Duress

Péter Cserne, Tilburg University
Duress in Contracts: An Economic Analysis


Péter Cserne®

Abstract
The purpose of this paper is to provide an overview of the economic analysis of contractual duress. The focus will be on the distinctive features of the economic perspective on the duress doctrine, as developed in the theoretical literature of law and economics. Before discussing the results of economic analysis, the legal background and some non-economic theories of duress are briefly presented.

Keywords: contact formation, holdup, duress, coercion, necessity, economic duress

JEL Codes: K00, K12

1. Introduction: economists learning from law?
In the law of contracts, duress is an excuse for non-performance. Virtually all modern legal systems provide that a party who concluded a contract under unacceptable pressure has a right to refuse performance or, if he has already performed, a right to restitution. In short, victims of contractual duress are allowed to avoid the offending agreement. The purpose of this paper is to provide an overview of the economic analysis of contractual duress. The focus will be on the distinctive features of the economic perspective on the duress doctrine, as developed in the theoretical literature of law and economics. Before discussing the results of economic analysis, the legal background and some non-economic theories of duress are briefly presented. This detour may be justified by an argument put forward by Robert Cooter and Thomas Ulen in the introductory chapter of their textbook. As they argue, the voluntariness of contracts is a subject on which economists should learn from the law. “[E]conomists frequently extol the virtues of voluntary exchange, but economics does not have a detailed account of what it means for exchange to be voluntary. [...] [C]ontract law has a complex, well-articulated theory of volition. If economists will listen to what the law has to teach them, they will find their models being drawn closer to reality.” (Cooter – Ulen 2008: 12)

Although this gesture of two economists towards lawyers is to be appreciated, it cannot be taken at face value. First, large overlaps notwithstanding, there are as many different “theories of volition” as national (as well as sub- and supranational) contract laws. More

® Senior Research Fellow, Tilburg Law and Economics Centre (TILEC) and the Faculty of Law, University of Tilburg, The Netherlands. The author can be contacted at peter.cserne@uvt.nl.
importantly, the “theory of volition” of contract law, well-articulated as it may be, is not well-founded theoretically. The legal understanding of concepts, such as voluntariness usually half-knowingly reflects philosophical or (pre-)scientific standpoints of earlier ages (“the metaphysics of the Stone Age” Hart – Honoré 1985: 2) or expresses some folk-psychological notions. On the other hand, what law as an institutionalized practice needs for its everyday operation is not a theoretically sound definition but a uniform understanding of voluntariness. Even naive or common sense “theories” are able to provide this, at least in so-called easy cases. Still, what the law says about voluntariness could be called a “theory” only in this practical sense. A further difficulty is that this “legal worldview” is not of much help in policy design. From a purely legal perspective, it is impossible to answer normative questions as to how the legally required degree of voluntariness of contract formation should be regulated. As we shall see, law and economics scholars have made considerable effort to re-conceptualize contractual duress in order to make it both amenable for the analytical tools and subject to the evaluative standards of economics.

Alongside Cooter and Ulen, commentators on comparative law and economics have also argued that economic analysis should take doctrinal details and the diversity of legal rules seriously. As they suggest, confronting details and diversity may improve economic analysis not merely by “drawing the models closer to reality” and thus increasing their practical relevance, but also by delivering empirical data for testing hypotheses, serving as a source of inspiration for new theories, and allowing the correction of the analyst’s home-country bias (De Geest–Van den Bergh 2004: xiii).

2. Law, doctrine, and philosophy
This section provides an overview of the legal rules on duress and briefly comments on the relevant doctrinal and philosophical literature.

2.1 Legal rules and doctrines
Virtually all legal systems impose threshold conditions for the making of enforceable contracts (Kötz 1997: 136–7, 209–213; Zweigert – Kötz 1998: 428–430, Probst 2001, Fabre-Magnan – Sefton-Green 2004, Schindler 2005). Even the rigid formalities of ancient Roman law, summarised in the maxim voluntas coacta tamen voluntas est (coerced will is nevertheless a will; Paulus D. 4.2.21.5), allowed for legal action against the party who forced another into a transaction by tortuous or criminal conduct (Hartkamp 1971; Zimmermann 1990: 651–662). Besides incapacity and fraud, the most common kind of abuse of the bargaining process is when the conduct of one of the contracting parties has been subject to threat or actual coercion. This may impair the enforceability of the resulting agreement.

In order to set aside the contract for duress under common law, traditionally a threat of bodily injury had to be established. Later, inspired by the equity doctrine of undue influence, courts widened the scope of the doctrine so that nowadays duress of a non-physical nature (i.e. affecting economic or reputational interests) suffices for the excuse to hold. In the United States, section 175(1) of the Restatement (Second) of Contracts provides: “If a party’s manifestation of assent is induced by an improper threat by the other party that leaves the victim no reasonable alternative, the contract is voidable by the victim.” A similar notion can be found in the Unidroit Principles (1994) for international commercial contracts. Article 3.9 permits the avoidance of a contract on the ground of a threat which is unjustified and “imminent and serious” so as to leave the
other party no reasonable alternative. As in modern English law, the threat does not need to be made against person or property but may also affect the reputation or purely economic interests of the other party. Similar provisions can be found in national contract rules of civil law jurisdictions under varied names. The actual terminology refers to this defect of consent as illegal threat (e.g. in Germany, the Netherlands, and Greece), violence (in France, Italy, and Spain), well-grounded fear (in Switzerland, Austria), or force and fear (in Scotland).

