N

Naïve Consumers: Contract Economics

Elena D'Agostino
Department of Economics, University of Messina, Messina, Italy

Abstract

Consumers are viewed as a weak contract party by both lawyers and economists, although with some distinctions. The unconscionability theory addresses consumers' naïvety, to be intended as partial or total incapacity of understanding contract terms, as a reason for public intervention. Economists as well agree that law must protect consumers against sellers' abuses, especially when contracts contain add-ons or are preprinted and no bargaining is allowed.

How law should intervene is still an open question. On the one hand, lawyers point out that courts should not enforce contract clauses literally but rather should replace them with terms that consumers could have reasonably expected and approved. On the other hand, economists focus on how sellers exploit naïve consumers and warn that regulation may turn out pejorative if it is not able to educate naïve consumers or if it allows the seller to raise the price up when forced to offer good terms.

The Notion of Naïvety in Law

Consumers can be viewed as a special category of buyers that has magnetized lawyers and economists' attention in the last decades. A distinction is, however, necessary to highlight what the two disciplines have in common and nevertheless why they differ.

Starting from a legal point of view, when lawyers refer to consumers, they have in mind a weak and nonprofessional contract party with limited or no contract power, and usually uninformed of every clause or consequence of the contract signed (see Korobkin 2003). This sort of considerations has driven the legal literature to formalize a general theory of unconscionability: accordingly, it is unconscionable a contract that mentally competent people would not sign and/or that no fair and honest person would accept. Not surprisingly, consumers' naïvety, to be intended as partial or total incapacity of understanding the contract content in each of its clauses, has been viewed as one of the main factors to support the theory of unconscionability.

Accordingly, since consumers may sign the contract without reading or understanding its content, their signature on the contract may not correspond to a meaningful consent to all its clauses. It comes out that courts should not enforce contract clauses literally, especially those so one sided

A. Marciano, G. B. Ramello (eds.), *Encyclopedia of Law and Economics*,
https://doi.org/10.1007/978-1-4614-7753-2

to turn out vexatious, but rather should replace them with terms that consumers could have reasonably expected and approved.

The Notion of Naïvety in Economics

On the other hand, economists make a second-order distinction between two categories of uninformed consumers: those who are simply uninformed but conscious of the potential risks hidden into a contract that has been prewritten by the counterparty and those who are not only uninformed but also unconscious of such risks. Consumers in the former category are labeled "rational" or "sophisticated" and are usually assumed to be able to fill their knowledge gap by paying a positive cost to read contract terms. Consumers in the latter category are labeled "boundedly rational" or "naïve" and are assumed to never read the contract simply because they do not realize the risk of signing without reading.

Ellison and Ellison (2009) analyze how firms can exploit consumers' in Internet transactions where price search engines make a price search more difficult and sometimes not convenient.

As pointed out by Armstrong and Chen (2009), it does not imply that naïve consumers do not care of contract clauses at all or less than sophisticated consumers. There might be also consumers who are naïvely pessimistic, that is, they do not trust the seller and believe that contracts always include unfriendly clauses. More generally, we can say that naïve consumers simply do not realize the risk involved in signing a contract without reading or they never trust the seller and never sign.

Similarly to the legal literature, there is a general consensus in the economic literature as well that court intervention is necessary to protect naïve consumers. To be more precise, economists have focused on how unregulated seller(s) exploit such consumers. The answer clearly depends on the beliefs that naïve consumers hold about the terms in preprinted contracts. Despite in some contexts a consumer might expect complex contracts to contain default terms (e.g., he might believe that the non-price terms in a complex contract address contingencies which are irrelevant to the buyer and are therefore regulated by default rules or uses), most of the literature prefers focusing on the case in which consumers are weaker and easier to exploit for a contract-drafting seller: it happens when consumers are naïvely optimistic, believing that preprinted contracts contain favorable terms.

The Effects of Regulation

A recent behavioral literature treats the infrequency of reading as indicative of buyer naïvety and argues that sellers lack an incentive to educate such buyers: cf. Gabaix and Laibson (2006) for the special case of add-on goods and D'Agostino and Seidmann (2016) in respect of contracts of adhesion. The intuition is the following: if all buyers believe that complex contracts contain favorable terms, then an unregulated seller would include the worst possible terms in preprinted contracts charging the price that buyers would be happy to pay for friendly terms according to those constraints imposed by the market structure in which they operate. Regulations may therefore benefit such naïve buyers.

Despite the general principle that parties are free to negotiate contracts, courts, legislatures, and regulators have sometimes overruled onerous terms in consumer contracts. There are two underlying and potentially conflicting rationales: Sect. 218 of the Uniform Commercial Code states that a clause is unenforceable if a buyer would not have traded had he known its contents, which suggests that naïve buyers should be protected. By contrast some courts, notably in Henningsen v. Bloomfield Motors (NJ 1960), have cited market share as an aggravating factor, which suggests that all buyers should be protected against sellers who exploit market power to offer onerous contracts.

It is also unclear whether regulations are designed to protect buyers who have already accepted onerous terms (and must necessarily gain) or to protect buyers who have yet to enter the market.

Suppose all consumers are, economically speaking, naïve in the sense that they believe that contract terms are favorable and do not take into account the risk of accepting a preprinted contract without reading its content. Sellers in a free market would therefore include onerous terms in their contracts charging the highest price that such consumers are willing to pay for favorable terms. Consumers therefore risk to get a negative payoff, to be intended as the difference between their evaluation for an onerous contract and the price they pay believing that it is rather favorable, under the supposition that they value more a favorable contract than an onerous contract.

Trivially, regulations which do not educate buyers have no effect on play: if naïve consumers remain unaware of the effect of regulation, they will not be able to call the seller in front of a court of law if terms are different from those legally acceptable. Conversely, effective regulations which either mandate terms turning out favorable to consumers or prohibit rather onerous terms induce both a monopolist and each competitive seller to offer a favorable contract, but the effect on price crucially depends on the market structure. Precisely, a monopolist will charge the highest price that consumers are willing to pay for those terms, whereas competition will lead price down to the production cost. As an effect, consumers get zero from buying and may be better off compared to a free market in which they would have been offered onerous contracts charging a price close or equal to their reservation price for a favorable contract.

Suppose now that some consumers are naïve and some others are rational. If rational consumers must pay a cost to read and understand the contract, a commitment problem arises and an unregulated seller has again no incentive to include favorable terms. Precisely, consumers cannot commit to read the contract and seller (s) cannot commit to include favorable terms. As a consequence, an unregulated seller cannot charge a price equal to consumers' reservation price for favorable terms if he plans to trade with rationals, but under some conditions he may find it profitable to include favorable terms with positive probability if rational consumers read with positive probability. Naïve consumers are therefore protected by rational consumers if the latter category is sufficiently large in the market to force seller(s) to reduce the price and possibly to include favorable terms with positive probability. Conversely, if naïves represent a large percentage of consumers into the market, then seller(s) will ignore rationals and equilibrium conditions correspond to those found when all consumers are naïve. Focusing on the former and more interesting case, regulation will force sellers to offer favorable terms with market structure dictating the equilibrium price. The effect on naïve consumers' payoffs is therefore ambiguous depending on what price they were charged in a free market given the probability of finding favorable terms.

References

Armstrong M, Chen Y (2009) Inattentive consumers and product quality. J Eur Econ Assoc 7(2–3):411–422

D'Agostino E, Seidmann DJ (2016) Protecting buyers from fine print. Eur Econ Rev 89:42–54

Ellison G, Ellison SF (2009) Search, obfuscation and price elasticities on the internet. Econometrica 77(2):427–452

Gabaix X, Laibson G (2006) Shrouded attributes, consumer myopia, and information suppression in competitive markets. Q J Econ 121:505–540

Korobkin R (2003) Bounded rationality, standard form contracts, and unconscionability. Univ Chicago Law Rev 70:1203–1295

National Identity

▶ National Locus

National Law

▶ National Locus

National Locus

Giuseppe Eusepi
Department of Law and Economics of Productive Activities, Sapienza University of Rome, Faculty of Economics, Rome, Italy

Abstract

The idea of nation grew out of the ancient idea of law and subsequently evolved in the form of national states as territorial boundaries. The unitary national state turned what was an internal conflict between a statesman and a merchant into an external conflict among nation states. As the USA and the EU demonstrates, the concept of nation state as a public good is either ambiguous or tautological. In a more or less distant future, the national locus is apt to move from a territorial to a procedural dimension.

Synonyms

National identity; National law; Nation-state

Definition

National consciousness as opposed to nationalism.

The idea of nation has been worked out by the Romantic movement, and it may be regarded as the force that set in motion movements engaged in asserting national identities in Europe such as the Italian Risorgimento (Salomone 1970) and the German Nationsbildung (Hroch 2005), which began during the first half of the nineteenth century and extended until the second half of the century and beyond.

The progenitors of the idea of nation are to be found in the idea of state, which marks the end of the Middle Ages and the beginning of the Modern Age, and much earlier in the idea of law that had its roots in ancient Greece and, above all, in ancient Rome. In ancient Rome, law was basically a private law: the ideas of law and order guaranteed private property and the related freedom of exchange.

Nineteenth-century idea of nation capsizes the private-public law relationship, which shaped the ancient Roman law. The conflict between public law and private law is best represented by the perennial conflict between a statesman and a merchant: the former operates for the purpose of bringing about adjustments of relations among individuals, and the latter is concerned with individuals' desire to get rich. The idea of nation emerged exactly to nuance this conflict by turning conflicts within the nation into conflicts outside the nation. The concept of national locus separates the nation as a unit from the rest of the world. In symbolic terms, the national anthem and the flag symbolize the national locus by definition, but from a more rational point of view, the clearest definition of national locus is the Constitutional Charter and constitutional political economy sensu lato (Buchanan and Tullock 1962). The clue to constitutional political economy is to be found in the consensual-contractual dimension, which is the outward form of the inward notion of metarule; thus, the notion of national locus becomes constitutional locus. In this context, the centrality of the constitutional law requires that control on application of the law be operated by a judge acting objectively and impartially just as a judge of the Constitutional Court.

Standard or mainstream economic thinking has been much concerned with a substitute for the procedural component, which was viewed as irrelevant. One such substitute has been seen in the objective dimension with the consequence that the concept of national locus loses its process connotation (including the democratic one) and becomes a measure of an optimal national quantity commonly lumped in the label GDP or better public goods. In its widest abstract sense, the concept of national locus can be pictured as a defense or protection line against external invasions by other nations as well as by foreign goods. Although the concept of national locus seems to have strengthened itself in public economics, especially in the form of Samuelsonian national public goods, nevertheless it is not free from

conflicts. And, in fact, the conflict arises, not only because the distinction between national public goods and local public goods is not so sharp but because the theoretical foundations of the national and local dimensions may be ambiguous. National public goods, if we are to understand what they really are, should be viewed in their territorial dimension – and in this case the distinction is tangible but tautological. There is in fact a difficulty in drawing a line between "local locus" and "national locus" because they shade into one another and give rise to a dimensionally ambiguous concept, which is not clearly determined. The same may be said of the national locus vis-à-vis the supranational locus. The United States and the EU are a flagrant example of this indeterminacy. While in the United States the federal government is the emblem of the national community (national locus/national constitution), in the EU, member states (national loci) are a subset of a supranational body.

Despite its manifest negativity, this ambiguity is tonic to the advancement of the very idea of national locus. No doubt it is too soon to foreshadow an era in which the concept of national locus is conceived of as separated from the concept of nation as territory. But there seems to be good ground for asserting that the national locus will increasingly move over from a spatial dimension to a procedural dimension (supranational alliances). Moving from a territorial dimension to a procedural dimension involves the abandonment of the concept of nation and a stricter role for law and rules.

In a nutshell, the national locus is a notion moving a great distance away from a territorial symbolic connotation of a nation clustering around its hymn and flag. Against this background, the national locus is apt to transform itself into the notional locus of the law (Eusepi 2008) built up over the two foundations of the law: civil law and common law (Pound 2000).

Cross-References

- ▶ Becker, Gary S.
- ▶ Constitutional Political Economy
- ▶ Independent Judiciary

References

Buchanan JM, Tullock G (1962) The calculus of consent. University of Michigan Press, Ann Arbor

Eusepi G (2008) Dura lex, sed lex? Insights from the subjective theory of opportunity cost. Eur J Law Econ 26:253–265

Hroch M (2005) Das Europa der Nationen. Die moderne Nationsbildung im europäischen Vergleich. Vandenhoeck & Ruprecht, Göttingen

Pound R (2000) The ideal element in law. Liberty Fund, Indianapolis

Salomone AW (ed) (1970) Italy from the Risorgimento to fascism: an inquiry into the origins of the totalitarian state. Anchor Books, Garden City

Nation-State

▶ National Locus

Negotiated Procedures in EU Competition Law

Frédéric Marty[1] and Mehdi Mezaguer[2]
[1]CNRS - UMR 7321 GREDEG, Université Côté d'Azur, Université Nice Sophia Antipolis, Valbonne, France
[2]CNRS – GREDEG, Université Côte d'Azur, Université Nice Sophia Antipolis, Valbonne, France

Definition

The enforcement of EU competition law in the field of antitrust, e.g., the sanction of abuse of dominant position and collusive agreements, increasingly uses negotiated procedures. Negotiating remedies with incriminated undertakings is a well-known practice in the field of merger control. The practice of settlements is also significantly developed in the United States. However, it remains a relative new approach under the EU competition law enforcement. This chapter presents the three main tools at the disposal of the EU Commission: the leniency program, the direct settlement, and the commitment procedures. It

analyses their main challenges and issues in both legal and economic fields.

Negotiated Procedures Under EU Competition Law: An Introduction

Negotiated procedures under EU competition law mainly encompass three procedures: leniency, commitments, and direct settlements. We mainly consider the negotiated procedures implemented for the application of Articles 101 and 102 of the Treaty on the functioning of the European Union. We draw some parallels with remedies in merger and acquisition control.

These procedures are provided under European Union law using a specific framework. Firstly, they were organized around both soft law and hard law after a period of insecure practice on the shadow of law: Sunlight is said to be the best of disinfectants and the twilight zone in which settlements currently are negotiated, desperately needs light (Van Bael 1986). Secondly, these procedures stemmed from the practice of the European Commission. The Commission increasingly tries to organize its intervention on the market using cooperation with companies for the application of EU competition law (in our case, it is mainly antitrust law). Thirdly, these procedures are organized under the notion of mutual concessions which allows an approach based on rationality. Between main legal instruments and the peripheral ones, a particular dynamic has been installed to the advantage of the public authority. Under EU law, the European Commission is clearly at the center of the proceedings. From that perspective, negotiated procedures have to be discussed above all around the question of their essential nature: Do these procedures pertain to a logic of cooperation with the public agency or to a real negotiation between two partners?

Understanding the reality of these procedures implies to put into light the interests for companies and the public authority, appreciate both legal and economic approach, and consider the problems of the procedures under fundamental rights.

An Overview of EU Negotiated Procedures for Articles 101 and 102 TFEU

The Negotiated Procedures Before the Negotiated Procedures

Understanding the specificities of EU competition law-related negotiated procedures supposes to consider the historical dynamic that led to them, especially the experience of the ad hoc agreements that were implemented in the field anticompetitive practices before the EU 1/2003 Regulation, and to make some connections with the case of remedies in merger controls.

The informal practice of the negotiated practices is a specific situation under EU competition law. Indeed, the European Commission has started in the 1960s to relay on informal arrangements following the fundamental principle that no situation is ever lawful or unlawful forever (see, e.g., cases *Nicholas frères* (Commission Decision 64/502/CEE, 30 July 1964) and *Henkel-Colgate* (Commission Decision 72/41/CEE, 23 December 1971). The European Commission was at this time in charge of a very innovative field, and this type of *cooperation* could have been seen as a will to implement softly antitrust law. The first decisions applying such an approach were mentioned in the reports on competition law by the European Commission and have mainly concerned agreement cases (abuses of domination cases appeared in the beginning of the 1980s). For example, in its first report in 1971, the Commission explained such situation by the need of saving resources. In its fifth report, issued in 1975, the Commission mentioned the "friendly" nature of its intervention. Anyway, in that period, the informal intervention of the European Commission was not well known by public or academics, and the Commission itself was not really able to explain the nature of such procedures (when questioned, e.g., by a resolution of the European Parliament in 1987).

As a consequence, the informal development of negotiated procedures has created some uncertainties, particularly since it was not possible to isolate each procedure according to its specificities. From this perspective, the creation of a formal architecture allowed an individual and global

understanding of negotiated procedures. The formal scheme also had to respond to the need for legal certainty and clarification regarding practices involving cooperation through reciprocal concessions. The resulting procedures, without claiming similarity, converge on several points but need at the same time to be presented over their singularity.

When working on the proposal of a new framework for the application of Articles 101 and 102 (former 81 and 82), the Commission provided that: "The second way in which the proposal will increase the protection of competition is by allowing the Commission to concentrate on the detection of the most serious infringements" (EU Commission 2000). Negotiated procedures aimed indeed to increase efficiency in law application by the liberation of available resources. This has to be considered as a real philosophy at the European level which leaded to the creation of a formal framework of three procedures qualified as negotiated (Mezaguer 2015).

Leniency Procedures

The first procedure corresponds to leniency programs. These procedures were created only under soft law. Indeed, in 1996, 2002, and 2006, the European Commission has adopted three communications for the organization of the proceedings. The first communication provided that these procedures were intended to apply competition rules with efficiency. The 1996 communication on leniency introduced the first formal instrument providing a legal framework for negotiated procedures under EU competition law. Nevertheless, in the first times of the regulation 1/2003, which constitutes the main instrument of EU antitrust law, no mention of leniency procedures was made. Finally, Regulation 2015/1348 has mentioned such procedures but only related to Regulation 773/2004 which provides only procedural rules. The legal situation of leniency is rather paradoxical.

