] \cdot q$$

would ensure that constraints (12) – (14) are met even with equality. For this specification of damages, in particular, it follows from the difference principle that the efficient reference profile $(a^*, y^*, \eta^*)$ forms a Nash equilibrium of the game in normal form and all Nash equilibria (if more than one exists) are payoff equivalent.

This efficiency result is in the spirit of efficient expectation damages (see Cooter, 1985). It also generalizes findings of Hewer and Goeller (2014) who have studied efficient incentives for takings in a setting with purely selfish investments, where the utility $V(a, \omega)$ from performance does not depend on party B’s investments and the costs $C(y, \omega)$ of performance do not depend on party A’s investments. In this special case, the above damages regime simplifies to

$$D(a, y, \omega, q) = V(a^*, \omega) \cdot (1 - q)$$

as the externality is of essentially unidirectional nature.

It also generalizes Schweizer (2006) on one-sided cooperative investments, where party A’s investments only affect the costs $C(a, \omega)$ of performance of party B. In this case, the above damages regime simplifies to

$$D(a, \omega, q) = V(\omega) \cdot (1 - q) - [C(a, \omega) - C(a^*, \omega)] \cdot q$$

such that the externality remains of bidirectional nature in spite of one-sided investments.

To quantify such damages in the general case, courts must observe actual investments $a$ and $y$, the actual performance decision $q$ and, most demanding, the state $\omega$ of nature. In particular, if party A has deviated by investing $a \neq a^*$, courts would need to know the counterfactual utility $V(a^*, y^*, \omega)$ from performance. The next section deals with quantifying damages in an informational setting where counterfactuals remain unknown. As it turns out, it is still possible to generate efficient incentives.
6 Counterfactuals unobservable

In the present section, the following modification of last section’s model is considered. At stage 1, parties A and B still reach their investment decisions \( a \in A \) and \( y \in Y \). A state of nature, however, is not explicitly introduced. Rather, investments directly affect the distribution of the actual utility \( v' \in V \) and costs \( c' \in C \) of performance, which become known at stage 2. The sets \( V \) and \( C \) denote the ranges from which \( v' \) and \( c' \) are realized and \( F(v, c, a, y) \) denotes their common distribution as a function of investments.

Before party B decides on performance at stage 3, he only learns the actual utility \( v' \) and costs \( c' \) of performance such that a complete contingent performance plan consists now of a mapping \( \eta : V \times C \to \{0, 1\} \) with the interpretation that \( \eta(v', c') \) is the performance decision which party B will take if confronted with \( v' \) and \( c' \).

Let \( \eta^+(v', c') \) denote the ex post efficient performance decision if, at stage 2, utility \( v' \) and costs \( c' \) of performance have been realized. It is defined as \( \eta^+(v', c') = 1 \) if \( c' \leq v' \) and \( \eta^+(v', c') = 0 \) else. Notice, since \( v' \) and \( c' \) are observable at stage 2, the ex post efficient performance plan would be feasible for the informational setting at hand.

In the normal form of the game, party A chooses investments \( a \in A \) whereas party B chooses \( b = (y, \eta) \in B = Y \times \{0, 1\}^{V \times C} \) where \( y \in Y \) denotes his investments and \( \eta \in \{0, 1\}^{V \times C} \) his complete contingent performance plan. The value functions in the normal form as a function of the strategy profile \((a, y, \eta)\) are

\[
m(a, y, \eta) = E[v \cdot \eta(v, c)]_{(a, y)} - g(a) - T
\]

and

\[
n(a, y, \eta) = T - E[c \cdot \eta(v, c)]_{(a, y)} - h(y)
\]

for party A and B, respectively. Subscript \((a, y)\) indicates that the distribution function and, hence, the expectation operator depend on the actually chosen investment profile \((a, y)\).

Anticipating the ex post efficient performance decision \( \eta^+ \), the efficient investment profile solves

\[
(a^*, y^*) \in \arg \max_{(a, y) \in A \times Y} m(a, y, \eta^+) + n(a, y, \eta^+).
\]
In the present section, the ex post efficient profile \((a^*, y^*, \eta^+)\) will serve as reference profile. As in the previous section, however, the non-contingent profile \((a^*, y^*, \eta^0 \equiv 1)\) with the unconditional duty \(\eta^0 \equiv 1\) to perform is still taken as obligation profile.