Duress can arise in two main contexts: pre-contractual negotiations and contract modification. In pre-contractual negotiations, it is a routine bargaining tactic to exert some pressure on the other party. For the legal system to disown the conduct of the contracting party and the contract to be set aside, this pressure has to be illegal or illicit, as well as of a certain magnitude. For example, threatening to abuse a dominant position in the market (as far as contract law is concerned); to beat or kill someone; to impound the goods of someone; or to get someone put behind bars by perjury, are unacceptable bargaining tactics which make the resulting contract voidable or void (Kötz 1997, Eidenmüller 2007).

Duress cases also concern the modification of contracts when one of the parties to an existing contract wants it modified, and backs his wish with a threat that otherwise he will not perform at all. If the threat is credible and in fact acted upon, the other party has the option to sue for breach of contract. But often this is not a realistic or expected response of the coerced party. Rather, he gives in to the pressure, agrees to the modification, and later wants to be excused from performing under the modified terms. Courts then take into account a number of circumstances in deciding whether he is allowed to avoid his agreement on the ground of duress (Kötz 1997, Hillman 1979). An ex post consideration is given to such diverse circumstances as whether he gave in under protest or without struggle; whether he (alternatively, an average person or a person with exceptional firmness) could have been expected to stand firm and sue; the time allowed to consider the suggested modification; his acquiescence or otherwise after the modification; and the reasonableness of the modification. Common law courts traditionally also had to inquire whether there was “fresh consideration” for the new promise. As discussed later in this paper (section 3.7) as well as elsewhere in this Encyclopaedia, the extensive economic literature on the hold-up problem and contract renegotiation suggests rather different criteria for determining whether the modification should be enforced.

The law usually imposes sanctions both when the coercer creates a desperate situation for the other party and when he merely exploits the necessity of the other (for this distinction see section 3.6 below). A special variant of the latter case is when the threat originates from an identifiable third party. This case is again regulated differently in national legal systems, depending on the state of mind (“good faith”) of the other party and the seriousness or imminence of the threat (Kötz 1997: 212–213).

As to the remedies, protection against duress is commonly afforded by allowing the abused party to avoid the transaction, restoring both parties to their pre-contractual positions. In other cases, the court does not suppress, rather modifies the contract. As we shall see below (section 3.6), economic analysis provides a justification for this distinction in remedies.

In sum, using a rather varied terminology, legal systems follow either a means oriented or a result oriented approach to duress, or combine the two. Means oriented systems focus on the conduct of the coercer (wrongdoer), especially on the nature of the threat as
illegal, unwarrantable, unlawful, or immoral. Result oriented systems focus on the effect of coercion on the coerced party (victim), and specifically on the quality of the fear induced; the reasonableness of the victim’s conduct in face of the threat; and the presence or absence of reasonable alternatives.

When judges provide relief or render a contractual clause unenforceable, their reasoning is often difficult to analyse even in terms of the doctrinal categories of the very system the court is bound to interpret. When it comes to practical applications, the function and domain of different legal doctrines is much less neatly separated than in theory. As doctrinal boundaries are neither clear nor uniform, this paper also covers, along with duress in the narrow sense, contract law doctrines which are sometimes used for similar purposes as duress. These different but related doctrinal categories include necessity (when one party takes advantage of the desperate situation of another), undue influence (one party abuses a position of trust or confidence), and certain types of procedural unconscionability (when the consent of one party is impaired). The doctrine of economic duress (the exploitation of necessity) will also be covered. This last doctrine was developed in England in the 18th century, to play a major role in the judicial control of contracts in the 20th century. As Dawson reconstructed (1947: 267–276), the main reason English courts developed the common law doctrine of economic duress was to bar the enforcement of credit contracts of “expectant heirs” (young aristocrats with a tendency to spend over their means), thus hindering their ability to take up large credits with their family estate as security. Initially, judges were motivated by and often made explicit reference to the policy purpose of conserving certain social structures through maintaining real estate in the hands of the aristocracy. Later, especially since the mid-20th century, economic duress has been used to void contracts which, in the court’s view, resulted from “structural inequality” or the “inequality of bargaining power”. Below (section 3.9) we will focus on this broader interpretation of economic duress.

2.2 Juristic (doctrinal) theories

The so-called “classical theory of contract” or “will theory” was based on the fundamental premise that a contract is the expression of the free will of two consenting individuals (Gordley 1991). The binding force of a contract derives from the mutual assent of the parties, i.e. “the meeting of their minds”. Starting from the 18th, dominant in the 19th and still influential well into the 20th century, the will theory held that contract is an expression of the concordant wills of two autonomous individuals.

In doctrinal literature, duress has been classified traditionally as a defect of consent (“overborne will” theory). The classical theory of contracts interpreted duress as one of the necessary conditions for the functioning of the actual free will of the parties. More sophisticated juristic theories accounted for the defects of voluntariness in contract formation such as mistake, fraud and duress as the “constitutive limits” of freedom of contract (Kennedy 1982, Stewart 1997).