However, leniency has become a major instrument of cartel resolution even in member states of the EU. Moreover, the European model is a reference for member states even if, in a 20 January 2016, *DHL Express*, C-428/14 case, the European Court of Justice has provided that the creation of leniency program was not an obligation for member states.

Leniency gives several ways of cooperation to the competition authority and to the incriminated undertakings from different perspectives. Firstly, the company denouncing the cartel runs for immunity if it shows a "true spirit of cooperation." This situation was generalized under the 2002 communication and differs from the US situation while even the first company is recognized guilty (but without any fine). The condemnation of the first company to a zero euro fine will probably have a specific importance for follow-on civil actions (even if the present framework is questionable, see, for instance, Cauffman (2011) and Mezaguer (2015)). Nevertheless, the first company to denounce must obtain a reward far superior to those who follow it so that the leniency program can be efficient. A leniency program should initiate a denunciation race among the cartelists to be efficient.

After this first company, leniency will allow other cartelists to receive rewards (between 5 and 50% depending on the period) for not disputing the facts, for giving valuable material for the proof, and for going through useful cooperation with the Commission. From this perspective, cooperation implies express, clear, and precise recognition of the facts (*Tokai Carbon*, Court of First Instance, 29 April 2004, joint cases T-236, 239, 244, 246, 251, and 252/01). Lower-rank leniency maintains incentives for the companies to engage a race for cooperation. From this perspective, companies are no longer supposed to help in the mere initiation of the investigation, as it is the case for the first-rank leniency, but must reinforce the Commission in its work of qualifying the infringement. Indeed, it is well known that "a reduction in the fine will be granted for a contribution during the administrative procedure only if that contribution enabled the Commission to establish an infringement with less difficulty and, where appropriate, to put an end to that infringement" (Conclusions of the Advocate General Geelhoed, *Commission v SGL Carbon*, 19 January 2006, case C-301/04 P). Finally, the Commission benefits from a wide margin of interpretation to reward companies. This range of

reductions distinguishes the leniency from the direct settlement procedure.

Direct Settlement Procedures

The transaction procedure corresponds to direct settlements. This procedure has been created through both a Communication of 2 July 2008 (2008/C-167/01) and a Regulation of 30 June 2008 (n°622/2008) of the European Commission. This Regulation has modified the specific procedural regulation 773/2004, which entirely relies on the European Commission, and not the general one. Even if the use of regulations contrasts with the soft law, the European Commission is still at the center of decisional power. Moreover, the direct settlement procedure is also seen as a piece of a general transactional scheme for cartel resolution with the leniency program. Under the direct settlement procedure, companies who plead guilty before the Commission can expect a 10% fine reduction. At the creation of the procedure, former competition commissioner Neelie Kroes (2005) suggested that "we may need to look at how some form of plea bargaining procedure could bring advantages." Under the actual framework for direct settlement, companies that "plead guilty" and take a commitment not to contest the conclusions of the Commission receive a 10% reduction of the fine as a reward (the reward is combinable with leniency ones). Consequently, direct settlement is a hybrid procedure between American "plea-bargaining" and French-style "non-contestation des griefs." It is more a simplified procedure than a negotiated one.

Commitment Procedures

Commitment procedures were the first ones which were included in the general Regulation 1/2003 since their creation (Article 9). Commitment procedure has to be distinguished from leniency and direct settlement first of all because it does not apply to cartels. Indeed, Regulation 1/2003 clearly provides that such a procedure is not possible when the Commission intend to impose a fine. As for other "negotiated" procedures, commitment procedure relies on an efficiency-based objective. Concerning the question of its application under Article 101 or 102 of the TFEU, The *Alrosa/De Beers* case (Court of Justice, *Commission v Alrosa*, case C-441/07, 29 June 2010) has shown that Article 9 could be used in both situations (Mezaguer 2015).

Under the commitment procedure, things are clearer. Indeed, when the Commission intends to adopt a decision (requiring that an infringement be brought to an end), the companies can offer commitments meeting the concerns of the public authority, which are expressed in a preliminary assessment. From that perspective, the Commission adopts a decision making those commitments binding on the companies. Moreover, Article 9 of Regulation 1/2003 provides that "Such a decision may be adopted for a specified period and shall conclude that there are no longer grounds for action by the Commission." Before issuing a decision making the commitments bidding the Commission must submit them to a market test procedure to invite all the interested parties to make comments. These ones aim at helping the Commission to discuss these initial proposals and to require if necessary the proposal of modified ones.

What Are the Reasons to Commit into These Procedures?

We analyze the determinants of the choice to opt for a negotiated procedure by taking successively the point of view of the competition authority and the one of the undertakings.

The Theoretical Advantages of Negotiated Procedures for a Competition Agency

We first present the common advantages of all the negotiated procedures in competition laws for an enforcement agency before considering specifically some advantages related to leniency programs and to commitment procedures (Wils 2008).

Firstly, negotiated procedures aim at reducing administrative costs for the enforcement agency. Indeed, the negotiated option may lead to a shorter duration and to reduce the burden of proof. Secondly, a negotiated settlement allows the competition authority to obtain more rapidly the end of the prejudicial practice. Such a result is also

obtained possibly more surely, as the firm chooses and implements itself the remedy that it had proposed on a voluntary basis. Thirdly, the negotiated procedure in itself limits the information asymmetries between the undertaking and the authority. The last one has not demonstrated the existence of an abuse or of an agreement and has not assessed its effects. In addition, for the commitment procedures, the uncertainties related to the adequacy and to the proportionality of the remedies proposed by the undertaking are limited through the market test procedure, which allows the authority to benefit from the opinions of the different stakeholders (consumers, competitors, etc.). Fourthly, it limits the risk induced by an adversarial procedure. To the extent that the competition authority has neither to asses the net effect of the market practice at stake on the consumer welfare nor to consider the defense of the incriminated undertaking, it reduces at the minimum the risk of being disavowed by the appeal court. Fifthly, negotiated procedures, especially the commitment ones, lead to implementation of compliance programs and by doing so favor the dissemination of a competition culture within the organization. Sixthly, the guarantees about the effectiveness of the remedies are enhanced by the fact that a failure in their implementation would lead to an automatic fine, irrespective of any assessment of the effect of this nonfulfillment. For instance, Microsoft was fined on this basis for its negligence in its implementation of the remedies negotiated in the MS Explorer case, as we will see below.

In a more specific way, we can consider that negotiated procedures have two main types of advantages for competition authorities. It avoids the burden of the characterization of competition law provisions infringement through material or economic evidences. For instance, in the case of leniency programs, the material evidences are brought by the undertaking applying for leniency. The second main advantages consist in the effect of the decision on market structure or on the dominant undertakings' future behavior. A fine mainly produces a deterrence effect. It does not lead to correct the effects of the previous market practices. In a very opportunistic spirit, the fine can be seen by a dominant firm or by the cartelists as the *cost of doing business*. On the contrary, a negotiated decision leads to behavioral or even structural measures that may remedy to an unsatisfactory situation on markets and to some extent to reestablish a level playing field on the market. It may ensure for the future a fairer and more effective competition.

An adversarial decision, as the *Google Shopping* one (June 2017) or the *Android* one (July 2018), for instance, might lead to equivalent corrective measures through the injunctions to provide an equal treatment to rival comparison shopping services and its own services in the first case or to end its tying practices and to allow the development of *fork Android* operating systems, in the second case. However it remains that these two Commission decisions are both appealed before the General Court. Even if, the General Court and the Court of Justice will uphold these injunctions, their implementation will must be strictly monitored. Such remedies will be challenged and at best implemented in a noncooperative way. At the opposite, a commitment procedure participates to a jointly constructed competition regulatory approach and not only to an ex post legal enforcement of competition law provisions. It remains that the negotiation at stake is not really a horizontal bargaining between two equal partners but rather a transactional agreement with a public authority within a vertical relationship. The EU Commission *Google Shopping* case demonstrated that the two "partners" were not in equivalent position.

The Advantages Considered on the Incriminated Firms' Side

For the incriminated undertaking, opting for a negotiated procedure may reduce significantly the procedure duration. Indeed, the shorter the case, the less expensive its costs in terms of financial resources, management attention, and reputational effect (especially on the stock market). In addition, avoiding any recognition of guilt may induce two main benefits: Firstly, it makes successful follow-on actions for damages less probable, and, secondly, it allows to stay away from any

fine increase in a future case grounded of the aggravating circumstance of repeated offense.

In the specific case of leniency programs, the undertaking applying for the leniency may hope exiting from a cartel agreement without any legal (defense) costs or administrative sanction, as fines. In the case of a direct settlement, the undertaking benefits from a fine reduction. In other words, the fine may be reduced outside the scope of leniency since the undertaking does not challenge the theory of damage presented by the competition agency or its assessment of effects. In addition, the commitment procedure allows the undertaking to avoid to be fined and to have to recognize to have breach the law. In addition, the undertaking still benefit from informational asymmetries allowing to propose the less costly remedies possible.

Considering all these advantages on both sides, negotiated procedures seem to pertain to a *win-win strategy* (Bellis 2013).

The US Origins of Negotiated Procedures in Competition Law: What Can We Learn for EU Ones?

The Antitrust Division head, Thurman Arnold, has played an essential role in this resolved use of these settlements rather than opting for adversarial procedures before courts (Waller 2004). In the competition law-related field, the negotiated procedures constitute a US transplant. It is worthwhile to consider the US antitrust history to put into relief the main features and the underlying logic of these procedures. Although the real start of the large implementation of settlement procedures by the Antitrust Division of the Department of Justice (DoJ) can be dated in the late 1930s and in the late 1940s, the fisrt Antitrust Division's consent decrees was the US v Otis Elevator Company in 1906.

How to explain this specific choice at this moment of American antitrust history? The F.D. Roosevelt administration was disappointed about the results of the cooperation with dominant firms during the *National Industrial Recovery Act* (NIRA) implementation before its invalidation by the Supreme Court. Big firms were suspected of having not fairly play the game of an economic coordination based on government support to promote investment and employment. The coordination among firms effectively led to stop price drops, but it mainly allowed firms to increase their markups without effective counterparts. This disappointment has pleaded for a public antitrust enforcement renewal, especially because, the government benefited from a large portfolio of easy-to-win cases. The NIRA agreements supervision had led government agencies to accumulate data and evidence about collusive practices. The government position was also comforted by the TNEC report conclusions. The Temporary National Economic Committee established in 1938 aimed at analyzing the monopoly powers in the US economy. While government accumulated evidence about collusions and monopolization practices, why did Th. Arnold opted for settlements and not for conventional antitrust procedures? The reasons were twofold. The first one was to obtain quickly effective results by avoiding endless legal procedures. The second one was also related to a procedural concern: averting the legal risk induced by a still conservative Supreme Court case law.

This historical experience is all the more relevant for us that the EU Commission itself has used these kinds of procedure after sector-specific enquiries in order to avoid General Court and Court of Justice legal control and to circumvent member states' reluctances to accept legal proposal aiming at liberalizing some economic sectors as utilities. We may provide the example of the energy sector enquiry in 2005–2007. A large number of formal procedures were opened after this enquiry against gas and electricity dominant operators across the EU (Hancher and de Hauteclocque 2011). For all except one, the final decision was not an infringement one but a negotiated one. Even nowadays, far-reaching and broad scope competition law remedies are obtained through commitment procedures in the EU energy sector as testified in the May 2018 *Gazprom* decision (case 39.816 Upstream gas supplies in Central and Eastern Europe, 24 May 2018).

However, US and EU procedures and practices remain specific. In the US case, for instance, for

the Antitrust Division of the DoJ, an antitrust settlement is a horizontal contract between the firm and the agency that requires judicial supervision. Indeed, the 1974 Tunney Act requires an ex ante judicial review of the DoJ consent decrees. Whatever the supervision of these agreements, it remains that the US antitrust enforcement realized a shift from a litigation-oriented model toward a more regulatory-based regime. The balance between courts and agencies has shifted dramatically toward the last ones. Ginsburg and Wright (2012) showed that by the 1980s, already 97% of the DoJ antitrust cases were settled. From 2004 to nowadays, almost all its cases were resolved through these negotiated procedures. The tendency is the same considered on the FTC side.

In cartel-related cases, the rise of negotiated procedures was later but equally important. The US corporate leniency program was first introduced in 1978. It was revised in 1993 and completed by an individual leniency program 1 year later. Its scope of application covers agreements aiming at setting prices, market or consumer sharing devices, and bid-rigging practices. From the initial 1978 procedure to the 1993, three major revisions were implemented: (1) leniency is now automatic for qualifying companies if there is no preexisting investigation; (2) leniency is still available even if cooperation begins after the investigation is underway; and (3) all employees who come forward with their company and cooperate are individually protected from criminal prosecution. The real start of the US leniency program can be dated 1993: from 1993 to 2010, US enforcement data reveal a 20-fold increase. Nowadays, in the USA 90% of the penalties imposed by the DoJ were linked to leniency-related cases (OECD 2018).

The institutional specificities of the US enforcement regime may explain this precocious and fast development. Opting for a negotiated procedure might be a rational choice for an enforcement authority while considering the judicial and political risks associated to unfruitful lawsuits. In addition, the burden of the rule of reason had grown earlier in the USA as the shift toward an effects-based approach in antitrust laws enforcement started at the late 1970s. This tendency also raises a theoretical question: are the settlements more efficient in terms of administrative performance (deciding cases quicker and easier) or in terms of obtaining concessions from the firms for *not-so-easy* cases for which the enforcement agency has no certainty about its own chances to be successful before a court? Such a game relies on the fact that the competition authority does not disclose its evidence and remains vague on the theory of damage or on the assessment of the damage to competition. As we have already mentioned for the EU competition law case, the enforcement agency does not issue a statement of objections but only a communication of competition concerns. The second player – the incriminated firm – does not know what the competition agency exactly knows and what its real chance of being convicted are. Because of the asymmetry of information that benefits to the authority, a risk-adverse undertaking may prefer enter in a commitment procedure. Indeed, the trade-off between prohibition and commitment decisions cannot be analyzed without taking into account uncertainties related dimensions (Gautier and Petit 2018).

Is the tendency observed for the EU competition law enforcement as significant as it is for the US Case? From 2007 to 2017, we count 19 Article 7-based decisions (antitrust prohibitions) and 32 Article 9-based ones (antitrust commitments). If we consider the case of cartels, we may count 40 cartel prohibition decisions (based on an adversarial procedure) and 25 based on settlements. However considering on the period from 2013 to 2017, the balance is rather different with 7 adversarial procedures and 18 negotiated ones (EU Commission 2018). The same constraints tend to produce the same results. In addition, we have to keep in mind that competition law enforcers are engaged in convergence process.

Economic and Legal Concerns Related to the Recourse to These Procedures

This conclusion illustrates some of the economic and legal issues raised by the recourse to negotiated procedures. We first consider the case of

leniency programs before considering the one of commitments. We finally consider the nature of these procedures in order to draw a dividing line between negotiated ones and transactional ones.

A Law– and Economic–Based Discussion of the Possible issues Raised by Leniency Programs and Commitment Procedures

The case of leniency programs raises several issues within the economic field. For instance, what about the incentives to collude while the expected cost of the sanction is reduced while a firm anticipates that she will be the first to betray its partners in crime? Do these procedures only lead to dismantling only poorly efficient cartels? Should it be necessary to grant a monetary reward to the cartel member who reports a cartel agreement before any investigation (Brisset and Thomas 2004)? Considered at the legal point of view, how to consider the capacity of a cartelist, possibly the initiator of the cartel, to escape at any sanction after a breaching of competition laws? How to articulate public and private enforcement, especially if we consider the follow-on actions aiming at obtaining damages? The near predominant part of leniency procedures in the cartel-related competition law enforcement undoubtedly raises issues in terms of restorative justice. At the same time, controversies over differences in the treatment of unilateral practices in the USA and the European Union can be read through the prism of relative weight differences between public and private enforcement (Cosnita-Langlais and Tropeano 2018). The rise of leniency programs in the US case and its consequences in terms of private enforcement do not raise the same issues as in the EU where follow-on actions have to be encouraged (see, for instance, Bueren and Smuda 2018).

The development of commitment procedures also leads to raise several issues in the economic field. For instance, how taking into account information asymmetries to assess the adequate, effective, and proportionate character of remedies? It raises significant adverse selection-related concerns. Do the behavioral or structural remedies proposed sufficient to address competitive concerns? For instance, in case of asset divestitures, are the values of these last ones properly selected? Will the future buyers able to exert an effective competitive pressure? The EU procedure of the *up-front buyer* aims at limiting the risk to transfer the asset to a market player who has not the financial or technically capacities to compete or has excessive incentives to collude or to avoid a too fierce competition with to dominant firm. For instance, a divestiture benefiting to a major competitor may create a symmetry among the main competitors within the relevant market and by doing so enhance the risk of collective dominance.

These concerns are not the only one induced by the information asymmetries at stake in commitment procedures. The supervision of the proper implementation of remedies also raises moral hazard-related issues. Will the undertaking implement properly and efficiently the remedies it has proposed? According to EU regulations, a failure to comply with commitments that a Commission's decision made binding leads to an automatic sanction. Microsoft experienced the situation with its €561 million fine imposed to its failure to provide European users a screen choice of web browsers from May 2011 to July 2012 despite its 2009 commitments (see the EU Commission decision, case AT.39530, *Microsoft*, 06/03/2013).