Damages have to be quantified (if at all) ex post. They may depend on actual investments, the realized utility and costs of performance and the actual performance decision \(q\) but not on the complete contingent plan underlying party B’s strategy in the normal form of the game. Let \(D(a, y, v', c', q)\) denote these damages as a function of the observed entities. In the normal form, they give rise to expected damages

\[
d(a, y, \eta) = E[D(a, y, v, c, \eta(v, c))]_{(a, y)}
\]

if party B makes use of performance plan \(\eta\). Subscript \((a, y)\) indicates again, that the distribution of \(v\) and \(c\) and, hence, the expectation operator depend on actual investments.

To be in line with the difference hypothesis (i.e. to satisfy (5) – (7)), the following conditions must be met in expected terms:

\[
E[D(a^*, y^*, v, c, \eta^*(v, c))]_{(a^*, y^*)} = E[v \cdot (1 - \eta^*(v, c))]_{(a^*, y^*)}, \quad (15)
\]

\[
E[D(a, y^*, v, c, \eta^*(v, c))]_{(a, y^*)} \leq E[v \cdot (1 - \eta^*(v, c))]_{(a^*, y^*)} + E[c \cdot \eta^*(v, c)]_{(a^*, y^*)} - E[c \cdot \eta^*(v, c)]_{(a, y^*)}
\]

for all \(a \neq a^*\) but \(b = (y^*, \eta^*)\) and

\[
E[v]_{(a^*, y^*)} - E[v \cdot \eta(v, c)]_{(a^*, y^*)} \leq E[D(a^*, y, v, c, \eta(v, c))]_{(a^*, y)} \quad (17)
\]

for all \(b = (y, \eta) \neq b^* = (y^*, \eta^*)\) but \(a = a^*\).

Again, constraints (15) – (17) allow for some flexibility as far as quantifying damages ex post is concerned. For example, damages

\[
D = D(a^*, v', q) = E[v]_{(a^*, y^*)} - v' \cdot q
\]

if party A has invested efficiently and

\[
D = D(a, c', q) = E[v \cdot (1 - \eta^*(v, c)) + c \cdot \eta^*(v, c)]_{(a^*, y^*)} - c' \cdot q
\]

if party B has invested efficiently, but A has not (i.e. \(a \neq a^*\)), would ensure that constraints (15) – (17) are met even as equalities.
To implement these damages, courts must know whether party A has invested efficiently or not. Moreover, if A has invested efficiently, courts must know the actual utility $v'$ from performance and the actual performance decision $q$ by party B. If A has deviated from efficient investments, courts must know the actual costs $c'$ of performance and the actual performance decision. But they need not know by how far party A has deviated from the efficient level. In both cases, neither counterfactuals nor party B’s investments need be known to quantify the above damages.

Since constraints (15) – (17) are met, the ex post efficient reference profile $(a^*, y^*, \eta^+)$ forms a Nash equilibrium of the normal form game as immediately follows from the difference principle.

It need not correspond to a subgame perfect equilibrium of the game in extensive form though. Yet, any subgame perfect equilibrium would correspond to a Nash equilibrium in the associated normal form and, hence, would be payoff equivalent to the ex post efficient reference profile nonetheless.

Alternatively, one might think of parties to renegotiate, off the equilibrium path, voluntarily. If parties anticipate such renegotiations, efficient incentives will still prevail and expected payoffs remain unchanged for reasons spelled out in section 2.

### 7 Conclusion

This is not a paper on the design of sophisticated mechanisms under asymmetric information but rather an economic analysis entirely driven by the legal principle behind the difference hypothesis. From the legal perspective, compensation of the creditor is considered as the primary goal of obligation law whereas the economic analysis sees the efficiency of incentives as the ultimate goal. The compensation principle provides a convenient link between the two goals.

Damages in line with the difference hypothesis even relative to an inefficient obligation profile ensure that the compensation goal relative to an efficient reference profile is met, provided that a cap on damages claims by the creditor is imposed whenever the debtor has breached efficiently. Such a damages regime ensures that the compensation goal relative to the efficient reference profile is achieved and hence, by the compensation principle, the
regime generates efficient incentives for both parties.

Establishing the efficiency of incentives with the aid of the compensation principle requires much weaker assumptions than the first order approach, predominant in other studies. To be sure, achieving the compensation goal is only a sufficient, not a necessary condition for efficient incentives. In fact, there exists efficiency results where the first order approach works but the compensation principle does not (see, e.g., Edlin and Reichelstein (1996), Ohlendorf (2009) or Stremitzer (2012)). But such efficiency results are of a less robust nature as they require more restrictive assumptions than the difference principle.

8 References