More or less influenced by these theories, national contract laws developed elaborate rules for checking the voluntariness of contractual agreements. While full voluntariness is an unattainable ideal, the consent of an individual can be considered substantially voluntary if three conditions are fulfilled (Pope 2004: 711–713). The individual is (1) in possession of an abstract capability of making choices (even if the decisions are foolish, unwise, or reckless, they can still be attributed to an autarchic subject); (2) substantially free from controlling external influences such as coercion, threat, or manipulation; (3) substantially free from epistemic defects such as ignorance of the nature of one’s
conduct or its foreseeable consequences. These conditions, in turn, can be linked to different legal doctrines regulating the validity of contracts. The rules of incapacity are supposed to regulate that only people being capable of making choices (in the above said abstract sense) can conclude a valid contract. The contract law rules on fraud, duress and undue influence serve to guarantee the lack of certain external controlling influences. Substantial freedom from epistemic defects is taken care of via rules on unilateral mistake, mandatory disclosure, and other rules making consent more deliberate. These rules, technically called formation defences, can be found among the basic contract provisions of practically every civil code and of the common law as well. However, as argued by many (starting with Savigny in the early 19th century, cf. Eidenmüller 2007: 23–24), coercion is not primarily a matter of voluntariness. Even in the textbook example of entering into contract under a your-money-or-your-life-type threat (signing an agreement at gunpoint), it is not the actual consent that is lacking: the victim of coercion is acting voluntarily in the sense of evaluating his reasons for action. If such a threat is illegal and the contract is void for duress this is because of a normative judgment about the quality of the choices available to the coerced party. Most coerced transactions are not lacking consent: rather they are instances when either consent is obtained improperly or the alternatives available for the coercee are such that consent has a different moral weight than under normal circumstances. The “overborne will” theory presents an implausible picture of what happens to a person subject to duress.

What voluntariness means in a given contract formation setting is not always easy to discern. Suppose there is full information, neither party is subject to cognitive deficiencies and the contract is complete. In such a case the question is whether the constrained choice set of one party renders his consent “involuntary”. If it does then practically every contract is “coerced” because of the scarcity of resources and opportunities. On the other hand, except for extreme cases such as actual physical force, psychic torture or hypnotic trance, almost every exchange can be viewed as voluntary in the psychological sense of reflecting the choice of the individual. Ultimately, the seemingly simple question of what constitutes voluntary consent is a not a factual but a normative matter. The question is how low or high the threshold of legally (or morally) acceptable constraints on individual choice should be set. Traditional duress doctrine sets this baseline relatively low so that only threats to the physical security of a contracting party are considered to be below the threshold. When threats to the reputation or the non-material interests of the party also establish duress, the threshold is drawn somewhat higher. If legal doctrines are not, is philosophy in the position to answer this normative question?

2.3 Philosophical (moral) theories

Political and legal philosophy have been traditionally interested in both conceptual and normative aspects of coercion, be it political or private. The current mainstream approach in analytical philosophy strives to conceptualize coercion as an action-directing action which eliminates or makes worse the alternatives available for the coerced party, thus changing his balance of reasons in a way preferred by the coercer. Starting with the classical article by Nozick (1969), philosophical analysis has been concerned with drawing a distinction between (illegitimate) threats and (legitimate) offers. While threats reduce the possibilities open to the recipient of the proposal, offers expand them. Philosophers concentrate their efforts on identifying those proposals or
directives which count as coercive and are, on this account, morally problematic (see e.g. Wertheimer 1987, Honoré 1990, Smith 1994, Bigwood 1996, Stewart 1997, Brady 1999, Taylor 2003, Owens 2007).

One of the difficulties with this approach is related to the specification of a baseline against which the proposal is to be measured. The positioning of this baseline is not self-evident. It may be statistical (what the offeree can reasonably expect), empirical or phenomenological (what he in fact expects) or moral (what he is entitled to expect). Whichever position is taken, “the distinction between threats and offers depends on whether it is possible to fix a conception of what is right and what is wrong, and to determine what rights people have in contractual relations independent of whether their contracts should be enforced” (Trebilcock 1993: 80). While philosophers suggest broader theories to answer these questions, economic analysts who are sceptical about this prospect criticise the philosophical approach to coercion as irrelevant, indeterminate or inconclusive, and suggest economic interpretations instead (Trebilcock 1993: 78–101, Craswell 1995, Bar-Gill – Ben-Shahar 2005).

3. Economic theories

The existing law and economics literature on duress focuses on two main questions. On the one hand, it provides an explication of the various rules and doctrines applied by courts, by identifying functions or goals they serve within contract law. In this explanatory mode, economic analysis is concerned with the impact of the duress excuse (under this or that interpretation) on efficiency and distribution. On the other hand, economic analysis also works out what efficiency as a normative criterion suggests as the best way to deal with those problems that legislators and courts are striving to solve with the help of duress and related doctrines. In this evaluative mode law and economics scholars criticise and eventually suggest changes to the existing legal rules so as to bring them closer to the normative goals contract law is supposed to serve.

3.1 The private ordering paradigm

Freedom of contract is an ideologically charged notion which attracts strongly-held political views amongst both defenders and critics. Modern Western legal orders attach a high value to freedom of contract by considering it a basic legal principle. Modern legal orders also set several limits to this freedom, going well beyond punctual exceptions. The primacy of autonomy has been traditionally supported by mainstream economic theory as well. In neoclassical economics the “predilection for private ordering over collective decision-making is based on a simple (perhaps simple-minded) premise: if two parties are to be observed entering into a voluntary private exchange, the presumption must be that both feel the exchange is likely to make them better off, otherwise they would not have entered into it.” (Trebilcock 1993: 7).