Symmetrically, information imperfections may lead the antitrust enforcement authority to require excessive remedies. Farrell (2003) illustrated this case for mergers remedies with the scalp, overfixing, and broad scope phenomena. The first one corresponds to an application of disproportionate remedy from a dominant firm in order to sanction its past behaviors or to address its structural dominance. The second one consists in requiring remedies going beyond what is necessary by taking into account the information asymmetries that the undertaking does benefit. It plays the role of a security margin. The third phenomenon corresponds to remedies unfitted to the damage theory presented in the communication of the competition concerns. It might correspond to a situation in which the competition authority takes advantage of the negotiated procedure to address several issues even if all of these ones are not closely related to the case.

Such phenomena may be difficult to observe in adversarial procedures mainly because of the judicial control exerted on them. However, this control is by far relaxed under the EU competition law by the Court of Justice judgment in *Alrosa* (EU Court of Justice 2010). To the extent that the remedies are voluntarily proposed by the undertaking, the EU Commission has not to check if less demanding or less intrusive remedies might be proposed. If the EU Commission has also to perform a proportionality test, this last one is limited to the remedies effectively proposed by the undertaking. In other words, if the dominant undertaking only proposes structural remedies, the authority has not to verify if behavioral ones may allow to obtain similar effects. These specificities may lead to concerns about competition authorities' capacity to obtain "excessive" remedies in these negotiated procedures. It may be illustrated, for instance, by the case of divestitures. These last ones are seldom used in adversarial decisions (Article 7) but can be observed in Article 9 ones. The case of the European energy sector is emblematic of such difference. The EU case may be all the more specific that the incumbent has special duties regarding the effectiveness of competition. This duty may increase the probability of being fined under a conventional procedure. It enhances the incentives to opt for negotiated procedures, and it simultaneously increases the potential cost of a negotiation failure if the Commission will decide to go back to a conventional procedure. The Google case is striking example of such a risk. Taking into account this one may induce some bias in the "negotiations" by leading firms to propose "disproportionate" remedies in order to have a greater probability to see these ones accepted.

Indeed, the use of negotiated procedures raises issues in terms of remedies proportionality. If we insist on the importance of information asymmetries, we may fear that a dominant and better-informed dominant undertaking may propose insufficient remedies or may behave strategically during their implementation in order to reduce their effect. In the same way, we might take into consideration the risk that the competition agency tends to opt for settlements while its expectations to win before courts are unfavorable. The risk in such a case is to negotiate unsatisfactory remedies to close the procedure. However, as we have underlined, the symmetrical risk cannot be minimized, especially if the dominant undertaking does prefer avoiding a prohibition decision, taking into account its reputational impacts, its induced risks in terms of follow-on actions, and possibly the possible fine increases in future decisions on the basis of the aggravating circumstances pronounced in case of recidivism.

This competition authority's strong position in the bargaining may lead to costly remedies for the incriminated dominant undertaking. We can provide an example in the domain of mergers with the *Bayer/Monsanto* case (decision of the European Commission; 11 April 2018, case M8084). Bayer took the commitment to sell to BASF a part of its activities in order to prevent a possibly non contestable position in the market of seed and pesticides. The Commission has required an up-front buyer condition. It led to limit the number of potential acquiring firm and possibly to reduce the transfer price. The up-front buyer requirement aims preventing to transfer the assets to a non-efficient market operator who cannot exert a long-term credible competitive threat on the dominant firm. Such a requirement makes sense in order to protect the competitive process, while it has an adverse effect on the dominant firm's interests.

At the legal point of view, commitments procedures may also raise several concerns. Firstly, what is the proper scope of competition law remedies? Should these ones concern prices? On the principle, it has not to. The competition authority has not the play the role of a price regulator. Nevertheless, as soon as an essential facility is at stake, commitments deal with access price-related issues. Compulsory licensing-based remedies also raise the same concern. An even more far-reaching question could be put into relief for remedies aiming at reinforcing the competitive situation of a given market.

Again, the *Gazprom* case can be relevant to illustrate this situation. In 2015, the Commission sent to Gazprom a statement of objections. According to the Commission's preliminary view, this company breached EU competition

rules by pursuing an overall strategy to partition gas markets along national borders in several member states in Central and Eastern Europe. This strategy may also have enabled Gazprom to charge higher gas prices. The remedies proposed by Gazprom and negotiated with the EU Commission both address these competitive concerns and deepen the gas internal market. They do not sanction an anticompetitive behavior by pronouncing a fine in a deterrence purpose as a prohibition decision would do. These remedies pertain to the building of competitive markets. The competition field will be not the same before and after the remedies. It is not an issue to reestablish the condition of a free and undistorted competition but an issue to create a competitive market.

What are the remedies in this specific case? A first one undoubtedly pertains to a logic of guaranteeing the end of the alleged anticompetitive practices. Gazprom commits to remove all the contractual provisions that imposed geographical restrictions or that limit the customer's capacity to resell Russian gas. A second remedy concerns gas prices. A structured process to ensure competitive gas price will be put in place in order to guarantee that Russian gas price will be aligned with the prices observed on Western European market places. The third remedy may contribute to change the competitive structure of the gas market itself: "Gazprom will enable gas flows to and from parts of Central and Eastern Europe that are still isolated from other Member States due to the lack of interconnectors, namely the Baltic States and Bulgaria." In other words, this commitment will oblige Gazprom to open its network for intra-EU gas exchanges. This remedy is a quasi-structural one. It will play as an alternative to new investment in new gas infrastructures within the EU.

A Legal Perspective on the Commitment Procedure

The rise of commitment procedures may raise other concerns in the legal field. Firstly, the negotiated procedures under the EU competition regulations lead to limit the scope of judicial control. It may question the effectiveness of the guarantees that protect the fundamental rights of the undertakings. Remedies affect their freedom of contract and their property rights. In the same vein, their entitlement to a fair trial may be altered. Secondly, we may wonder what could be the undesirable collective consequences of a near from generalized recourse to these procedures. It may reduce the quality of the jurisprudence itself. Case law is a public good at the economic sense of the term. A commitment decision reveals a poor information compared to a prohibition decision. The disappearance of the adversarial stage of the procedure hinders discussions about the theory of damage, about the balance of the effects, and about the adequacy and the proportionate character of the remedies. Altering *the struggle for law* deteriorates the quality of the signal produced by the case law. It may have several consequences. A first one is to reduce the capacity of the different stakeholders to anticipate the decisions. It impairs the legal certainty attached to the legal rule definition and its implementation. By doing so it increases the dominant undertaking's propensity to opt for such procedures. Eventually, the higher the number of cases settled, the higher the possibility to observe the development of parallel case law, specific to negotiated procedures and all the more robust that it is not balanced by any judicial control exerted by the General Court and by the Court of Justice.

We might finally wonder if negotiated procedures under EU competition regulation do really imply an effective negotiation. We have noted that a commitment procedure is not a private contract that must be validated by a court as it is the case in the USA. The bargaining between the incriminated undertaking and the authority cannot be only conceived in a horizontal way. Indeed, there is a significant verticality at stake. The EU Commission is not a player as another one. The Commission benefits from a large margin of discretion in favoring this kind of procedure for a given case (through communicating competition concerns and not issuing a statement of objections). The Commission also decides unilaterally to accept or not the commitments proposed by the undertaking and to come back to an infraction decision.

The *Google Shopping*, *Android*, and *AdWords* cases are particularly striking. Considering that the initial negotiated procedure was unsuccessful

(e.g., that the commitments proposed by Google were not sufficient to address its competitive concerns), the Commission unilaterally decided to go back to an Article 7 procedure and to send in 2015 a statement of objections. Three procedures were launched. A first one led to the *Google Shopping* decision of June 2018 (with a €2.42 million fine). A second one came to the *Android* decision of July 2018 (with a €4.34 million fine). A third one, corresponding to *AdWords* related practices is still expected.

On the one hand, it tends to enhance the credibility of the Commission. Commitments are not suggested to dominant undertakings for weak cases, and the Commission demonstrates that it may refuse insufficient proposals. The Commission's behavior may be analyzed as a reputational investment. On the other hand, these decisions contrast with the Recital n°13 of Regulation 1/2003 according to which: "Commitment decisions are not appropriate in cases where the Commission intends to impose a fine." Negotiated procedures in antitrust appear more as a competition policy tool (with all the characteristics associated to public policy in terms of verticality) than a "contractualization" of the competition law enforcement in a pure horizontal logic.

References

Bellis J-F (2013) Article 102 TFEU: the case for a remedial enforcement model along the lines of section 5 of the FTC Act. Concurrences 1:54–61

Brisset K, Thomas L (2004) Leniency program: a new tool in competition policy to deter cartel activity in procurement auctions. Eur J Law Econ 17(1):5–19

Bueren E, Smuda F (2018) Suppliers to a sellers' cartel and the boundaries of the right to damages in U.S. versus EU competition law. Eur J Law Econ 45(3):397–437

Cauffman C (2011) The interaction of leniency programmes and actions for damages. Comp Law Rev 7:181–220

Cosnita-Langlais A, Tropeano J-P (2018) How procedures shape substance: institutional design and antitrust evidentiary standards. Eur J Law Econ 46(1):143–164

E.U. Commission (2000) Proposal for a council regulation 2000/0243, COM (2000) 582 final, 27 Sep 2000, p 6

E.U. Commission (2018) Report on competition policy, COM(2018) 482 final, 18 June

EU Court of Justice (2010) case C-441/07P, Commission v Alrosa, judgment of 29 June

Farrell J (2003) Negotiation and merger remedies: some problems. In: Lévêque F, Shelanski H (eds) Merger remedies in American and European Union competition law. Edward Elgar, Cheltenham, pp 95–105

Gautier A, Petit N (2018) Optimal enforcement of competition policy: the commitments procedure under uncertainty. Eur J Law Econ 45(2):195–224

Ginsburg DH, Wright JD (2012) Antitrust settlements: the culture of consent. In: William E. Kovacic: an antitrust tribute, liber amicorum. Concurrences, Paris, ed. 407p

Hancher L, de Hauteclocque A (2011) Manufacturing the EU energy markets: the current dynamics of regulatory practice. Comp Regulat Netw Ind 11(3):307–334

Kroes N (2005) 'The first hundred days', Speech at the International forum on EC competition law, Brussels, 7 April 2005

Mezaguer M (2015) Les procédures transactionnelles en droit antitrust de l'Union européenne – Un exercice transactionnel de l'autorité publique, Bruylant, Bruxelles, 582p

OECD (2018) Challenges and co-ordination of leniency programmes, DAF/COMP/WP3 (2018)1

Van Bael I (1986) The antitrust settlement practice of the EC commission. Common Market Law Rev 23:61–90

Waller SW (2004) The antitrust legacy of Thurman Arnold. Saint John's Law Rev 78:569–614

Wils W (2008) The use of settlements in public antitrust enforcement: objectives and principles. World Comp 31(3):335–352

Neuro Law and Economics

Emanuele Lo Gerfo[1] and Ferruccio Ponzano[2]

[1]Economics, Management and Statistics Department, University of Milano Bicocca, Milan, Italy

[2]Department of Law and Political, Economic and Social Sciences, University of Piemonte Orientale, Alessandria, Italy

Abstract

Neuro law and economics is a very young discipline that aims to study law and economics phenomena towards the help of neuroscientific techniques. These studies are the result of bringing together two different approaches that start from different disciplines in order to analyze, in particular, some aspects of the punishment. These two approaches are behavioral and experimental economics and the analyses deriving from the so-called neuro-law. These

two disciplines have been developed independently in the last decades and only in the last years they are jointly used to implement scientific analyses of punishment.

Introduction

Neuro law and economics can be considered as a very young discipline that studies law and economics phenomena towards the help of neurosciences. These studies can be seen as the result of bringing together two different approaches that start from different disciplines in order to analyze, in particular, some aspects of punishment. They implement some experimental and neuroscientific techniques. These two approaches are the studies of behavioral and experimental economics and the analyses deriving from the so-called neuro-law. These two disciplines – that incorporate interdisciplinary approaches – have been developed independently in the last decades and only in the last years they are jointly used to study some scientific phenomena. Then, we have to present these different approaches to understand what neuro (law) and economics is.

Behavioral and Experimental Economics

First of all, we present how the development of Behavioural and Experimental Economics allowed to shed light, among other topics, on the study of punishment in a very new perspective for economists. Let us consider a simple environment as the one represented in the ultimatum game, where we have the presence of two subjects that face a simple economic dilemma. The first player has to choose how to divide a given amount of money between himself and the second player. Then, the second one may or may not accept the choice made by the first one. If the second one accepts the monetary endowment is divided as the first player decided. If the second player does not accept the choice of the first one both the players earn an amount of money that is equal to zero. The classical economic theory predicts that the second player will accept each positive monetary offer from the first one, even if the offer is very unfair. Moreover, the second player is indifferent between the acceptance or the rejection when the first player offers zero. Güth et al. (1982) show – using an the ultimatum game in an economic experiment – that people – the second players – are ready to reject a positive monetary offer if they consider it unfair, even if this choice reduces to zero their monetary endowment. Then, people are ready to spend money to sanction other people if they behave in an unfair way. In particular, Güth et al. (1982) show that people are ready to spend money to sanction an unfair action that damages them. This pioneering result has been confirmed by many other studies in the following decades, also with the implementation of different experimental protocol as the public good game with punishment (Fehr and Gacther 2000, 2002) and the trust – or the investment – game (Berg et al. 1995).

An important ancillary result has been reached after that a new experimental protocol using the third party punishment game has been developed. In his simplest version, in this game we have the presence of three players. A first player can decide how to divide a given amount of money between himself and a second player. After this choice, a third player must decide whether or not to sanction the first one if she/he thinks that the division between the first two players has been unfair. Obviously, the sanction has a positive monetary cost for the third player. Also in this case, the classical economic theory never predicts that a third-party will be ready to spend money to sanction an unfair behavior. As Fehr and Fischbacher (2004) show for the first time, people are ready to sanction an unfair behavior in this case also, even if it damages a person that is not the punisher. Then, they show that people are ready to spend money to sanction an unfair action even if they are not directly damaged by that. Also in this case, other studies confirm this result (see, for instance Henrich et al. (2006) and Marlowe and Berbesque (2007)) also in environments with the presence of a jury that acts as a third party (Ottone et al. 2015).

Obviously, the study of behavior related to sanction is really wide. For a survey on this topic and some suggestions for future research, the readers can refer the paper by Mulder (2016).

Neurolaw

Secondly, we have to present the development of neurolaw. In fact, law and science over time had to meet and to confront each other. Scientific discoveries have always raised a strong international juridical debate. From this meeting and clash between law and science, more or less suddenly, innovations and changes arise into legal system. Suffice it to think to legal issues about the declaration of human death or when we have before us a brain death. Other examples could be permission to use new informatics technologies inside the courtrooms or to use biological evidences such as blood or DNA trails.

Neurolaw stems from this continuous debate between law and science, particularly from Neuroscience.

Neuroscience is a discipline that studies the nervous system. It arises from the relationship of biology with medicine, psychology, mathematics, engineering, chemistry, thus evolving as an interdisciplinary science. Aim of neuroscience is to understand how the interconnections among single nervous cells (neurons) produce different perceptive and motors acts and different cognitive processes such as decision making or problem solving.

In the last years of 1900 different disciplines apparently unrelated to neuroscience begin to incorporate into their researches neuroscientific data and methodologies giving rise to new disciplines such as Neuroeconomic, Neuromarketing, or Neuroesthetics.

Neurolaw explores the effects of neuroscientific discoveries on legal rules (Petoft 2015) investigating law-relevant mental states and decision-making processes in defendants, witnesses, jurors, and judges.

Aim of law is to regulate individual's conducts, to respect human dignity, and to create a just and fair legal system. Moreover, law judges evidences about the causes of human behaviors, and neuroscientific study of human brain functions can make a very important contribution, so that the interaction/integration of neuroscientific discoveries in Law is inevitable.

The term "neurolaw" first appeared in 1991 in a scientific paper by Taylor et al. entitled "Neuropsychologists and Neurolawyers." In this paper, Taylor and his co-authors analyzed how neuroscience, and in particular Neuropsychology, could provide probative data during a trial with defendants with cerebral lesions.

In subsequent years, thanks to the advent of new technique of neuroimaging such as functional magnetic resonance imaging (fMRI) or the not invasive technique of brain stimulation such as transcranial magnetic stimulation (TMS) and Transcranial Direct Current Stimulation (TDCS), the attention to neurolaw increased. In 2007, Law and Neuroscience Project was born from The John D. and Catherine T. MacArthur Foundation. The interdisciplinary project has the goals to help the legal system in order to avoid an improper use of neuroscientific evidence in some law contexts. Furthermore, it tries to deploy neuroscientific insights to improve the fairness and effectiveness of the criminal justice system.

The neuroscientific techniques usually adopted in neurolaw are electroencephalography (EEG), fMRI and TMS and TDCS. EEG is a not invasive technique which measures brain activity during the rest or during some cognitive tasks towards the information provided by superficial electrodes placed on various regions of the scalp.

fMRI is a correlational methodology which enables to detect the so called BOLD (blood oxygenation level dependent) signal. It is an indirect measure of neural activity and represents a small change in the levels of oxygenated red blood cell when a muscular or mental activity is performed. This technique is very useful because it is characterized by a high spatial resolution (the identification of volumetric area of 3 mm).

TMS and TDCS are causal methodologies, which enable to interfere with the cerebral activity and to observe functional changes. The first uses the electric field elicited by a magnetic field, and the second uses low intensity of electric current. Both have a moderate spatial resolution and TMS has an optimal temporal resolution.

It is possible to identify several topical areas on which the research of neurolaw scientists is focused.