From an economic perspective, this presumption can be rebutted by identifying either a contracting failure or a market failure. These two kinds of failure provide an economic justification for the rules and doctrines of contract regulation (Cooter – Ulen 2008: 232–238). Contracting failures are problems of individual rationality. They are either cases of bounded rationality, addressed by the rules on (in)capacity or cases of constrained choice, addressed by the doctrines of duress, necessity, or impossibility. Market failures can be explained by three types of transaction costs and addressed in contract law accordingly. Negative externalities often justify the unenforceability of contracts which derogate public policy or violate a statutory duty. Failures deriving
from imperfect information are addressed as fraud, failure to disclose, frustration of purpose, or mutual mistake. The third type is structural or situational monopoly which leads to the lack of competition, and is addressed by doctrines such as necessity, unconscionability or lésion.

The operation of a modern market economy relies on freely negotiated enforceable contracts. This not only requires but implicitly assumes some “constitutive limits” on freedom of contract (Kennedy 1982), namely those minimal limitations which are necessary for the working of even a libertarian (unregulated) contract regime. As Milton Friedman put it: “The possibility of coordination through voluntary cooperation rests on the elementary – yet frequently denied – proposition that both parties to an economic transaction benefit from it, provided the transaction is bilaterally voluntary and informed.” (Friedman 1962: 13) This quote shows once again how one of the most basic insights of economics rests on a standard of voluntariness. As long as this standard cannot be substantiated within economic theory itself, it refers back to law, philosophy or common sense.


3.2 Economic “definitions” of duress: incentive-based theories
Duress thus provides a constitutive limit to the private ordering paradigm. From an economic perspective, it refers to one class of cases when a contract is most probably not welfare-enhancing and therefore should be presumably void.

In their textbook Cooter and Ulen (2008: 281–284) characterise duress along similar lines. In their view, duress is a threat to destroy value (“Pay me $3000 or I will shoot you.”). It should be distinguished from bargains where one party threatens not to create value (“$3000 for my car is my final offer, take it or I walk away.”). This distinction has two aspects. First, the two types of threat differ in what happens when bargaining fails. In the first case, failure to agree leads to destruction. In the second case, failure to reach a bargain results in a failure to create a cooperative surplus. Second, while successful bargains tend to create value, contracts concluded under duress tend to shift resources from one person to another. In sum, Cooter and Ulen suggest the following economic interpretation of duress: a promise should be enforceable if it was extracted as the price of one’s cooperation in creating value; a promise should be unenforceable if it was extracted by a threat to destroy value. In order to deter destructive threats, contracts made under duress should not be enforced.

Most economic theories are concerned with the incentive effects of the duress doctrine. By and large uncontroversial are the cases of actual or threatening physical or psychic coercion by the other party. Here, legal rules, moral intuitions, and economic theory equally suggest the non-enforcement of such contacts, be it for retribution, prevention or for some other reason. In economic terms, voiding a contract concluded at gunpoint is efficient because it maximizes social welfare. “A points a gun at B saying, ‘Your
money or your life”; B accepts the first branch of this offer by tendering his money. But
a court will not enforce the resulting contract. The reason is not that B was not acting
of his own free will. On the contrary, he was extremely eager to accept A’s offer. The
reason is that the enforcement of such offers would lower the net social product by
channelling resources into the making of threats and into efforts to protect against
them.” (Posner 2007: 115)

The main argument of the incentive theory is this: the non-enforcement of contracts
concluded under duress deters potential coercers from threatening with socially wasteful
coercion, as well as economizes on preventive measures on the part of potential victims
of coercion. In terms of this incentive theory of duress, anti-duress rules have the
function of discouraging parties from taking excessive and wasteful precautions against
being subjected to extortionate contractual terms. In still other words, the economic
justification of the duress doctrine “is found in the phenomenon of rent seeking: [...] we
wish to discourage investments in coercion or against coercion” (Hermalin–Katz–
Craswell 2007: 54).

In general, rules against duress “guarantee that resources remain in the hands of the
highest-value user.” (Esposto 1999: 145) Even if involuntary transfers might improve
the allocation of resources in particular cases, if they were tolerated or supported by law
as a rule, this would induce people to insure and defend themselves against coercion
privately. This, in turn, would be undesirable because it would be more costly than
public law enforcement. The incentive effects generated by the non-enforcement of
certain contracts provide the ultimate criterion as to which threats should or should not
be discouraged. By assessing the welfare effects of investments in threats and in
precaution against threats, the theory relies on assumptions about the relative costs of
private and public measures against coercion. To the extent that these variables are
quantifiable, the theory can also be tested empirically.

Instead of conducting an empirical analysis, economists usually refer to relatively
simple criteria or particular narrower contexts as to when non-enforcement is
economically justified. Based on such considerations some commentators claim that the
criteria applied by courts roughly corresponds to what economic analysis would dictate
(Shavell 2007). Thus the incentive effects of the duress doctrine justify the voiding of
contracts where one party deliberately created (“engineered”) the lack of adequate
alternatives for the other. Still, “[n]ot every threat will ground a claim of economic
duress. [...] For example, a seller’s assertion of a right to withdraw from negotiations is
a threat of non-contracting. Were it to constitute economic duress, property owners
would be required to sell their goods to anyone who offered a derisory price for them.
Since this would destroy property rights, a distinction must be made between
permissible threats and economic duress.” (Buckley 1991: 38). In still other scenarios
the net incentive effect is less straightforward: here both legal rules and commentators’
views diverge. Besides “economic duress” this applies to necessity and contract
modification. In these cases, discussed in more detail below, the validity of an ex ante
welfare-enhancing transaction is at stake.