Mind Reading or Lie Detection

To detect deception and truth in an individual is the dream of many judges. The use of lie detector tests has become a popular and legendary symbol from crime dramas to comedies to advertisements. The lie detector test provides for the use of three physiological signals recorded with a Polygraph: heart rate/blood pressure, respiration, and skin conductivity. Courts, including the United States Supreme Court, had repeatedly rejected the use of Polygraph because of its inherent unreliability.

Recently EEG method was employed in two forensic techniques: Brain Fingerprinting (BF) and Brain Electrical Oscillations Signature (BESOS) (for a description of these technique see Chauhan (2016) and Puranik et al. (2009)). The first was admitted in 2003 in a legal proceeding of Terry Harrington in Iowa. BF detects particular brain waves called p300. According to its inventor Lawrence Farwall (neuroscientist at Harward university), these waves are correlated to memories of the past. Terry Harrington was sentenced to life for a murder, but defense subjected him to BF which resulted in a negative outcome, as if memories of the murder were not in his memory. Iowa District Court admitted BF test as scientific evidence and the murder case was re-opened. Later Terry Harrington was acquitted after the confession of a key witness. The case of Terry Harrington is the first and only one where a Court admitted the BF test. This is relevant because, although the BF has not been exposed to a peer review, its admission challenges previous judgment. This case is the proof that a neuroscientific evidence can influence the judge's decision making.

Different neuroscientific researches investigated the brain networks that underlie lie production (Meijer et al. 2016 for review). fMRI is the most used method and it is employed by several companies such as No Lie MRI. Some studies conducted by Daniel Langleben showed how lies are distinguished from truth by increased prefrontal and parietal activity (Langleben et al. 2005). Although numerous fMRI neurolaw researches on lie detection are continuing, the Courts still exclude fMRI lie detection evidence from trials. As an example, Judge Eric M. Johnson of the Maryland Sixth Judicial Circuit had refused to admit potential exculpatory fMRI evidence during the murder trial of State v Gary Smith, claiming that "the use of fMRI to detect deception has not achieved general acceptance in the scientific community"(MacArthur Foundation 2016).

In fact, fMRI studies have different general limitations: because of its correlational nature, this methodology cannot be used to prove causation. When an fMRI study shows that a brain region is active when a person is trying to lie, one cannot exclude that the same brain region might be active because it is involved and activated during a state of anxiety or in any other cognitive processes. One meta-analysis study showed that the regions usually related to deception were the ventrolateral and dorsolateral prefrontal cortex, inferior parietal lobe, anterior insula, and medial superior frontal cortex (Christ et al. 2009). These brain regions are also activated during general executive processes as planning, working memory, and emotional process. Therefore, a limitation is to discern whether the neural activity measured by fMRI is associated with lying or with other neural processes. Moreover, many scientists argue that lie detection data recorded with fMRI are valid only within the context of controlled laboratory experiments, which often poorly approximate the real word.

Neuroscientists agree that more studies are needed before starting to use fMRI lie detection methodology.

Criminal Responsibility

Differently from fMRI, MRI was frequently admitted in courtrooms when it is relevant to obtain behavioral information correlated to specific brain pathologies. In 2007 a court of New York have authorized the scan of the defendant's brain, the popular journalist Peter Braunstein indicted of kidnapping and sexual assault. The defense intended to show that the brain illness "Schizophrenia" might have made it difficult for its client to control violent impulses. At the end of the trial, Braustein was utterly convicted: this case demonstrates how a neuroscientific evidence can be useful but not necessarily

lead to acquittal. Neuroscientific evidence could be able to indirectly address questions of intent in defendants with neurological or psychological illness that might reduce the ability of judgment during the crime. Another important contribution of neurolaw regards the brain development in adolescence. In fact, neuroscientific evidence has demonstrated that the adolescent brain has a slow development of the cerebral regions necessary for cognitive control as Prefrontal cortex (Blakemore and Robbins 2012). For this reason, adolescence is characterized by low risk aversion and impulsiveness. The contribution of neuroscience to law about the behavior of adolescents brought in 2005 the US supreme court to the abolition of the death penalty.

Another important contribution of neuroscience to law is represented by studies on memory focused on memory limitations. These limitations may affect not only witnesses but also juries.

Bringing Together Behavioral and Experimental Economics and Neurolaw

Now neurolaw researches also deal with neuroscience evidence to understand and to improve decision making of the judge. Human societies universally expect that criminals will be punished, usually by impartial third-party decision makers. Then, the integration of Neurolaw together with behavioral economics can use various experimental protocols to investigate the punishment. Obviously, paradigms commonly used by neuroscientist are those used by experimental economists such as the third-party punishment or the hypothetical crime scenarios. The first is the economic game presented before where a potential punisher, the third-party, observes how a dictator share an amount of money with a receiver. In this case he may choose to punish the dictator spending his money or he can simply observe the scene. In this case, neuroscientific tools can be used to observe what happens in the brain of the three players. In the second paradigm, the subjects make various decisions about some different scenarios where a crime can be committed. These paradigms are used together with neuroscientific methodologies as fMRI or not invasive brain stimulation (TMS, TDCS) in order to investigate the neuronal underpinning of punishment. The data showed different brain networks, with an involvement of anterior Insula and dorsal and anterior cingolate cortex and Amigdala. This first network seems to be involved in the detection of norm violation (Sanfey et al. 2003; Krueger and Hoffman 2016). A second network is formed by medial prefrontal cortex, ventromedial and dorsomedial prefrontal cortex, posterior cingulate cortex, and tempo parietal junction. The second brain network connected to the first one is generally associated to self-monitoring, mentalizing (Bressler and Menon 2010), emotional processes related to the harm to the victim. Both networks are connected to a central executive network engaging posterior parietal cortex and dorsolateral prefrontal cortex. The central executive network underlies the final decision making and then selects a specific punishment, integrating the information processed in previous networks with the contextual facts as circumstances surrounding the crime. The neuroscientific evidence about punishment could elucidate the neuronal and contextual variables that can influence the decision making.

Moreover, Neuroeconomics, another neuroscientific discipline, helps the juridical system and several times neuroeconomic studies are borderline with neurolaw studies. Pisoni et al. (2014) in a TMS study assessed that fair or unfair economic decisions, and context, in terms of different interaction styles, may modulate the corticospinal excitability. Authors show that economic exchange, indeed, especially when imposed by only one part (proposer), can elicit positive or negative emotions according to the context in which it occurs, implying different moral evaluation and reactions. These results are very important for a legislator and for policy, because explore the people's reactions to injustices.

Future Developments

Although there are the a number of studies about legal decision making, lie detection, mind reading, and criminal responsibility, legal experts and

lawyers still try to understand the best way to integrate neuroscience, behavioral economics, and law. Today, the best neuroscientific data that neurolaw can deliver to law are the evidence about criminal responsibility of individuals with neurological diseases. Some neurological pathologies indeed can elicit different criminal behaviors, especially diseases caused by damage to frontal brain areas involved in control of instinctive behaviors and decision making. The other topics of neurolaw so far have not gained scientific validations external to laboratories. The law is good to be prudent. Anyway interdisciplinarity will lead us to the law of future.

Cross-References

- ▶ Behavioral Law and Economics
- ▶ Cognitive Law and Economics
- ▶ Crime: Economics of, the Standard Approach
- ▶ Crime and Punishment (Becker 1968)
- ▶ Experimental Law and Economics
- ▶ Prosocial Behaviors

References

Berg J, Dickhaut J, McCabe K (1995) Trust, reciprocity and social history. Games Econ Behav 10:122–142

Blakemore SJ, Robbins TW (2012) Decision-making in the adolescent brain. Nat Neurosci 15(9):1184–1191

Bressler SL, Menon V (2010) Large-scale brain networks in cognition: emerging methods and principles. Trends Cogn Sci 14(6):277–290

Chauhan JSD (2016) Brain fingerprinting. Int Res J Eng Technol 3(07):494–497

Christ SE, Van Essen DC, Watson JM, Brubaker LE, McDermott KB (2009) The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analyses. Cereb Cortex 19(7):1557–1566

Fehr E, Fischbacher U (2004) Third-party punishment and social norms. Evol Hum Behav 25:63–87

Fehr E, Gacther S (2000) Cooperation and punishment in public goods experiments. Am Econ Rev 90(4): 980–994

Fehr E, Gacther S (2002) Altruistic punishment in humans. Nature 415:137–140

Güth W, Schmittberger R, Schwarze B (1982) An experimental analysis of ultimatum bargaining. J Econ Behav Organ 3(4):367–388

Henrich J, McElreath R, Barr A, Ensminger J, Barrett C, Bolyanatz A, Cardenas JC, Gurven M, Gwako E, Henrich N, Lesorogol C, Marlowe F, Tracer D, Ziker J (2006) Costly punishment across human societies. Science. New Series 312(5781):1767–1770

Krueger F, Hoffman M (2016) The emerging neuroscience of third-party punishment. Trends Neurosci 39(8): 499–501

Langleben DD, Loughead JW, Bilker WB, Ruparel K, Childress AR, Busch SI, Gur RC (2005) Telling truth from lie in individual subjects with fast event-related fMRI. Hum Brain Mapp 26(4):262–272

MacArthur Foundation research network on Law and Neuroscience (2016) www.lawneuro.org/LieDetect.pdf

Marlowe FW, Berbesque JC (2007) More 'altruistic' punishment in larger societies. Proc R Soc B Biol Sci 275(1634):587–590

Meijer EH, Verschuere B, Gamer M, Merckelbach H, Ben-Shakhar G (2016) Deception detection with behavioral, autonomic, and neural measures: conceptual and methodological considerations that warrant modesty. Psychophysiology 53:593

Mulder LB (2016) When sanction convey moral norms. Eur J Law Econ. https://doi.org/10.1007/s10657-016-9532-5

Ottone S, Ponzano F, Zarri L (2015) Power to the people? An experimental analysis of bottom-up accountability of third-party institutions. J Law Econ Org 31(2): 347–382

Petoft A (2015) Neurolaw: a brief introduction. Iran J Neurol 14(1):53

Pisoni A, Lo Gerfo E, Ottone S, Ponzano F, Zarri L, Vergallito A, Lauro LJR (2014) Fair play doesn't matter: MEP modulation as a neurophysiological signature of status quo bias in economic interactions. Neuroimage 101:150–158

Puranik DA, Jospeh SK, Daundkar BB, Garad MV (2009, November) Brain signature profiling in India: it's status as an aid in investigation and as corroborative evidence-as seen from judgements. In: Proceedings of XX all India forensic science conference, pp 815–822

Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD (2003) The neural basis of economic decision-making in the ultimatum game. Science 300(5626):1755–1758

Next-Generation Access Networks

Claudio Feijóo

Sino-Spanish Campus, Tongji University - Technical University of Madrid, Shanghai, China

Abstract

Next-generation access networks refer to telecommunications infrastructures – and the

technologies that support them – able to provide final customers with data rates above 30–50 Mbps.

Synonyms

Ultrafast broadband networks; Ultra-broadband networks

Sometimes the "access" is missing and the term is used simply as "next-generation networks (NGN)," although strictly speaking NGN would refer to the whole infrastructure and not only to the part closer – the access – to the final user.

Acronym

NGAN

Definition

Next-generation access networks refer to telecommunications infrastructures – and the technologies that support them – able to provide final customers with data rates above 30–50 Mbps.

Reference Framework: Broadband and Economic Development

The rise of the knowledge economy has reinforced the role of telecommunications as a strategic investment:

> . . . the ability to communicate information at high speeds and through various platforms is key to the development of new goods and services. Broadband enables new applications and enhances the capacity of existing ones. It stimulates economic growth through the creation of new services and the opening up of new investment and jobs opportunities. But broadband also enhances the productivity of many existing processes, leading to better wages and better returns on investment. (EC 2006)

Next-Generation Networks and Next-Generation Access Networks

Next-generation networks (NGN) are the supporting infrastructure of ubiquitous broadband. They are defined as networks based on the Internet Protocol able to deliver multiple data applications – whether originally based on voice, data, and video – to multiple devices, whether fixed or mobile. In addition, the provision of applications is decoupled from networks facilitating the introduction of innovations (De-Antonio et al. 2006).

An NGN can be divided into two main parts (Knightson et al. 2005):

- A backbone transport network that interconnects the local nodes where data traffic from the final users is gathered to be switched and further transported. The backhaul from distant nodes to the core network is typically included as part of the backbone, although it is convenient at times to consider it separately (the so-called middle-mile).
- An access network that links final users with local nodes, the next-generation access network (NGAN), colloquially called the last-mile.

NGAN Requirements

There is nothing like a strict definition of the minimum access speeds provided by an NGAN or any other defining parameter. A tacit agreement at the industry level seems to put this figure at 50 Mbps or beyond, but to prove the vagueness of the case, there are no indications whether this number refers to both the upstream and downstream parts or it should be applied just to the downstream channel. Even less is mentioned about the quality of service-guaranteed data rates per customer. Also, while the above figure may represent as of May 2014 some consensus, there are a number of regulatory decisions and digital strategy plans that implicitly address figures from 30 Mbps to 100 Mbps; see the Digital Agenda Europe for 2020 as a main example. In any case, these figures are beyond conventional broadband capabilities and therefore the name of ultra-broadband networks.

Some sources provide a narrower definition in relation to NGN to include exclusively wired access networks which consist wholly or partly in optical elements and are capable of providing enhanced broadband access compared to services provided over existing copper networks.

According to this type of definitions, wireless networks are not part of NGAN; see below.

NGAN Technologies

In general, broadband access technologies can be classified by the physical medium into two major groups: wired (or fixed line) technologies and wireless technologies. The main wired technologies are based on fiber, coaxial, copper wire (or any combination of them), and power line. Wireless technologies can be either satellite based or terrestrial. Terrestrial wireless solutions can be either fixed or mobile. Due to their niche market prospects and limitations, power line communications, satellite solutions – or other recently proposed airborne solutions such as balloons – and fixed wireless are not usually accounted among the NGAN technologies.

Therefore, the list of NGAN technologies as of 2014 reduces to:

- *Fiber to the home* (*FTTH*). In this technology fiber runs all the way to the customer premises from the local node. FTTH technology is the ultimate fixed solution, supplying the highest data rates possible per household (Kramer et al. 2012).
- *Fiber to the basement/building/cabinet/curb/node/premises* (*FTTx*). FTTx is a generic term for those technologies which bring fiber from the central office closer to the subscriber. They come in many varieties depending on the termination point of the optical network. In all of these architectures, the fiber from the central office is brought down to a node where equipment is housed in a cabinet to convert signals from the optical network (fiber) into electronic (copper wire, wireless connection, or even coaxial cable). The main advantage of these solutions is reusing part of the existing legacy network.
- *Digital Solutions on subscriber loop* (*xDSL*). From the point of view of the architecture of the technology, xDSL solutions are equivalent to FTTx solutions mentioned above, their only difference being the perspective adopted: from the copper wire or from the fiber optics side. Among them VDSL is the most widely used technology over copper wire. With new technological developments able to increase data rates, copper lines will continue to be a strategic asset well into the midterm. Not only are they able to provide data rates that would fall into the NGAN category, but, in addition, they also allow for a smoother and more scalable path in the transition from existing broadband to FTTH.
- *Upgrade of cable television networks* (*DOCSIS*). Typically cable networks use HFC ("hybrid fiber-coaxial") technology. DOCSIS ("Data Over Cable Service Interface Specification") is the name of the series of standards developed for high-speed data transmission over cable television networks with release 3.0, the most widespread as of 2014. In general it could be said that the role of cable networks is particularly relevant in the NGAN competition scenario as the only different infrastructure from those of historical telecoms operators.
- *4th generation mobile communications* (*4G*). The case of mobile wireless is controversial regarding its inclusion into NGAN. Mobile technologies are approximately 3–5 years behind fixed technologies in terms of sustained data rates per user. However, they already deliver peak data rates well above 100 Mbps, and they are not far from reaching the 10 Mbps level per user with some consistency. Therefore, mobile broadband connections are considered here as a suitable (stand-alone or complementary) technological alternative to fixed access technologies. The most relevant technology is LTE, labeled as 4G by the International Telecommunications Union in 2010 (Ghosh et al. 2010). 4G plays a fundamental role: not only is it the cheapest solution for rural areas, but it can also complement or even replace fixed broadband in urban and suburban areas, especially as wireless technologies fit mobile lifestyles better.

NGAN Deployment

The choice of access technology is simply a matter of deployment costs (which in turn depend

basically on socio-demographics and geographic variables and possible reuse of existing infrastructures) and the user's requirements (and expectations).

Regarding the status of deployment, in general terms it can be said that as of 2014 NGAN is still in a relatively early stage of deployment – particularly out of main urban areas. In any case, the transition from copper to fiber access networks is underway, and it is expected to result in the replacement of most copper access networks over the next two decades.

Future Directions

From a technical perspective, future prospects for NGAN include some form of fixed-mobile convergence, where fiber networks will be complemented by heterogeneous wireless networks (Raychaudhuri and Mandayam 2012). The rationale is that no access technology enjoys the optimal characteristics for satisfying all the requirements demanded by users in every circumstance. Therefore, this case is leading operators to create platforms capable of integrating different access technologies over the same backbone network. The future market of the ICT sector, characterized by "comprehensive" operators, would be quite different from the current one, where there is clear separation between technologies.

From an economic perspective, the conditions for the deployment of NGAN are currently on the forefront of the debate about the role of telecommunication markets, the best regulation for them, the conditions for the return on investments, the type and level of competition, the requirements for sustained innovation, and the level and modes of potential public involvement (Bauer 2010).

Cross-References

- ▶ Public Investments: Broadband
- ▶ Telecommunications

References

Bauer JM (2010) Regulation, public policy, and investment in communications infrastructure. Telecommun Policy 34(1):65–79

De-Antonio J, Feijóo C, Gómez-Barroso JL, Rojo D, Marín A (2006) A European perspective on the deployment of next generation networks. J Commun Netw 5:47–55

EC (2006) Bridging the broadband gap. European Commission, Brussels. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2006:0129:FIN:EN:PDF

Ghosh A, Ratasuk R, Mondal B, Mangalvedhe N, Thomas T (2010) LTE-advanced: next-generation wireless broadband technology [Invited paper]. Wirel Commun IEEE 17(3):10–22

Knightson K, Morita N, Towle T (2005) NGN architecture: generic principles, functional architecture, and implementation. Commun Mag IEEE 43(10):49–56

Kramer G, De Andrade M, Roy R, Chowdhury P (2012) Evolution of optical access networks: architectures and capacity upgrades. Proc IEEE. https://doi.org/10.1109/JPROC.2011.2176690

Raychaudhuri D, Mandayam NB (2012) Frontiers of wireless and mobile communications. Proc IEEE. https://doi.org/10.1109/JPROC.2011.2182095

Further Reading

Astely D, Dahlman E, Furuskar A, Jading Y, Lindstrom M, Parkvall S (2009) LTE: the evolution of mobile broadband. IEEE Commun Mag 47(4):44–51

Bourreau M, Cambini C, Hoernig S (2012) Ex ante regulation and co-investment in the transition to next generation access. Telecommun Policy 36(5):399–406

Brito D, Pereira P, Vareda J (2010) Can two-part tariffs promote efficient investment on next generation networks? Int J Ind Organ 28(3):323–333

Cave M (2010) Snakes and ladders: unbundling in a next generation world. Telecommun Policy 34(1):80–85

Cave M, Martin I (2010) Motives and means for public investment in nationwide next generation networks. Telecommun Policy 34(9):505–512

Charalampopoulos G, Katsianis D, Varoutas D (2011) The option to expand to a next generation access network infrastructure and the role of regulation in a discrete time setting: a real options approach. Telecommun Policy 35(9–10):895–906

Coomonte R, Feijóo C, Ramos S, Gómez-Barroso J-L (2013) How much energy will your NGN consume? A model for energy consumption in next generation access networks: the case of Spain. Telecommun Policy 37(10):981–1003. https://doi.org/10.1016/j.telpol.2013.09.002

Elixmann D, Ilic D, Neumann KH, Plückebaum T (2008) The economics of next generation access-final report.

WIK-Consult Report for the European Competitive Telecommunication Association (ECTA)
Essiambre R, Tkach RW (2012) Capacity trends and limits of optical communication networks. Proc IEEE. https://doi.org/10.1109/JPROC.2012.2182970
Fredebeul-Krein M, Knoben W (2010) Long term risk sharing contracts as an approach to establish public-private partnerships for investment into next generation access networks. Telecommun Policy 34(9): 528–539
Ghosh A, Mangalvedhe N, Ratasuk R, Mondal B, Cudak M, Visotsky E, . . . Novlan TD (2012) Heterogeneous cellular networks: from theory to practice. Commun Mag IEEE. https://doi.org/10.1109/MCOM.2012.6211486
Given J (2010) Take your partners: public private interplay in Australian and New Zealand plans for next generation broadband. Telecommun Policy 34(9):540–549
Gomez-Barroso JL, Feijóo C (2010) A conceptual framework for public-private interplay in the telecommunications sector. Telecommun Policy 34(9):487–495. https://doi.org/10.1016/j.telpol.2010.01.001
Hoernig S, Jay S, Neumann K-H, Peitz M, Plückebaum T, Vogelsang I (2012) The impact of different fibre access network technologies on cost, competition and welfare. Telecommun Policy 36(2):96–112. https://doi.org/10.1016/j.telpol.2011.12.003
Huigen J, Cave M (2008) Regulation and the promotion of investment in next generation networks – a European dilemma. Telecommun Policy 32(11):713–721
Inderst R, Peitz M (2014) Investment under uncertainty and regulation of new access networks. Inf Econ Policy 26:28–41
Kazovsky LG, Shaw WT, Gutierrez D, Cheng N, Wong SW (2007) Next-generation optical access networks. J Lightwave Technol 25(11):3428–3442
Kazovsky L, Wong S-W, Ayhan T, Albeyoglu KM, Ribeiro MRN, Shastri A (2012) Hybrid optical-wireless access networks. Proc IEEE. https://doi.org/10.1109/JPROC.2012.2185769
Marsden C (2010) Net neutrality. Bloomsbury Academic, London. Retrieved from http://www.bloomsburyacademic.com/pdffiles/NetNeutrality.pdf
Nitsche R, Wiethaus L (2011) Access regulation and investment in next generation networks – a ranking of regulatory regimes. Int J Ind Organ 29(2): 263–272
Noam E (2010) Regulation 3.0 for telecom 3.0. Telecommun Policy 34(1–2):4–10
Ruhle E-O, Brusic I, Kittl J, Ehrler M (2011) Next Generation Access (NGA) supply side interventions – an international comparison. Telecommun Policy 35(9–10):794–803. https://doi.org/10.1016/j.telpol.2011.06.001
Wallsten S, Hausladen S (2009) Net neutrality, unbundling, and their effects on international investment in next-generation networks. Review of Network Economics 8(1):90–112
Wong E (2012) Next-generation broadband access networks and technologies. J Lightwave Technol 30(4): 597–608

No-Fault Revolution and Divorce Rate

Bruno Jeandidier
Bureau d'Economie Théorique et Appliquée (BETA), CNRS and University of Lorraine, Nancy, France

Abstract

Concomitantly with the no-fault divorce revolution in the United States, there has been an increase in the divorce rate. From this observation, the question emerged of whether divorce law had a neutral effect on divorce behavior. Economists have conceptualized the issue by using an approach based on the "Coaseian" negotiation process, comparing unilateral divorce to divorce by mutual consent. The theoretical model predicts the neutrality of the law. But in the case of divorce, these assumptions are questionable. A determination as to the neutrality or non-neutrality of divorce law can thus be made based on empirical analysis. Early empirical studies came to conclusions favoring neutrality, but little by little, as methodological controversies accumulated, those empirical conclusions became more refined. Finally, a certain consensus emerged that the revolution in divorce would seem to have had a positive short-term impact on the divorce rate on the one hand, due more to the transition to unilateralism than to the abandonment of any reference to fault, and to have had a negative effect on the longer-term divorce rate on the other, because divorce law reform would seem to have prompted better-quality marriages.

Introduction

The term "no-fault divorce revolution" comes from the United States. It originated in the debates that came about in the 1980s as one after another the various American states transitioned from at-fault divorce regimes to no-fault divorce regimes (about economic origins of no-fault

divorce revolution; see Leeson and Pierson 2017). This change in American civil law stirred fierce conflict between the advocates of traditional family law and advocates of more progressive and feminist positions. The origins of this polemic are doubtless to be found in the stir created in the media and the political sphere by the alarmist book by Weitzman (1985). Based on data drawn from California State divorce rulings, the author showed that the introduction of no-fault divorce in that state would seem to have resulted in declines in the standard of living for the average female divorcee that were far more significant than had been the case prior to divorce law reform. Subsequent work (including Peterson 1996) showed Weitzman's (1985) estimates to have been largely exaggerated; nevertheless the debate had been launched around the negative consequences of the no-fault divorce revolution. Parallel to questions about the estimation of consequences relative to the standard of living of ex-spouses following a divorce, the debate also focused heavily on whether or not this new divorce regime would in fact encourage divorce. In other words, was it the abandonment of fault as a condition of divorce that explained the increase in the divorce rate observed during that same period? Economists have seized upon this question, seeing it as an interesting application of the more general issue of the neutrality of law. Starting from the pioneering article by Peters (1986), the no-fault divorce revolution thus spread to scientific journals in economics and sociology as well, giving rise to a fairly extensive series of articles in which more recent articles challenged the preceding ones and so on.

Peters' Pioneering Article Establishes the Theoretical Framework for Economic Analyses of the No-Fault Divorce Revolution

For Peters (1986), and subsequently for other economists, the discussion does not center on the concept of fault in itself but on the decision-making process: unilateral decision versus decision by mutual consent. In a no-fault divorce regime, either spouse may apply for divorce unilaterally on the grounds that he or she no longer wishes to live as a couple, even if their partner does not agree. In a fault-based divorce regime, an injured spouse cannot apply for a divorce if the offending spouse does not agree to it; mutual consent is required unless fault is proven in court. It was in particular the observation of frequent difficulties in proving fault (with negative effects) that led American legislators to abandon the at-fault divorce regime.

Peters (1986) then uses Coase's theoretical model as an analytical framework to express the hypothesis of the neutrality of the law in regard to divorce: under certain conditions, negotiations can be arranged between the spouses, resulting only in efficient divorces, regardless of the divorce regime in force. The divorce regime in force would therefore be considered neutral.

Broadly, and making a simplified use of the notation proposed by Peters (1986), the gain derived from marriage R can be expressed as follows:

$$R = M(1-p) + E\left(A_f + A_m\right)p \qquad (1)$$

where p is the probability of divorce, M is the benefit derived from living as a couple, and A_f and A_m are the estimated postdivorce opportunities for the wife and husband, respectively, with the following assumptions:

- M is fixed, known, and perfectly divisible.
- Although A_f and A_m are not known with certainty, each spouse is fully aware of the opportunities available to their partner (informational symmetry), and these postdivorce gains are perfectly divisible.
- A division of M was agreed upon at the time of marriage, attributing X to the wife and $(M - X)$ to the husband.
- The negotiations between the couple take place without transaction costs (because people in couple relationships know one another very well).

The divorce is thus a fairly simple matter: it will be an efficient divorce if $M < (A_f + A_m)$; the

husband will want a divorce if $(M - A_m) < X$, and the wife will want a divorce if $A_f > X$.

If both spouses want a divorce and the divorce is efficient, or, conversely, if both spouses wish to remain married and the divorce is not efficient, the question of the neutrality of the law does not arise. In the other two cases, the following reasoning will apply. The situation we will study will be one in which the husband wants a divorce but not the wife; the same reasoning may however be applied symmetrically, with the wife desiring the divorce and not the husband.

We will first look at a situation where the divorce would be efficient. If the divorce law in force is unilateral, the husband will ask for a divorce without taking his partner's negative opinion into consideration; this behavior is efficient. If, on the other hand, mutual consent is required, the husband will have to negotiate his wife's acceptance of the divorce: in an "at-fault" divorce regime, the spouse seeking the divorce thus negotiates the "right to remain married" held by the other party. The object of this negotiation is to grant compensation (C_f) to the wife in order at least to cancel out the loss that the divorce represents for her ($X - A_f$):

$$C_f = \left(X - A_f\right) + \alpha\left(A_f + A_m - M\right) \qquad (2)$$

The negotiation therefore concerns α, a coefficient that represents the extent to which the husband is willing to share the gains to be obtained from the divorce with his former spouse.

Secondly, in the case of divorce by mutual consent, the husband will not be able to obtain the divorce if the divorce is not already efficient prior to negotiation, because the gains he will derive from the divorce will be insufficient to compensate for the loss the divorce will represent for his partner, and it will thus be efficient to remain married. In the case of unilateral divorce, the husband will request a divorce a priori, with no need for negotiation. However, the wife is still able to negotiate in regard to her husband's "right to divorce" and may in fact renegotiate the division of marriage gains X so that it is more in the husband's interest to remain married than to separate:

$$(M - X) = A_m + \beta\left[M - \left(A_m + A_f\right)\right] \qquad (3)$$

Ultimately, whatever the divorce regime in place (unilateral or mutual consent), if negotiations are held under the assumed conditions (symmetry of information, no transaction cost, divisibility), the Coase model shows that only efficient divorce situations ($M < (A_f + A_h)$) will arise. Thus the negotiated compensation system makes it possible, on the one hand, to avoid the inefficient divorces that might occur under a unilateral divorce regime and, on the other hand, to successfully conclude the efficient divorces that in a mutual consent divorce system could be blocked by one of the two spouses. The particular type of divorce law in force would thus be neutral with respect to divorce decisions, and so, empirically, there should be no increase in the divorce rate associated with the change in divorce regime.

Are the Hypotheses of the Coase Model Suited to an Analysis of the No-Fault Divorce Revolution?

Peters (1986) acknowledges that the hypotheses of the Coase negotiation model may be strong in the case of divorce. On the one hand, it is quite possible that symmetry of information regarding postdivorce opportunities for each of the spouses ($A_m + A_f$) may not be respected (one strategy may consist in concealing this information, and such information may be costly to obtain). In this case, negotiation in the event of a unilateral divorce may result in the non-avoidance of inefficient divorces (insufficient negotiation of the share-out of marriage gains). And, in case of divorce by mutual consent, it is possible that efficient divorces may not be implemented (insufficient negotiation of compensation). If this is the case, an imperfect symmetry of information would lead to an increase in the divorce rate.

On the other hand, it is possible that marriage gains M may not initially be well known (at the time of marriage). In this case, in a unilateral divorce regime, since the wife cannot oppose the divorce and cannot expect to obtain compensation, it is possible that she will self-insure

(by making less of an investment in the domestic sphere), which is likely to reduce marriage gains M and hence increase divorce probability p. In this scenario, the type of divorce is thus not neutral with regard to the divorce rate.

Zelder (1993a, b) analyzes two other limitations of the thesis of the neutrality of divorce law in a discussion of two other hypotheses. On the one hand, the author shows that in the presence of transaction costs, the divorce law is not neutral. The author first studies the situation of unilateral divorce regimes, where transaction costs are much lower (even zero) than those in mutual consent divorce regimes. This is the hypothesis most often retained in sociological studies based on the fact that unilateral divorce is simple and inexpensive. In the context of efficient divorce, in the first case (unilateral), the wife is not able to fully induce the husband to remain in the marriage for lack of sufficient means (which is in line with the assumption of legal neutrality). In the second case (mutual consent), the husband cannot induce the wife to divorce because the transaction costs would absorb her divorce surplus. There would thus be more divorces under unilateral divorce. The author then proceeds to study situations where transaction costs are prohibitive regardless of the divorce regime. In a unilateral divorce regime, if divorce is inefficient, the wife cannot induce the husband to remain in the marriage (whereas she could do so in the absence of transaction costs), and in a mutual consent divorce regime, the husband no longer has the means to induce his wife to divorce (in spite of his divorce surplus). There again, the divorce rate should be higher in unilateral divorce.

On the other hand, Zelder (1993a, b) also looks at the fact that marriage gains M may not be divisible and/or transferable (at least partially) between spouses, when said gains consist in a public good. Now, the primary marriage gain is often the child, who is, in fact, a public good. Under a unilateral divorce regime, when a child is present, even if the divorce is inefficient, the spouse who wishes to remain married will not be capable of inducing his or her partner to give up the divorce because their asset (the "consumption" of the child) is not transferable because it is not divisible. Again, under this assumption of indivisibility, the divorce rate should therefore be higher in unilateral divorce regimes.

These studies show that the neutrality of divorce law, as conceptualized through the Coase negotiation model, is highly dependent on the assumptions applied in the theoretical model. Thus empirical observation must ultimately be used to attempt to make a determination in regard to the neutrality of divorce law, in an investigation of whether or not the divorce rate observed does depend on the type of divorce regime in force.

Empirical Controversies Regarding the Impact of the No-Fault Divorce Revolution on the Divorce Rate

Because the various American states adopted the no-fault regime on different dates, the United States has provided an exceptional natural experimental site to study the relationship between the divorce rate and the type of divorce regime. Again in his pioneering article, Peters (1986) empirically demonstrates that divorce law is neutral, based on data from the *Current Population Survey* for 1979. His demonstration is based first of all on the observation that, all matters being otherwise equal, the fact of residence in a state that had opted for no-fault divorce was not significantly related to the probability of having had a divorce in the 4 years prior to the investigation. Secondly, his demonstration is supported by the observation that, on the other hand, residence in a State that had opted for no-fault divorce was positively and significantly correlated with the amount of compensation obtained (child support, alimony, etc.), which would thus support the negotiation hypothesis of the theoretical model.

Allen (1992), using the same data as Peters (1986), then challenged this demonstration by advancing three methodological criticisms. On the one hand, the status of states that made changes to divorce legislation during the 4-year observation period would appear to be ambiguous. On the other hand, the categorization of States according to the classification "at-fault versus no-fault" would appear highly debatable,

insofar as there is a whole continuum of situations between strict at-fault regimes and strict no-fault regimes; for example, certain states had barred the attribution of fault for the submission of a divorce application but retained the notion of fault based on negotiations regarding the distribution of assets and compensation. This critique was also made by Brinig and Buckley (1998). Lastly, it would be a mistake to introduce State-based indicators into the specifications because of the correlation with the indicator "at-fault versus no-fault." Taking these criticisms into account, Allen (1992) then shows that, contrary to the results obtained by Peters (1986), residence in a state that had opted for no-fault divorce is in fact positively and significantly related to the probability of divorce: the divorce regime would therefore not be considered neutral.

Peters (1992) responds to Allen (1992), accepting the first two criticisms and thus adopting the corrections proposed by Allen (1992) but rejecting the third criticism on the grounds that this specification allows consideration of an unobserved heterogeneity according to which it is possible that it was the States with the highest divorce rate that were the first to opt for a no-fault divorce regime (reversal of causality). In reassessing his model with two of the three corrections proposed by Allen (1992), Peters (1992) then reaches the same conclusion obtained in his original 1986 work, namely, that the type of divorce law is neutral with regard to the probability of divorce.