To illustrate the divergent views, Anthony Kronman argued for a “modified Paretian
approach”, in terms of which a coerced contract should be enforced when the
enforcement of the same kind of contract would make most parties better off (Kronman
1980). In other words, Kronman “would ask whether the welfare of most people who
are taken advantage of in a particular way is likely, in the long-run, to be increased by
permitting the kind of advantage-taking in question in the particular case” (Trebilcock
To note, this approach reverses the conventional argument of economists for voluntary exchanges and private ordering (voluntariness implies welfare-improvement): in order to establish whether a certain transaction should be considered voluntary, one has to address its hypothetical welfare effects.

In his *The Limits of Freedom of Contract*, Michael Trebilcock argued that the enforceability of a coercive contract should be decided by asking the following question (“literal Pareto principle”): “Does this transaction render both parties to it better off, in terms of their subjective assessment of their own welfare, relative to how they would have perceived their welfare had they not encountered each other?” (Trebilcock 1993: 84) For instance, even if an offer is exploitative in the sense that the contract divides the gains from cooperation very unequally, as long as both parties gain from the contract, the literal Pareto principle would dictate to uphold it.

This approach has to face some difficulties too. As subjective assessments of welfare are difficult to prove in judicial procedures, at the end of a complex line of argument, Trebilcock tends to take the transactions on a relevant competitive market as the normative benchmark for establishing whether the coerced transaction should be enforced. “The competitive market price functions in Trebilcock’s account not just as a measure of the substantive fairness of the bargain, but as a determinant of the voluntariness of the transaction.” (Stewart 1997: 230). For this comparison to work in practice, one needs data about the prices on an actual, relevant and essentially competitive market. Another concern is that while the competitive market price is arguably the closest practical measure of objective value in such situations, it is at best correlated to the subjective welfare of individuals.

More generally, when economic theories focus on the effects of duress rules on future contracts they completely disregard the particulars of the interaction between coercer and victim (Stewart 1997: 224–225). The reason for this neglect of the particulars lies in the *ex ante* perspective of economic analysis and its focus on questions of institutional design. Economic theories provide a prima facie case for freedom of contract: unless transaction costs and information imperfections are prohibitive, individual exchanges and the private ordering are preferable to the forced transfer of wealth, be it through individual coercion or government redistribution. In the exceptional cases when voluntary transactions are impossible or impracticable, tort law or the rules of restitution and unjust enrichment should provide proper incentives (Calabresi – Melamed 1972, Bouckaert – De Geest 1995, Wonnell 2000).

### 3.3 Shavell: duress as holdup

In recent years, at least three further particular ways have been suggested to reconceptualise duress in economic terms. In the following sections these approaches shall be briefly discussed in turn.

In economics, the term most commonly used to capture duress and necessity is *holdup*. Besides information asymmetry and externalities, holdup situations provide a third general economic justification for limiting freedom of contract (Shavell 2007). From an economic perspective, duress, necessity, usury, contract modification and several other problems all represent cases of holdup and raise some common concerns.

Shavell’s analysis of contractual holdup (2004: 235–237; 2007) focuses on the incentive effects of holdup and of legal intervention, as well as the welfare effects of risk bearing. His model has a clear normative starting point: law should minimize social costs: “the costs of any efforts made prior to the occurrence of situations of need, the costs of
furnishing aid in situations of need, losses sustained in situations of need, and risk-bearing costs where parties are risk averse” (2007: 330). He then identifies five incentives generated by the possibility of holdup (2007: 330–331). First, holdup can lead contractors to invest in wasteful efforts to engage in holdup. Second, it can lead victims to invest in inefficient precautions to avoid holdup or mitigate its consequences. Third, it can dilute motives of potential victims to invest in their socially beneficial activities. Forth, holdup may also represent significant risk to risk-averse potential victims. On the other hand, enforcing contracts with high holdup prices has a possible socially beneficial effect by giving incentives to search for victims in situations of need, and to make related investments. He then distinguishes two scenarios: engineered and non-engineered holdup (2007: 325–326). The first refers to the case when A creates an opportunity for himself to exploit B. The second refers to the case when A does not create but only exploits the necessity of B.

Based on a formal model, Shavell then argues that in these terms, contractual holdup may justify legal intervention. In case of engineered holdup the contract should be voided since that will remove the prospect of profit from it. In case of non-engineered hold-up, for example a rescue situation on high sea, cost-minimization would dictate price control. “In these circumstances, the policy of controlling the contract price is preferable, as that policy can reduce the problems of holdup but still allow contracts to be made.” (2007: 326). Although such administrative or judicial price control faces severe practical difficulties, Shavell claims that courts in practice solve holdup problems both in fresh contracts and in contract modifications roughly along the lines his economic model suggests. We will come back to non-engineered holdup in section 3.6.

3.4 Bar-Gill and Ben-Shahar: credibility theory
Recently, in a series of papers Oren Bar-Gill and Omri Ben-Shahar (2004a, 2004b, 2005) argued that the credibility of threats should be the key variable in determining whether contracts concluded under threat should be enforced. At the same time they criticized both philosophical commentators and legal doctrine for neglecting the issue of credibility and in this way worsening the conditions of victims.