A year later, Zelder (1993a) published results based on different data (*Panel Study of Income Dynamics*). In the first estimate, the author confirms the conclusions reached by Peters (1992), by showing the absence of significant correlation between the probability of divorce and residence in a state that had opted for no-fault divorce. Then, in the second estimation, the author demonstrates the relevance of his public good hypothesis. To do so, the author crosses the indicator "at-fault versus no-fault" with an estimation of the public good (the share falling to the children – estimated based on expenditures for the children – in the total assets of the couple). The fact that this cross-variable has a regression coefficient that is significantly positive thus shows that residing in a State which has opted for no-fault divorce would encourage divorce when there are children present: divorce law would therefore not be neutral for married parents.

Another controversy then arose among researchers no longer using individual data but divorce rate time series by state. Nakonezny et al. (1995), using average 3-year divorce rates, show that divorce law is not neutral, insofar as these rates, observed before and after the transition to the no-fault divorce regime, are significantly different. Glenn (1997) criticizes this approach on the grounds that these estimations need to be corrected taking into account the general trend in divorce rates; the author shows that the differences highlighted by Nakonezny et al. (1995) are actually four times lower. Rodgers et al. (1997) then respond to Glenn (1997), accepting his criticism and introducing a corrective to reflect the trend but asserting that the new version is not adequate to reverse their initial conclusion of the non-neutrality of divorce law. Brinig and Buckley (1998) arrive at an identical conclusion, with different specifications. Without going directly into the controversy opposing Glenn with Rodgers et al., Ellman and Lohr (1998) propose a more elaborate methodology for correcting the trend, by taking into account peaks occurring before and after the legislative reform (the wait-and-see effect before, the surge effect after); yet their conclusions are mixed, suggesting that each State has a certain specificity. Glenn (1999) then once again revived the controversy with Rodgers et al. by questioning the method they used to calculate the trend (linear, over 10 years), insofar as the general progression of the divorce rate was not linear; it in fact accelerated at the end of the period, which with a linear specification would result in the attribution of part of that acceleration to the divorce law reform. Finally, Rodgers et al. (1999) responded to this criticism by means of a rather lengthy graphical analysis carried out on a state-by-state basis, an analysis which does not, however, lead to a definitive conclusion.

The third wave of controversy primarily involved Friedberg and Wolfers. Friedberg (1998) gathered all prior criticisms so as to better

integrate them into his analysis of individual data: the issue of endogeneity, the issue of classifying states based on the types of divorce legislation in force, the issue of trend correction, etc. His conclusions are as follows: whereas on the one hand there would indeed seem to be an effect due to divorce law (the divorce rate was 6% lower in the States that did not make a change to their legislation), this impact would appear to have been due primarily to the transition to a completely no-fault divorce regime (when the notion of fault is partially preserved, the impact is weak), and the effect would seem to be fairly permanent (it was strong just after the reform and then strengthened over the next 2 years). Wolfers (2006) then criticizes Friedberg's (1998) specification, primarily from the perspective considering the fixed "state-year" cross-effect (which mixes the fixed per-state effect and the effect of the reform) and the observation window being too short to take the trend into account. From the perspective of the results, his primary difference with Friedberg (1998) is that while the effect of the reform would indeed appear to be significant immediately after its entry into force, this effect nevertheless stabilizes after 8 years and then declines. This conclusion is consistent with that of Gruber (2004), using aggregate data drawn from several successive population censuses. Divorce law would thus appear to be non-neutral in the short term but neutral in the longer term. Drewianka (2008) then adds nuance to this conclusion. By separating no-fault divorce reforms from unilateral divorce reforms (since the two concepts are not strictly identical), the author shows that in fact the latter type of reform has an impact on the probability of divorce, and not the former.

The studies conducted by Wolfers (2006) are significant, because they open up a new research question: why does the impact of the no-fault divorce revolution diminish over time? The studies conducted by Rasul (2005) and Mechoulan (2006) then demonstrate the double impact of divorce law: in the short term, a transition to no-fault (easier) divorce leads to an increase in the divorce rate, and in the longer term this effect is offset by a negative impact on the divorce rate. Indeed, divorce becomes less probable because new marriages are of better quality (better, but later matches) in anticipation of the possibility of unilateral divorce easier to obtain and thus reducing marriage gains. In sum, the impact of the no-fault divorce revolution would appear fairly limited. In a certain way, 20 years later we came "full circle," since Peters (1986), in his pioneering article, had suggested that the no-fault divorce revolution could ultimately lead to a drop in the divorce rate if, anticipating an easier divorce and thus a conduct of self-insurance on the part of wives (synonymous with decreased marriage gains), behavior on the marriage market (marriage and remarriage) tended toward greater selectivity in matching.

The debate over the no-fault divorce revolution has not really crossed the Atlantic. Few studies have examined the neutrality of divorce law in Europe. The few European works were late in coming and therefore have benefited from the methodological advances made in North American studies. Taking up the methodologies used by Friedberg (1998) and Wolfers (2006) and applying them to 14 European countries experiencing divorce reforms between 1950 and 2003 (excluding countries that legalized divorce during this period), Gonzales and Viitanen (2009) conclude that in Europe, a transition to no-fault divorce would seem to have had a positive and permanent effect on the divorce rate, whereas a transition to unilateral divorce would seem to have had a short-term positive effect, fading after 5 years (results quite similar to those obtained by Wolfers (2006) for the United States). Kneip and Bauer (2009) confirm the latter results by conducting a similar analysis of 18 European countries for the period 1960–2003. These authors, however, contribute an original clarification, by distinguishing de jure unilateral divorce from de facto unilateral divorce. Divorce is de facto unilateral when mutual consent is no longer required at the end of a (usually short) separation period undertaken by the couple. They then proceed to show that, without calling into question the very short-term positive effect (on the divorce rate) of de jure unilateral divorce reform, the move to de facto unilateral divorce would also seem to have had a positive effect on divorce rates but a

permanent one. In sum, for the period 1970–1990, divorce law reforms in Europe would seem to account for one-third of the growth in the divorce rate (which grew by 2% in the same period), but would not at all explain the growth in the divorce rate before 1970 and after 1990. Bracke and Mulier (2017) discuss the classification of Gonzalez and Viitanen (2009) between no-fault divorce and unilateral divorce and propose to introduce in addition the reforms of procedure that make divorce easier. Reforms that reduce the duration of divorce proceedings (unilateral divorce being conditional on a period of de facto separation) would be also important. On the basis of data relating to Belgium and using a method of cointegration, they show that the reduction in the duration associated with the simplifications of 1994 had the same impact on the divorce trend as the introduction of unilateral divorce in 1974. They also estimate that a reduction of 1 month of legal divorce process would increase the divorce trend by 1.4%.

Conclusion

Economic analysis of the impact of the no-fault divorce revolution on divorce behavior has been a very interesting application, combining theory and empirical analysis, of the issue – very central to the economic analysis of law – of the neutrality of the law relative to individual behaviors. The refinement of econometric methodologies and the historical hindsight provided by the gradual accumulation of rich databases have led researchers to obtain nuanced conclusions, in particular by showing that the apparent neutrality of the divorce law relative to the divorce rate resulted, in fact, from the combination of a positive short-term effect and a negative longer effect (which is indirect, through marriage behavior). But as this revolution began to mature and the debate around the growth in the divorce rate was no longer really topical, other research questions emerged. How might divorce law impact negotiations with regard to specialization within couples, regarding premarital cohabitation, fertility, conjugal violence, female suicide, etc.?

Cross-References

- ▶ Alimony
- ▶ Becker, Gary S.
- ▶ Coase, Ronald

References

Allen DW (1992) Marriage and divorce: comment. Am Econ Rev 82(3):679–685

Bracke S, Mulier K (2017) Making divorce easier: the role of no-fault and unilateral revisited. Eur J Law Econ 43(2):239–254

Brinig M, Buckley FH (1998) No-fault laws and at-fault people. Int Rev Law Econ 18(3):325–340

Drewianka S (2008) Divorce law and family formation. J Popul Econ 21:485–503

Ellman IM, Lohr SL (1998) Dissolving the relationship between divorce laws and divorce rates. Int Rev Law Econ 18(3):341–359

Friedberg L (1998) Did unilateral divorce raise divorce rates? Evidence from panel data. Am Econ Rev 88(4):608–627

Glenn ND (1997) A reconsideration of the effect of no-fault divorce on divorce rates. J Marriage Fam 59(4):1023–1025

Glenn ND (1999) Further discussion of the effects of no-fault divorce on divorce rates. J Marriage Fam 61(3):800–802

Gonzalez L, Viitanen TK (2009) The effect of divorce laws on divorce rates in Europe. Eur Econ Rev 53(2):127–138

Gruber J (2004) Is making divorce easier bad for children? The long-run implications of unilateral divorce. J Labor Econ 22(4):799–833

Kneip T, Bauer G (2009) Did unilateral divorce laws raise divorce rates in Western Europe? J Marriage Fam 71:592–607

Leeson PT, Pierson J (2017) Economic origins of no-fault divorce revolution. Eur J Law Econ 43(3):419–439

Mechoulan S (2006) Divorce laws and the structure of the American family. J Leg Stud 35:143–174

Nakonezny PA, Schull RD, Rodgers JL (1995) The effect of no-fault divorce law on the divorce rate across the 50 states and its relation to income, education, and religiosity. J Marriage Fam 57(2):477–788

Peters EH (1986) Marriage and divorce: informational constraints and private contracting. Am Econ Rev 76(3):437–448

Peters EH (1992) Marriage and divorce, reply. Am Econ Rev 82(3):687–693

Peterson R (1996) A re-evaluation of the economic consequences of divorce. Am Sociol Rev 61(3):528–536

Rasul I (2005) Marriage markets and divorce laws. J Law Econ Org 22(1):30–69

Rodgers JL, Nakonezny PA, Schull RD (1997) The effect of no-fault divorce legislation: a response to a reconsideration. J Marriage Fam 59(4):1026–1030

Rodgers JL, Nakonezny PA, Schull RD (1999) Did no-fault divorce legislation matter? Definitely yes and sometimes no. J Marriage Fam 61(3):803–809
Weitzman LJ (1985) The divorce revolution: the unexpected social and economic consequences for women and children in America. The Free Press, New York
Wolfers J (2006) Did unilateral divorce raise divorce rates? A reconciliation and new results. Am Econ Rev 96(5):1802–1820
Zelder M (1993a) The economic analysis of the effect of no-fault divorce law on the divorce rate. Harv J Law Public Policy 16(1):240–268
Zelder M (1993b) Inefficient dissolution as a consequence of public goods: the case of non-fault divorce. J Leg Stud 22:503–520

Non-controlling Minority Interests

► Passive Minority Interests

Non-market Valuation

Sébastien Roussel[1] and Léa Tardieu[2]
[1]Department of Economics, CEE-M, Université Montpellier, CNRS, INRA, SupAgro, Université Paul Valéry Montpellier 3, Montpellier, France
[2]BETA, Université Lorraine, INRA, AgroParisTech, Nancy, France

Definition

Non-Market valuation is a set of techniques that aims at reflecting the economic value of changes, in the availability or quality, of goods and services that are not intended to be traded in the market (e.g., health care, education, environment). The objective is to estimate the impacts of these changes on one's utility and by extension on the social welfare, in order to manage these goods and services by considering their true value to society.

Economic Values

In a neoclassical perspective, goods and services are valued in a utilitarian framework, i.e., individuals are rational, have various categories of desires and wishes (e.g., food, housing clothes…), and are able to classify them according to their preferences. They aim at reaching a maximum level of individual welfare according to their income constraint (utility maximization under budget constraint). Moreover, individuals are able to value the impact of an additional unit of good on their own welfare, which is decreasing as the units aggregate (decreasing marginal utility law). In the end, the economic value of any good increases with its usefulness and scarcity.

Markets generate the value of all traded goods and services as relative prices. Prices are therefore very useful in comparing different effects, as they give some indications of goods and services relative scarcity. It is assumed that values may be reflected in individuals' willingness to pay (WTP) or willingness to accept (WTA) associated with a change in the availability or quality of the goods (see Table 1). The social value of goods is then the sum of the marginal utilities of each unit for all of its users.

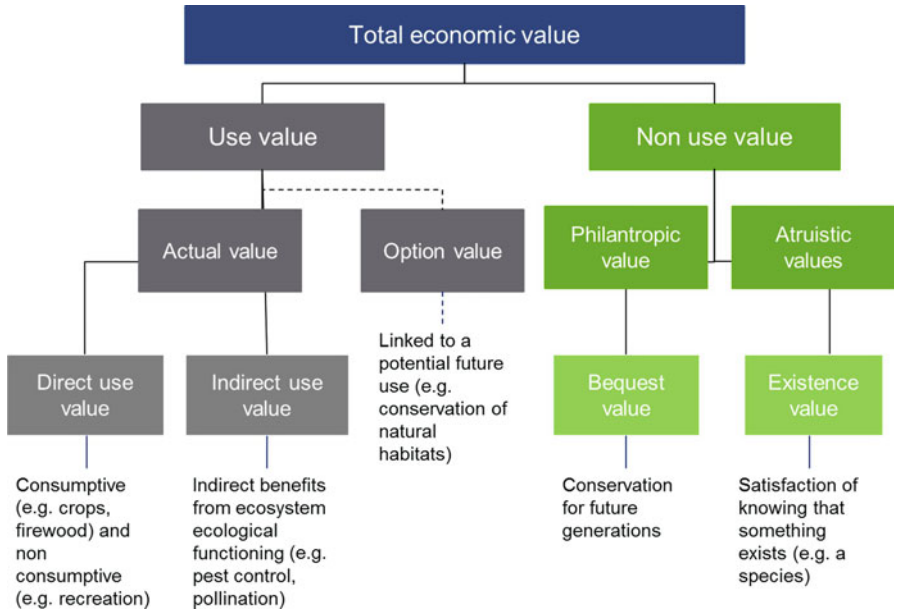
In a nutshell, economic values are mainly:

- Anthropocentric, since only the contribution to human well-being is taken into account
- Instrumental, as it is based on the utilitarian setting of the consequences of choices and actions and not on intrinsic or "deontological" values
- Subjective, because the value depends on individual's preferences

Economic values are classically divided into two types of values that make up the total economic value (TEV), in particular for environmental goods and services: use values, which include the benefits derived by an agent from a direct or

Non-market Valuation, Table 1 Willingness to pay (WTP) and willingness to accept (WTA)

	Improvement	Damage
Willingness To Pay (WTP) to…	Benefit from	Avoid a
Willingness To Accept (WTA) to…	Give up	Offset a



Non-market Valuation, Fig. 1 Use and nonuse values. (Adapted from TEEB 2010)

an indirect use of the good or service at stake, and also a potential option value; nonuse values, which reflect ethical or altruistic preferences (see Fig. 1).

Why Using Non-Market Valuation?

When the market is distorted or when goods and services have no market, market prices are poor indicators of the true economic value. As a consequence, how can one assign a monetary value to changes of goods and services that are not intended to be traded in the market (e.g., health care, education, environment)? The conventional view would be that nonmarketed goods and services are priceless, and a purely market rationale would act as if they had no value, especially while considering externality mechanisms. To tackle this issue, economists have developed a variety of methods that allow the construction of monetary indicators to value loss or gain associated with changes in the availability or the quality of these goods or services (see section "Nonmarket Valuation Techniques").

In terms of law and economics, nonmarket valuation plays a crucial role to provide information to support public and private decisions in a cost-benefit analysis (CBA) perspective (Hanley and Barbier 2009), in particular regarding high probabilities of harm and environmental justice concerns (Hsu 2005). For the effects at stake to be co-measurable, they are usually monetarized. Nevertheless, we may note that as argued by Hanley and Spash (1993), this is merely a device of convenience rather than an implicit statement that only money is meaningful.

In addition, nonmarket valuation can also help in environmental and resource economics at:

- Fixing the level of conservation effort
- Highlighting priorities in cost-effectiveness analyses by answering the question: "what is the best use of one euro invested in conservation?" in *ex-ante* analysis
- Communicating on the magnitude of global issues to promote awareness to policy makers and the general public
- Evaluating offset amounts in *ex-post* analysis

Non-Market Valuation Techniques

Nonmarket valuation techniques are used to re-build individuals WTP/WTA towards changes

Non-market Valuation, Table 2 Nonmarket valuation techniques. (Adapted from Tardieu 2017)

Economic techniques	Description
Market costs approaches	
Avoided damage costs	Uses the costs associated with the mitigation of a damage as the proxy for the value
Replacement costs	Uses costs of replacing a nonmarket service as a proxy for the value
Opportunity costs	Explicitly considers the value that is lost in order to protect, enhance or create a nonmarketed asset
Production function	Focuses on the (indirect) input costs of a particular good or service for the production of a market good corresponding to the nonmarket one
Revealed preferences	
Travel cost method	Uses data on people's actual behavior in real markets that are related to an environmental good. For example, the behavior studied is the number and distribution of trips that people make to outdoor recreation sites as a function of the cost of a trip. The travel cost is the weak complement (a complementary marketed good) of the outdoor recreation value
Hedonic pricing	Weak complementarity is assumed between the price of a property and the quality of the surrounding environment. For example, the nonmarket value is revealed through observations on the demand of residential properties
Stated preferences	
Contingent valuation	Estimates values by constructing a hypothetical market and asking survey respondents to directly report their willingness to pay to obtain a specified good or their willingness to accept giving up a good
Choice modeling	Based on a hypothetical market where respondents have a series of choice tasks in which they are asked to choose their preferred option (including *status quo*). Each option is described in terms of a set of attributes describing the good (including a price attribute) presented at various levels according to an experimental design. The analysis of the respondents' choices is based on Random Utility Maximisation (RUM) theory
Secondary valuation technique	
Benefit transfer	Uses economic information collected at a given area (study site), at a given time to make inference on nonmarket goods or services in another location (application site)

in the availability or quality of nonmarketed goods and services (see Table 2).

These techniques can be either direct or indirect. Direct methods are based on observable costs (avoided damage costs, replacement costs, opportunity costs) or allow deriving values from existing markets of close substitutes or inputs (production function). Indirect methods are based on individuals' revealed preferences through a substitute market (hedonic pricing, travel cost method), on individuals' stated preferences on an hypothetical market (contingent valuation, choice modeling), or on value transfer regarding nonmarkets goods and services from locations with similar characteristics (benefit transfer). By definition, each method has its flaws, as they either rely on individuals which may be misinformed, on proxies that may poorly represent agents' behavior, or finally on markets in which prices are poor indicators of the economic value. Moreover, all presented methods are not indicated to assess all types of values. For instance, nonuse values can be exclusively assessed by stated preferences methods. Last, revealed preferences methods are indicated for *ex-post* analysis, while stated preferences methods are suitable for *ex-ante* analysis.

Nonmarket Valuation Limits and Criticisms

Although nonmarket valuation plays a critical role to provide information to support decisions, this gives rise to conceptual and ethical issues especially with regards to environmental goods and services. These issues lead to limits and criticisms with regards to, respectively, anthropocentrism, which is at the core of value building, and how

heterogenous components of the total economic value (TEV) can be effectively aggregated; value discounting regarding future uses; and the spatial dimension of environmental goods and services valuation, including ecosystem services (Tardieu 2017; Roussel 2018). Furthermore, issues arise also in law and economics regarding how the information from nonmarket valuation can be used to value environmental damages and how suitable financial offsets can be defined.

With regards to anthropocentrism, some limits do appear linked to possible confusion between an intrinsic value of nature, independent of humans, and the one derived from marginal changes linked to individual preferences. Moreover, there is also a reluctance to add benefits and costs without reference to their social distribution. In terms of values aggregation, one of the main difficulties is related to the heterogeneity of the methods used to derive these values (see section "Nonmarket Valuation Techniques"), which can be considered as not directly comparable. For example, the aggregation of use values measured by revealed preferences and nonuse values measured by stated preferences is relatively tricky. Connected to these issues arise technical difficulties as double-counting or the simultaneous economic valuation of uses that are socially or technically mutually exclusive (e.g., wood provisioning service and regulation service regarding carbon sequestration from forests ecosystems), as well as the degree of substitutability between so-called manufactured capital and of natural capital.

If we now turn to value discounting regarding future uses, there is a choice uncertainty when taking time into account, particularly when we consider environmental issues with far-reaching consequences in the future. This has a twofold consequence on valuation (Chevassus-au-Louis et al. 2009): on the one hand, the costs and benefits linked to far-reaching consequences tend to see their weight in decisions reduced; on the other hand, there is an uncertainty on these consequences as well as the behavior and expectations of future generations. Accordingly, the selection of a discount rate can be justified following two main reasons: present preference and the decrease in marginal utility with higher incomes for future generations. Discounting can also be ethically criticized because discarding far-reaching consequences may lead to neglect critical issues with threshold (e.g., biodiversity).

In addition, the spatial dimension has often been excluded from economic valuation, leading to results far removed from reality. For example, taking into account the spatial dimension of ecosystem services in environmental economics implies measuring these services (supply and demand) according to their spatial context. The ecosystem services supply is indeed influenced by spatial variables such as climate variables, topography, soil type, hydrological conditions (Nelson et al. 2009). Simultaneously, the ecosystem services demand changes depending on the location of the ecosystem providing this service as well as the location of the potential recipients, the number of substitutes, and the accessibility (Bateman et al. 2006). In brief, the value will be influenced by three variables: supply (quantity and quality), demand (number of beneficiaries, uses, socio-economic characteristics), and the spatial context (complementary or substitutes goods).

Last, one can wonder about the reliability of the information from nonmarket valuation and how this can be used to value environmental damages and to define suitable financial offsets (Thompson 2002; Carson et al. 2003). The famous example of the Exxon Valdez oil spill in 1989 then led to nonuse value loss ranging from $2.8 billion to $7.19 billion (Carson et al. 2003), whereas use values loss were evaluated $5 million (Hausman et al. 1995). The ensuing debate with the Panel chaired by Kenneth Arrow and Robert Solow (Arrow et al. 1993) is thus an iconic illustration of controversies over the past decades and on-going debates on nonmarket valuation and Nature monetization (Gómez-Baggethun et al. 2010; Costanza et al. 2014).

Cross-References

- ▶ Cost–Benefit Analysis
- ▶ Ecosystem Services
- ▶ Externalities
- ▶ Public Goods

References

Arrow K, Solow R, Portney P, Leamer E, Radner R, Schuman H (1993) Report of the NOAA panel on contingent valuation. Fed Regist 58(10):1–67

Bateman IJ, Day BH, Georgiou S, Lake I (2006) The aggregation of environmental benefit values: welfare measures, distance decay and total WTP. Ecol Econ 60(2):450–460

Carson RT, Mitchell RC, Hanemann M, Kopp RJ, Presser S, Ruup PA (2003) Contingent valuation and lost passive use: damages from the Exxon Valdez oil spill. Environ Resour Econ 25:257–286

Chevassus-au-Louis B, Salles J-M, Pujol J-L (2009) Approche économique de la biodiversité et des services liés aux écosystèmes. Contribution à la décision publique. La Documentation Française, Paris (Rapports et documents. Centre d'analyse stratégique)

Costanza R, de Groot R, Sutton P, van der Ploeg S, Anderson SJ, Kubiszewski I, Farber S, Turner RK (2014) Change in the global value of ecosystem services. Glob Environ Chang 26:152–158

Gómez-Baggethun E, de Groot R, Lomas PL, Montes C (2010) The history of ecosystem services in economic theory and practice: from early notions to markets and payment schemes. Ecol Econ 69(6):1209–1218

Hanley N, Barbier EB (2009) Pricing nature: cost-benefit analysis and environmental policy. Edward Elgar, Cheltenham

Hanley N, Spash CL (1993) Cost benefit analysis and the environment. Edward Elgar, Cheltenham

Hausman JA, Leonard GK, McFadden D (1995) Utility-consistent, combined discrete choice and count data model assessing recreational use losses due to natural resource damage. J Public Econ 56(1):1–30

Hsu S (2005) On the role of cost-benefit analysis in environmental law. Environ Law 35(1):135–174

Nelson E, Mendoza G, Regetz J, Polasky S, Tallis H, Cameron D, Chan KMA, Daily GC, Goldstein J, Kareiva PM, Lonsdorf E, Naidoo R, Ricketts TH, Shaw M (2009) Modeling multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs at landscape scales. Front Ecol Environ 7:4–11

Roussel S (2018) Ecosystem services. In: Marciano A, Ramello GB (eds) Encyclopedia of law and economics. Springer, New York. https://doi.org/10.1007/978-1-4614-7883-6_711 (in press)

Tardieu L (2017) The need for integrated spatial assessments in ecosystem service mapping. Rev Agric Food Environ Stud 98:173–200

The Economics of Ecosystems and Biodiversity (TEEB) (2010) In: Kumar P (ed) The economics of ecosystems and biodiversity: ecological and economic foundations. Earthscan, London/Washington, DC

Thompson DB (2002) Valuing the environment: courts' struggles with natural ressource damages. Environ Law 32(1):57–89

Non-observability

▶ Information Deficiencies in Contract Enforcement

Non-price Predation

▶ Raising Rivals' Costs

Non-verifiability

▶ Information Deficiencies in Contract Enforcement

Norms and Standardization

Mitja Kovac
Department of Economic Theory and Policy, Faculty of Economics, University of Ljubljana, Ljubljana, Slovenia

Abstract

The purpose of this chapter is to survey the academic literature on the economics of norms and standardization in the domain of administrative law and to synthesize their main themes. The chapter begins by introducing basic economic framework and continues addressing the issues of norms and tax compliance. Next, the chapter surveys some of the most important areas in the administrative law and economics literature. Topics include norms and environmental compliance, standards in traditional law and economics scholarship, the issues of norms, standards and regulatory functions of the state and the occupational heatly and safety standards and norms.

Definition

Standard is a written definition, limit, or rule, approved and monitored for compliance by an authoritative agency as a minimum acceptable benchmark. Standardization is a framework of agreements to which all relevant parties in an industry or organization must adhere to ensure that all processes associated with the creation of a good or performance of a service are performed within set guidelines. Standards occupy a middle position between information measures, representing low intervention and prior approval, representing high intervention. Standard is also the legal or social criterion that adjudicators employ to judge actions under particular circumstances. Standards provide a greater degree of flexibility to judges and allow them to consider fact-specific circumstantial evidence. Norms are behavioral regularities associated with a feeling of obligation supported by normative attitudes.

Introduction

One of the pathbreaking insights offered by the law and economics scholarship is the notion that contracts, laws, and all human relationships are, due to bounded rationality, positive transaction costs, and asymmetric information problem, necessarily incomplete. However, the legislator may choose the level of identified incompleteness of the enacted laws by formulating legislation with different degree of specificity. The law and economics literature refers to this choice as a choice between standards and rules (Luppi and Parisi 2011). Kaplow (1995) argues that legislator is actually, while choosing between rules and standards, balancing between ex ante or ex post content and should calculate costs associated with each option. Whereas Schäfer (2001) suggests that the use of rules rather than standards has the advantage of reducing corruption, concentrating human capital, and cutting down court delays caused by complex decisions. Obviously, his suggestion rests on an assumption that legislators are inherently less corrupt and better informed than judiciary. Moreover, one may argue that norms and standards should be analytically seen as an economic institution developed to address the notorious positive transaction costs problem (Coase 1961; Marneffe et al. 2015) and the overwhelming market failures problem (Gómez-Barroso 2016).

According to the dominant law and economics scholarship quality standards which subject suppliers of goods and services to behavioral controls and which penalize those who fail to conform are the prevailing form of social regulation (Ogus 2004). Ogus in his encyclopedical work on economics of regulation stresses that standards have been applied, while pursuing different goals and employing different techniques, to a vast variety of commercial, industrial, and social activities (Ogus 2004). Standards can be subdivided into three categories which represent different degree of intervention: (a) target standards, (b) specification standard, and (c) performance standard. Extensive law and economics literature investigates different aspects of standards (more than 400 references) in relation to public interest justification, cost-effectiveness models for standard setting, legal formulation, and even, for example, on private interest considerations (Ogus 2004). Careful examination, however, reveals the lack of rigorous law and economics analysis of standards in narrowly defined administrative legal setting.

However, while one may argue that the economic role of standards in the domain of administrative law calls for further law and economics treatment, an overview of law and economics scholarship devoted to the study of norms reveals a vast amount of literature. McAdams and Rasmusen (2007) report that since the early 1990s, considerable scholarship in law and economics has turned its attention to norms. Numerous articles have addressed the power of social norms and their relevance for law (Ellickson 1991, 1998; Cooter 1996; Posner 2000a), have investigated the nature of norm's definition (Ellickson 1998; Picker 1997; Mahoney and Sanchirico 2003; Kaplow and Shavell 2002; Kahan 2001; Scott 2000), have shed the light on how norms work (Hirschelifer and Rasmusen 1989; Ellikson 1998, McAdams 1996; Buckley

2003), and have stressed the importance of norms to legal analysis (Posner and Rasmusen 1999; Epstein 2001). Moreover, there is an extensive scholarship on how norms affect welfare (Sunstein et al. 2000; Shavell 2002; Kaplow and Shavell 2002) and on the interaction between the law and norms (Posner 1996a, b; Shavell 2002). Furthermore, extensive law and economics scholarship addresses the specific application in the fields of tort law, contracts and commercial law, property, intellectual property, criminal law, family law, international law, and even constitutional law (McAdams and Rasmusen 2007).

In addition, norms have been fairly addressed also in the economic analysis of administrative law. For example, Hasen (1996) offers norms as a possible explanation for rational choice theories of voting behavior. Hasen argues that as with other types of socially beneficial behavior, individuals might receive small social rewards for voting or small sanctions for not voting, and hence voting rights might reflect the degree to which a community succeeds in socially beneficial behavior (Hasen 1996). Hansen suggests creation of a legal duty to vote and supplement informal, social incentives with legal sanctions (Hasen 1996).

Norms and Tax Compliance

Posner (2000b) started the discussion on whether strict enforcement of tax law is complementary with social norms and introduced a signaling as a compromise between the standard model and the approaches that try to make sense of social norms by complicating utility functions. His signaling model differs from the standard model only by introducing the plausible assumption that people have private information about their own tastes, including their discount rates (Posner 2000b). Posner's signaling model implies that if information were costless, social norms would not exist and consequently rejects the claim that social norms are internalized or that people feel guilty when they violate social norms (Posner 2000b). Lederman (2003) takes investigation further by exploring the relationship between enforcement and compliance norms. She suggests that enforcement and compliance norms are complementary and that the enforcement can buttress norms-based appeals for compliance (Lederman 2003). Her study offers empirical evidence that there is a general societal norm of tax compliance in the United States but that, among certain groups, there may be a norm of noncompliance (Lederman 2003). She also suggests that enforcement will increase the number of people who obey the laws for prudential reasons (Lederman 2003). If one would create such a critical mass of taxpayers than the violation of the law will, according to Lederman (2003), it will create a disutility on the rest of population and hence create a tax payment incentive stream.

Norms and Environmental Compliance

The impact and interplay of norms have been also widely discussed in the relation to environmental compliance as part of general administrative procedure. Vandenbergh (2003), for example, proposes a conceptual framework that accounts for the influence of norms on environmental decision-making and on corporate environmental compliance. Vandenbergh assesses substantive norms of law compliance, human health protection, environmental protection, and autonomy and suggests that administrative enforcement policies should strive to harness those norms (Vandenbergh 2003). Carlson (2001) analyzes types of the US local governments' strategies framed for inducing higher levels of individual environmental compliance. Local governments have actually, in order to achieve such targeted behavior, employed different novel forms of recycling. Carlson (2001), while assessing these new forms, argues that a lawmaker should focus on the policies that reduce the cost of recycling rather than those that try to change people's preferences. By employing such policies, lawmakers will, analytically speaking, make signaling more effective and will also introduce direct esteem toward those citizens that recycle their daily waste (Carlson 2001).

Standards in Traditional Law and Economics Scholarship

The traditional general economic rationale (outside contract law) for government involvement in standardization, as with several other government activities, has derived from the possibility of "market failure" and the public good character of standards. Left to its own devices, the market produces too little or too much standardization or standardization of the wrong sort (Swan 2000). Swan (2000) also points out that an important role of standardization is creation of a strong, open, and well-organized technological infrastructure that serves as a foundation for innovation-led growth. Standardization also increases competition and enable firms to reduce costs and increase quality. Standards might also spur the development of product and service markets that are based on the newest technologies (Swan 2000). Moreover, Swan (2000) also provides the most comprehensive source of estimates of the macroeconomic benefits of standardization and emphasizes the following ones as the most important: (a) competitiveness cannot be achieved by innovation alone, but requires efficient diffusion of innovation, and standardization plays a key role in that, (b) standards provide a positive stimulus to innovation, (c) standards contribute at least as much as patents to economic growth, (d) standards have a positive effect on trade and do not seem to act as barriers to trade, (e) international standards are more important than national standards in encouraging intra-industry trade, (f) standards enhance international competitiveness, and, finally, (h) the macroeconomic benefits of standardization exceed the benefits to companies alone.

Norms, Standards, and Regulatory Function of the State

It is well acknowledged that informal of formal norms, principles of conduct, tacit agreements, and standards play a pivotal role in the coordination and harmonization of economic activity (Weigel 2006). Sometimes, as Weigel (2006) points out, norms and standards represent a framework for a more formalistic setup of legal rules, directives, and decrees. Such informal rules frequently delineate collective action which brings them into the domain of public law. As previously discussed standards and norms are instrumental in understanding of the most puzzling economic and social issues. Widespread technical standards and certifications which represent (or which should represent) a transaction costs minimizing mechanism could be, for example, listed as such examples, as such institutions that pursue allocative efficiency.

The most widely discussed regulatory regimes imposing standards are the ones of occupational health and safety, consumer protection/consumer products, and environmental pollution. In the domain of administrative law, Choi (1996) offers a two-period model (with merely two agents) and investigates the trade-off between ex post standardization and ex ante standardization. His model implies that in ex post period each user selects a technology from a random distribution of technologies (Choi 1996). Choi's model actually identifies Nash equilibrium that indicates an excess of ex ante standardization and a deficit of ex post standardization if users choose to experiment in ex ante period.

Jeanneret and Verider (1996) in their subsequent investigation analyze conditions for compatibility in vertically differentiated demand where imported high-quality good compete with low-quality domestic good. They derive a range of tax rates which make it profitable for both suppliers to opt for compatibility. In such compatibility only and only the compatible goods will prevail. Jeanneret and Verider (1996) also note that these ranges depend on the size of the quality augmenting effect of compatibility. They also show that EU policies that subsidize switching costs are likely to fail if the subsidy scheme is not established before foreign and domestic firms decide on compatibility (Jeanneret and Verider 1996).