According to their approach, the “enforcement of an agreement, reached under a threat to refrain from dealing, should be conditioned solely on the credibility of the threat. When a credible threat exists, enforcement of the agreement promotes both social welfare and the interests of the threatened party. If agreements backed by credible threats were not enforceable, the threatening party would not bother to demand a concession, and would simply refrain from dealing – to the detriment of the threatened party. The doctrine of duress, which predominantly controls such agreements, only hurts the “coerced” party. By denying enforcement in cases where a credible threat exists, duress doctrine precludes the threatened party from making the commitment that is necessary to reach agreement. Paradoxically, it is in those circumstances where a threatened party has no alternative options or adequate remedies that, under duress doctrine, she cannot secure an agreement.” (Bar-Gill – Ben-Shahar 2004a: 391)

Their model leads to the radical conclusion that “ex-post anti-duress measures, rather than helping the coerced party, might in fact hurt her[...] Anti-duress relief can be helpful to the coerced party only when the threat that led to her surrender was not credible, or when the making of threats can be deterred in the first place.” (Bar-Gill – Ben-Shahar 2005: 717) This result is driven by how the authors model legal intervention. They compare “the policy of enforcing [the contract or the] modification at
the agreed price with the policy of flat voiding of [the contract or] the modification, not with the policy of price-conditioned voiding. [...] But if courts can pursue a policy of price-conditioned voiding, courts can lower the price without [removing the incentive to contract when it is socially beneficial or, in the context of modification, without] causing breach.” (Shavell 2007: 340 n. 23)

3.5 Craswell: institutional competence

Richard Craswell, one of the few law and economics scholars open to a serious dialogue with philosophers, has taken an ambitious approach to contract rules with interesting implications for duress. Craswell does not attack philosophical theories of contractual duress frontally; rather, he questions their relevance for legal policy. He argues that while these theories are concerned with the philosophical justification of the morally binding force of contracts (promises), they have “little or no relevance to those parts of contract law that govern the proper remedies for breach, the conditions under which the promisor is excused from her duty to perform, or the additional obligations (such as implied warranties) imputed to the promisor as an implicit part of her promise.” (Craswell 1989: 489) In short, he claims that pure autonomy-based theories are indeterminate on important aspects of contract law.

A related idea by Craswell concerns what he calls relative institutional competence. He suggests that instead of waiting for philosophers to determine by abstract reasoning whether a contract was concluded (in)voluntarily, policy-makers should look at the capacities and competences of the legislator, the judiciary and the contracting parties on the one hand, and the available and desirable remedies on the other. Competence and remedies, in turn, should determine the enforceability or otherwise of the contract (term) in question. In this setting, the relevant question is whether legislators or judges and juries have the resources, especially the expertise to assess (establish, measure, qualify) the variables that a theory of legislation or adjudication would require them to do.

Methodologically, Craswell’s analysis builds on the property rule–liability rule framework suggested by Calabresi and Melamed (1972). Adapting this framework to contract formation problems he suggests that in order to determine whether a contract was concluded voluntarily or not, one first has to look at the remedies available, and then infer back to the enforceability of the problematic term. A property rule protection of contractual consent would mean that the contract is either enforced or voided (unenforceable). A liability rule protection would mean that the judge replaces the unreasonable terms with reasonable ones.

Craswell’s radicalism comes from the idea that it is not substantive issues but the available remedies that determine whether a contract term should be declared unconscionable or a contract should be voided for economic duress. As far as law is concerned, a contract should be deemed lacking voluntary consent when and only when the choice among the available remedies, based on their respective costs and benefits, dictate that. In determining which way to choose, Craswell explicitly speaks of two factors: (1) the relative institutional competence of the judge and the legislator to determine what is efficient; and (2) the position of the party offering the contract (term) to modify his behaviour.

For duress this theory implies the following. When A makes B sign a contract at gunpoint, the availability of remedies dictate that the contract should be voided and B should be protected by property rule. The reason is that in such situations A can easily change his behaviour and the circumstances surrounding the contract formation can be
proven with relative ease in front of a judge. On the other hand, when this is not the case, the contract should be considered voluntary in the eyes of the law. In some formulations Craswell goes as far as to suggest that the definition of duress should be directly based on this comparative analysis of competence. This would put the discussion in a clearly instrumental framework, by detaching it from the moral discourse.

### 3.6 Duress and necessity

Law usually imposes sanctions both when a coercer creates a desperate situation for the other by a threat of harm and when he merely exploits the pre-existing necessity of the other party. From an economic perspective, it is important to distinguish these two kinds of threats: duress in the narrow sense and the exploitation of necessity (Cooter–Ulen 2008: 285–287). This duality of duress and necessity roughly corresponds to Steven Shavell’s distinction between *engineered* and *non-engineered* duress (2004: 235–237) viz. holdup (2007: 225–226), as discussed above or, in still other terms, to the distinction between the *endogenous* (duress) and the *exogenous* (necessity) origin of the dire situation (Bar-Gill–Ben-Shahar 2005).

Cooter and Ulen contrast duress and necessity in the following way (Cooter–Ulen 2008: 285). Both constitute dire (as opposed to moderate) constraints (they inhibit rational behaviour) and can therefore justify promise breaking (non-enforcement). While duress is attributable to the promisee who is acting in a threatening manner, necessity is attributable to an action or omission of the promisor which made his situation desperate. This, in turn provides a situational monopoly for the promisee who can threaten the promisor with destructive inaction. An example of contracting under necessity is the transaction between a driver running out of gas on a remote road and a passerby offering to sell him gas at an exorbitant price. They summarise the economic analysis of the two doctrines in the following table (Cooter–Ulen 2008: 308, Table 7.5):

<table>
<thead>
<tr>
<th>Legal doctrine</th>
<th>Fact triggering legal doctrine (problem)</th>
<th>Incentive (solution)</th>
<th>Legal solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duress</td>
<td>Promisee threatens to destroy</td>
<td>Deter threats</td>
<td>No enforcement of coerced promises</td>
</tr>
<tr>
<td>Necessity</td>
<td>Promisee threatens not to rescue</td>
<td>Reward rescue</td>
<td>Beneficiary pays cost of rescue plus reward</td>
</tr>
</tbody>
</table>

While the enforceability of necessity contracts is not regulated as uniformly as duress, the economic literature analyses the regulation of both types of contract in a similar manner, in light of the goal to minimise the social costs (Shavell 2007). As discussed above, the main difference is that the socially beneficial activity of potential rescuers should be incentivized by rewarding rescue. Models used are similar to standard economic models of tort law: they take into account the incentive effects of necessity contracts on the activity level and the level of care by the victim, the rescue costs and the investment made by potential rescuers, as well as the welfare effect of risk-bearing. There are two main policy recommendations derived from such models (Cserne–Szalai 2008). First, in contrast to duress, contracts concluded in necessity may increase social welfare (rescue may reduce net social losses). Such contracts should not be declared void. Rather, an *ex post* regulation of the contract price is suggested. Second, the contract price should reflect whether the other party (the rescuer) invested *ex ante* in increasing his capacity to rescue or came to help the one in necessity accidentally. Professional rescuers should be compensated with a higher reward.
By focusing on the latter issue, Cooter and Ulen distinguish three types of rescue: fortuitous, anticipated and planned (Cooter–Ulen 2008: 285–287). In order to give the right incentives for rescue in the three cases, an increasing amount of rescue reward is required. Fortuitous rescue uses resources that were available for the rescuer by chance. This is the case of a passerby who happens to have a full tank when he meets the desperate driver and then offers to sell the driver some of his gas. In this case the rescuer needs to be compensated for the resources actually used in order to ensure that he prefers to rescue the victim rather than drive off. Anticipated rescue uses resources that were deliberately set aside ex ante, in case they were needed for rescue. This is the case of a passerby who carries an extra five-gallon can of gas in case he runs into stranded drivers. Anticipated rescue requires a larger reward in order to give incentives for potential rescuers to set some resources aside ahead of time. Finally, planned rescue is provided by someone who is deliberately looking for people to rescue. This is the case of a patrol service set up to rescue travellers who had ran out of gas in remote areas. As planned rescue requires active searching for people in distress as well as large ex ante investments, this is the most costly of the three kinds of rescue. Consequently, it requires the largest reward.

Shavell’s model (2007) also implies that if the contract price is higher than the costs of the rescue plus a small reward, it should be reduced to this amount. While this policy conclusion seems to be in line with how some actual legal orders regulate the matter, it can be shown that cost-based price control is optimal only in very specific circumstances (Buckley 2005: 194 n40, Cserne–Szalai 2008). More comprehensive models suggest that the minimization of the social costs of necessity may require a significantly higher contract price. Price control in general and cost-based contract pricing in particular, only lead to efficiency under specific assumptions.

3.7 Duress in contract modification

Besides bargaining for a “fresh contract”, duress is also relevant for the modification of contracts. As discussed above, many contract modifications are entered into under circumstances where one of the parties is dissatisfied with the original contract and is threatening to breach. While the underlying economic considerations are similar, both the contractual techniques available to handle the issue and the legal doctrines regulating contract modification significantly differ in the two contexts. Economists usually focus on how the parties can design an optimal contract which is opportunism-proof, i.e. only allows for renegotiation when it is welfare-increasing. “In many circumstances, parties will be reluctant to make specific investments in settings in which contract renegotiation is possible; accordingly, it is, in principle, beneficial for them to commit not to renegotiate if they can credibly do so.” (Hermalin–Katz–Craswell 2007: 55)

From a law-and-economics perspective the main issue about contract modification is how the law should distinguish welfare-decreasing duress from welfare-increasing adaptation of the contract to changed circumstances. In general, when a contract is renegotiated under duress the modification should not be enforced; if the contract is renegotiated under changed circumstances, enforcement is efficient (Muris 1981, Al vazian–Tre bilcock– Penny 1984, Graham–Pierce 1989, Jolls 1997, Triantis 2000, Hermalin–Katz–Craswell 2007: 55).

An example of duress in contract modification is the American case Alaska Packers’ Ass’n v. Domenico [117 F. 99 (9th Cir. 1902)]. The captain of a boat hired a crew in
Seattle for a fishing expedition to Alaska. Once they reached Alaska, the crew demanded a substantial wage increase. After their return to Seattle, the captain refused to pay the higher wages, one of his claims being that he the agreement to pay them was made under duress. While the court justified its decision with the doctrine of consideration, an economic reconstruction of the case would focus on whether such contract modifications are welfare-increasing. The crew, being in a situational monopoly, was threatening to destroy the value of the captain’s investment if they do not get the wage increase. When a party makes specific investments in the initial contract she becomes vulnerable to holdup in case of contract modifications. In such cases efficiency dictates that opportunistic modifications should not be enforceable. On the other hand, to modify the example above, if circumstances had changed in such a manner that the crew had to work more (e.g. because weather conditions deteriorated), their request for higher wages should have been enforceable. In most modern legal systems, “a modification made in proportionate response to new circumstances, unanticipated at the time of contracting, would generally be enforced, while an outright attempt to rewrite the original terms of the bargain would not.” (Hermalin–Katz–Craswell 2007: 55)

3.8 Unconscionability – a proxy for duress?
A number of legal systems allow for the judicial control of contractual fairness. Although such control is performed under different doctrinal labels in different legal systems, it has two basic aspects: procedural and substantive (Hatzis – Zervogianni 2006). The former (in American terminology: procedural unconscionability) refers to the absence of a meaningful choice by one party; the latter (substantive unconscionability) refers to an allocation of the risks and burdens of the contractual bargain which is unexpected or objectively unreasonable. While many commentators have criticised the unconscionability doctrine generally and substantive judicial control particularly, a number of law and economics scholars have argued that to the extent that the doctrine is interpreted as a proxy for involuntariness in contract formation, it is justified economically (Epstein 1975, Schwartz 1977).

Richard Epstein interprets the unconscionability doctrine in this way. Inefficient (welfare-reducing) contracts should not be enforced. Duress provides a presumptive evidence of some underlying problem that justify the non-enforcement of the contract but duress is sometimes too hard to prove directly. If the consent was defective but due to the evidentiary burden, some technicalities or practical difficulties of proof the requirements of the duress defence are difficult to meet, unconscionability can provide relief. As an empirical generalization, gross disparity or value inequality between the two parties’ performances may signal involuntariness. Based on this generalization, duress might be ascertained indirectly through a combination of procedural and substantive fairness rules (Epstein 1975).

Some American courts interpret the inadequacy of consideration in a similar vein, as a proxy for some formation defect. “[I]nadequacy of consideration is always potentially relevant as circumstantial evidence of duress, mistake, fraud, or some other ground for setting aside a contract. The less adequate it is, the stronger the evidentiary effect will be.” (Posner 2007: 101)

3.9 Economic duress, exploitation, and fairness
There are many kinds of pressure on a person entering into a contract: the law has to determine which are serious enough to authorize the claimant to avoid his contract. As discussed above, when allowing formation defences such as duress the law may be more or less lenient. While the normative justification of the duress doctrine is less problematic within its traditional narrow boundaries (i.e. when threat or actual coercion was used), it is more controversial whether the doctrine should be used for paternalistic purposes or to promote substantive fairness and ambitious policy objectives, such as redistribution. In one view, the doctrine of duress should protect the conditions of an autonomous choice only. Others argue that the range and quality of opportunities available to the contracting parties should also matter. Looking at the case law, the boundaries of the duress doctrine are also relatively unclear (Giesel 2005).

In particular, it has been discussed in the legal literature since several decades under what circumstances the law should void contracts for “economic duress” (Hale 1943, Dawson 1947). This is a particular instance of the normative question indicated above as to where to draw the threshold of voluntariness. To simplify a bit, the alternative is the following. Should the “non-coercive exploitation” (Feinberg 1983) of economic necessity trigger the remedies for duress? Or should the lack of alternatives for one contracting party, rather, count among those “legitimate inequalities of fortune” (Feinberg 1986: 196–197) that contract law should be neutral towards?

Philosophically minded commentators who base their definition of duress on the moral evaluation of the coerciveness of individual actions sometimes argue for a broader construction of duress, implying a larger regulative role for courts (e.g. Zimmerman 1981, Radin 1996). To recall, the main argument in the philosophical approach to coercion is that the consequences of an exercise of autonomy depend on the opportunities available to the individual. On this basis, many autonomy theorists argue that what is called economic duress, i.e. the lack of alternative ways to procure income, should be considered a case of coercion.

While economic analysis cannot provide a definitive answer to this question, it nevertheless provides arguments for a narrow construction of the duress doctrine. These arguments refer, inter alia, to the long-term incentive effects of invalidating contracts which are both ex ante and ex post inefficient, and to the inadequacy of contract law for systematic redistribution. We will focus on the first argument.

It follows from the very binding nature of contracts that the realization of a previously known risk or in other words, ex post regret is not a sufficient reason to allow withdrawal. From an economic perspective the question is about Pareto efficiency: if the transaction rendered both parties to it better off, in terms of their subjective assessment of their own welfare, relative to their position before their encounter, then the contract should be held valid. Thus, regardless of the range of opportunities, when the contract has improved the situation of the person in economic hardship, the contract should be upheld.

Those who argue that individuals are entitled to some minimum level of economic well-being or that certain resources should not be the subject of contractual exchange and infer that the law of contracts should invalidate contracts when the scope of choice for one party is limited by economic deprivation, forget at least two things. The first is that invalidation is not the only remedy available. Price control, administrative or judicial, may be alternatives. The other point is related to what Margaret Radin (1996), in a slightly different context, called the double bind effect. This term refers to the problem that in many cases the prohibition of a transaction may actually worsen the plight of the
individual whose welfare is the central concern. For example, banning prostitution may eliminate an income-earning option of poor women. Even in a case of deprivation one can say that the opportunity of choice does not reduce, but rather increases the individual’s welfare, relative to the other options available. On the other hand, prohibition is unlikely to increase the welfare of those concerned, and it “will almost never have the effect of enlarging the available choice set” either (Trebilcock 1995: 374). If the normative criterion of contract enforcement is based on individual preferences then invalidating contracts for “economic duress” (or “exploitation” or “commodification”) is arguably an instance of unjustified paternalism or moralism.

**Bibliography on Duress**


Other references
Feinberg, Joel 1983. ‘Noncoercive Exploitation’ in Ralf Sartorius (ed.) Paternalism Minneapolis: University of Minnesota Press, 201–235


Owens, David 2007. ‘Duress, Deception, and the Validity of a Promise’ *Mind* 116 (No. 462), 293–315