Moreover, such subsidy scheme has to be binding for a longer period so that policy is credible for the decisions horizon of the suppliers (Jeanneret and Verider 1996). Goerke and Holler (1998) interpret their insights and argue that EU standardization should show continuity. Goerke and Holler (1998) also argue that there might be a high potential for rent seeking in the organization of European standardization activities and standard setting. They stress the unwillingness of private agents to contribute to the financing of the standardization procedures as an obvious rent-seeking behavior's indicator. They also criticize the lack of financial arrangements on the side of EU commission and the lack of clearly assigned rights as the main sources of inefficient standard setting procedures (Goerke and Holler 1998). Hence, European Union, while providing an uncertain institutional framework (lack of exact duties, responsibilities, and influence in the standardization process) and while retaining the power to revert to detailed, interventionist standard setting, prevents companies, research institutions, chambers of commerce, consumer organizations, trade unions, and other stakeholders from making sufficient, standard-setting, commitments (Goerke and Holler 1998).

In addition, Weigel (2006) argues that the most crucial issue regarding norms and standardization in administrative law, that still awaits critical law and economics assessment, is the enforcement issue.

Occupational Health and Safety

The control of risks arising at the workplace occupies a special position among regulatory regimes and has a well-documented history (Thomas 1948). Ogus (2004) identifies the optimal loss abatement (optimal care) as the public interest economic goal of regulatory standards. He defines optimal care as the one where the marginal benefits equal marginal costs. However, as Ogus (2004) correctly observes, such an optimal care is rarely met. Employers have imperfect information of the risks involved in particular jobs and cannot be assumed to make rational decisions. Hence, an unregulated market would rarely generate optimal safety standards of occupational health and safety. In order to mitigate identified market failures, administrative agencies developed several different regulatory strategies that should produce optimal occupational health and safety standards. The predominant strategy is the agency approach where an independent public agency is entrusted with the task of determining exclusively the standards which will correspond to the optimal level of care in the particular workplace circumstances. Agencies actually mimic the idealized labor market, taking into account also distributional considerations (Ogus 2004). Under the quasi-market approach the primary responsibility for setting the standards is entrusted to the employer. In such model, administrative agency plays merely a residual role and scrutinizes the standards emerging from such practices to verify that they are consistent with the goal of optimal safety (Ogus 2004). Such type of standard-setting attempts to foster market-type solutions and external, by administrative agency-imposed standards, is only enforced in the instances of failures of such a quasi-market approach.

Concluding Remarks

The traditional general economic rationale (outside contract law) for government involvement in standardization, as with several other government activities, has derived from the possibility of "market failure" and the public good character of standards. Standards have been actually applied, while pursuing different goals and employing different techniques, to a vast variety of commercial, industrial, and social activities. However, while standards and norms have been thoroughly addressed in more than 400 references in economic literature and also produced a vast amount of law and economics literature, the assessment of public/administrative law actually reveals surprising lack of systematic assessment. Identified gap in the law and economics of

standards and norms in the domain of administrative law opens doors for an extensive law and economics treatment that would offer sets of legal and economic arguments for an improved regulatory response.

References

Buckley FH (2003) The morality of laughter. University of Michigan Press, Ann Arbor

Carlson A (2001) Recycling norms. Calif Law Rev 89:1231–1300

Choi JP (1996) Standardization and experimentation: Ex ante vs. ex post standardization. In:Holler MJ, Thisse JF (eds) The economics of standardization, Special issue of the European journal of political economy, vol 12. Elsevier, Amsterdam, pp 273–290

Coase R (1961) The problem of social cost. J Law Econ 3:1–44

Cooter RD (1996) Decentralized law for a complex economy: the structural approach to adjudicating the new law merchant. Univ Pennsylvania Law Rev 144:1643–1696

Ellikson RC (1991) Order without Law: how neighbors settle disputes. Harvard University Press, Cambridge

Ellikson RC (1998) Law and economics discovers social norms. J Leg Stud 27:537–552

Epstein JM (2001) Learning to be thoughtless: social norms and individual computation. Comput Econ 18(1):9–24

Goerke L, Holler MJ (1998) Strategic standardization in Europe: a public choice perspective. Eur J Law Econ 6:95–112

Gómez-Barroso JB (2016) Market failure (analysis). In: Marciano A, Ramello GB (eds) Encyclopedia of law and economics. Springer, New York, pp 1–5

Hasen RL (1996) Voting without law? Univ Pennsylvania Law Rev 144:2135–2179

Hirschelifer D, Rasmusen E (1989) Cooperation in a repeated prisoner's dilemma with ostracism. J Econ Behav Organ 12:87–106

Jeanneret MH, Verdier T (1996) Standardization and protection in a vertical differentiation model. In: Holler MJ, Thisse JF (eds) The economics of standardization, Special issue of the European journal of political economy, vol 12. Elsevier, Amsterdam, pp 273–290

Kahan DM (2001) The limited significance of norms for corporate governance. Univ Pennsylvania Law Rev 149:1869–1900

Kaplow L (1995) A note on subsidizing gifts. J Public Econ 58(3):469–477

Kaplow L, Shavell S (2002) Fairness versus welfare. Harvard University Press, Cambridge

Lederman L (2003) The interplay between norms and enforcement in tax compliance. Ohioa State Law J 64: 1453–1513

Luppi B, Parisi F (2011) Rules versus standards. In: Parisi F (ed) Production of legal rules. Encyclopedia of law and economics, vol 7. Edward Elgar, pp 43–53

Mahoney PG, Sanchirico C (2003) Norms, repeated games and the role of law. Calif Law Rev 91: 1281–1329

Marneffe W, Bielen S, Vereeck L (2015) Transaction costs. In: Marciano A, Ramello GB (eds) Encyclopedia of law and economics. Springer, New York, pp 1–6

McAdams RH (1996) Groups norms, gossip and blackmail. Univ Pennsylvania Law Rev 144: 2237–2292

McAdams RH, Rasmusen EB (2007) Norms and the law. In: Polinsky AM, Shavell S (eds) Handbook of law and economics, vol 2. North-Holland, Amsterdam

Ogus AI (2004) Regulation: legal form and economic theory. Hart Publishing, Oxford

Picker RC (1997) Simple games in a complex world: a generative approach to the adoption of norms. Univ Chic Law Rev 64:1225–1287

Ponser EA (1996b) The regulation of groups: the influence of legal and non-legal sanctions on collective action. Univ Chic Law Rev 63:133–197

Posner EA (1996) The legal regulation of religious groups. Legal Theory 2(1):33–62

Posner EA (2000a) Law and social norms. Harvard University Press, Cambridge

Posner EA (2000b) Law and social norms: the case of tax compliance. Virginia Law Rev 86:1781–1820

Posner RA, Rasmusen E (1999) Creating and enforcing norms, with special reference to sanctions. Int Rev Law Econ 19:369–382

Schäfer HB (2001) Enforcement of contracts. DSE World Bank. Inst Found Mark Econ 3:141–146

Scott RE (2000) The limits of behavioral theories of law and social norms. Virginia Law Rev 86: 1603–1645

Shavel S (2002) Law versus morality as regulators of conduct. Am Law Econ Rev 4:227–257

Sunstein CR, Schkade D, Kahneman D (2000) Do people want optimal deterrence? J Leg Stud 29: 237–253

Swan GMP (2000) The economics of standardization. Standards and Technical Regulations Directorate Department of Trade and Industry, United Kingdom

Thomas MW (1948) The early factory legislation: a study in legislative and administrative evolution. Thames Bank Pub. Co, Leigh-on-Sea

Vandenbergh M (2003) Beyond elegance: a testable typology of social norms in corporate environmental compliance. Stanf Environ Law J 22:55–144

Weigel W (2006) Why promote the economic analysis of public law? Homo Econ 23(2):195–216

Novelty

► Innovation

Nudge

A Critical Perspective

Angela Ambrosino[1], Valeria Faralla[2] and Marco Novarese[3]
[1]Department of Economics and Statistics Cognetti de Martiis, University of Turin, Torino, Italy
[2]Dipartimento di Giurisprudenza e Scienze Politiche, Economiche e Sociali, Università del Piemonte Orientale, Alessandria, Italy
[3]Department of Law and Economics, University of Eastern Piedmont, Centre for Cognitive Economics, Alessandria, Italy

Abstract

In recent years, the nudge approach and the choice architecture design have become popular in policy and academic circles. The aim of this entry is to present a general overview of the theory of nudges and a critical appraisal of its application to practice, in policy and program as well as in research design, also with respect to possible alternative approaches.

Following the publication of "Nudge: Improving Decisions about Health, Wealth, and Happiness" (Thaler and Sunstein 2008), the term *nudge* has entered the economic and juridical focus in relation to public policies. Nudge refers to the idea that people's behavior can be mildly, or gently, pushed towards a certain course of action. As a result, nudge can be considered as a tool of political economics, using individual cognitive characteristics to stimulate people towards a positive action without restricting their freedom of choice. This is an alternative form of intervention to other government tools of public policies, such as incentives, rules, and constraints, or education and empowerment. The proponents of this theory consider nudge as a less coercive tool than regulatory efforts and incentive-based systems, although it is considered to be associated with better results if compared with free markets or traditional informative campaigns. The debate is deeply connected with the debate on paternalistic policies and the regulation approach and is also tied with bounded rationality. However, the interplay between all these aspects makes the definition of nudge problematic. Tahler and Sunstein, indeed, propose a definition which is still not overwhelmingly accepted (Rebonato 2012).

Policies inspired by nudge appear to be cheap and have the advantage of being experimentally testable and measurable (even if these policies can be tied with cultural and political issues such as the effects of the electricity consumption). As a result, nudge represents a novelty in the field. On the other hand, however, the use of nudges in real-life situations may involve the need to lie (or at least hide something) about the aim of actions that people are asked to perform. This clearly represents a point against the use of nudges.

The nudge theory is built on behavioral economics and on the idea that people have a bounded rationality, often do not have well-defined preferences, and are subject to a number of biases which cause people to make choices not in their own interest, sometimes also against their voluntary will. As for willpower, several examples exist: the desire to save and study, to consume healthier food, or avoid gambling activities. The desired action can be frustrating when lack of willpower is in action (Elster 1979).

But nudge is also based on other aspects linked to the decision-making mechanism. In many situations, indeed, people do not have well-defined preferences and, therefore, decisions become more complex. When there is a lack of clear preferences or inability to perform statistical computations and follow logic rules, people seem to be guided by simple rules (i.e., heuristics; see Kahneman 2012) in order to save mental energies. Specifically, the rational system of decision-making appears to be challenged (and often won) by a more impulsive system, faster, which is guided by simplifying strategies, such as information availability (the ease with which the side of a problem can be visible), and would give rise to distortions and simple heuristics. These distortions and heuristics tend to lead to systematic errors. The rational system (i.e., system 2) is connected with conscious and deliberate

decisions, whereas the other system (i.e., system 1) is highly unconscious and automatic. As a result, aspects which characterize the context of choice such as the number of alternatives available to individuals, the order of presentation of these alternatives, the way in which the choice is presented and framed, the existence of a default option, or the presence of a status quo alternative, which are typically considered as apparently irrelevant features become significant.

Nudge is then added to the debate over human rationality and about the influences on the decision-making process. A rational agent, as described by standard neoclassical normative models (von Neumann and Morgenstern 1947), is not affected by nudge intervention. However, in real-life situations, people appear to be not so rational (Kahneman 2003). In particular, the way through which the choice is designed and presented, the architecture of choice, becomes relevant.

Accordingly, the same mechanisms connected with the systematic errors can be used to direct people towards more healthy behaviors.

Nudge, which in this view represents a development of behavioral economics, has been already anticipated by a series of papers on the framing effect which has been found to elicit certain types of choices, rather than others, that are not necessarily for the good of individuals (e.g., marketing purposes). In this case, knowledge about decision-making processes can be used to persuade people (see Cialdini 2001a, b). The relationship between behavioral economics, law, and public policies developed early, also thanks to Kahneman (e.g., Kahneman et al. 1998; McCaffery et al. 1995). More recently, Parisi and Smith (2005) review the various implications of behavioral and cognitive economics for the law, such as the possibility to use cognitive biases to develop an approach that could be more effective than using monetary and nonmonetary rewards and punishments.

The subtitle of Thaler and Sunstein's book, "Improving Decisions about Health, Wealth, and Happiness," mainly refers to the measures aimed at pushing the limits of willpower. In front of a plate of peanuts, people tend to not restrain and eat, even when they do not want or do not need to. The act of pushing the plate in an unhandy place, which can be considered as a pre-commitment strategy, is an example of nudge. In their book, the authors suggest to take away the peanuts, but this appears to be a command-and-control form regulation, which is a mandate that cannot be avoided. This is in fact an example of the difficulty to define a nudge.

Moreover, the two authors propose simple examples of nudge that do not seem to improve individuals' well-being, at least in a direct and clear way. For example, the choice of an opt-in mechanism, rather than opt-out (i.e., explicit versus presumed consent), seems to increase organ donation rates because of the human tendency to remain at the status quo (Thaler and Sunstein 2008). The choice of a certain type of mechanism represents an example of choice architecture through which a public policy can be implemented and that can be considered as an alternative to public awareness campaign or strong enforcement. In fact, it seems to be true that nudging leaves margins of freedom to individuals and does not bind people to an action unless they have a clear opinion in one direction or another. People seem to be influenced by nudges when they have not a clear opinion and, rather, tend to conform to social norms and what they perceived as normality.

Considering the difference between various policies, this example shows the difficulty with the current definition of nudge but also the complexity of the discussion around the architecture of choice. A more general definition considers nudge as a tool for implementing public policies by means of behavioral economics.

Nudges' problems and critics come from different perspectives.

In their website, Thaler and Sunstein display the ballot used by Hitler as an example of bad nudge (Nudge Blog 2010). As stated before, nudge can indeed be used in order to persuade people to the right, as well as the wrong, direction. The problem is clearly that of understanding who decides what is the right/wrong direction. Thaler and Sunstein lead the main debate to the issue of strong paternalism and soft paternalism and to the

fairness of one approach over another. Strong paternalism assumes that the state knows and decides what is the best for people and then constrains them to have a coherent, consistent behavior. Strong paternalism is also connected with the idea of an omniscient legislator that acts rationally and is perfectly capable of using nudges and push people towards efficient behaviors. Soft paternalism, or libertarian, is instead based on the idea that we have to help people in getting their best interest in terms of well-being. Here is the heart of the debate between strong and soft paternalism and about the possibility to implement a soft, libertarian paternalism by means of a theory of nudges.

Meanwhile, the opponents of this theory consider nudge as a dangerous tool because, differently from more explicit and direct forms of regulation, it operates through unconscious channels (Kahneman 2012).

In line with other critical-oriented works, referring to different example of nudging, Rayner and Lang (2011) claim that nudge is not a new policy. In particular, they highlight how similar tools tend to deny the general idea of politics that requires informed choices and discussions, and deals with problems in an explicit and direct way. They also underline the fact that nudge can be applied to avoid taking actual choices (Rayner and Lang 2011).

Bovens (2009) precisely criticizes the lack of transparency of nudges and, particularly, the fact that choice architecture does not trigger real changes in individual preference or improve the use of willpower. More recently, Vallgårda (2012) criticizes the libertarian paternalism and denies the possibility to find an ethical reason for nudge, at least from a libertarian perspective. Furthermore, nudge would be anything new.

The fact that the theory of nudges exploits the bounded rationality is also controversial. The definition of nudge given by Rebonato (2012) underlines this aspect, which is instead not corroborated by Sunstein (2014).

Is it really impossible to develop consciously accepted forms of nudge and then give legitimacy to this type of intervention? Loewenstein et al. (2015) warn their subjects of the presence of a nudge, and in particular of a default option, in choosing between different medical treatments. The effect of the default option seems to persist, even when individuals are informed about its use. As a result, the effect was not necessarily connected with a lie.

Colander and Chong (2010) discuss instead about the benefits of giving directly to people, rather than to the government or an economist, the choice of being nudged to obtain the best result. This is the path to an actual libertarian paternalism, where individuals can freely choose their choice architecture.

A different criticism refers to the effectiveness and the strength of nudge. For instance, Loewenstein and Ubel (2010) address this question in an article published in the *New York Times*. The question was about the reduction of electricity consumption and the preference for a nudge, rather than an incentive-based intervention (i.e., by charging higher costs), to approach the problem. Does the fact that a particular nudge intervention has been found to have an effect in controlled experiments suggest that these effects can be observed in real-life situations? Even more important, does statistical significant results imply a real impact of nudge interventions?

This article also proves the importance of this issue, considering that it has been taken out of academic publishing and issued in one of the most important newspapers in the USA. In the UK, for instance, a nudge unit has been created. Furthermore, the Obama administration has pointed out the need for a wider use of behavioral economic techniques in the implementation of public policies.

But if people are wrong, is there a guarantee that the governments are not going to make similar errors? Other than ethical issues, with what legitimacy can they inspire paternalistic policies?

Grüne-Yanoff and Hertwig (2016) criticize the use of nudges in the perspective of an evolutionary rationality. After all, the bounded rationality is a critic to the economic mainstream which is considered as a reference point for a hypothetical rationality, one from which we go away. However, according to the evolutionary rationality, this critic should not imply that people make systematic mistakes and are basically unable to take

decisions. Moreover, people are not necessarily driven by nonmodifiable and unconscious mechanisms, as those represented by the alleged system 1. In certain circumstances, indeed, simple rules and heuristics can be smart and effective. In everyday life, we need rapid decisions and, differently from the theory, logic is not always applicable. Rather than rationality, we need the ability to survive and to be clever and also to interact with the environment. Errors can be made because the environment has changed over millennia, for human beings as well as other living creatures. Furthermore, errors can also depend from an incorrect way of framing problems. Specifically, the use of conditional probabilities in describing the Bayes' Theorem makes hard to handle it, even for experts. Considering the probabilities as relative frequencies can instead make people understand the actual risk they face, giving them an effective freedom of choice. Problems need to, and can, be formulated in a more clear and understandable way. The theory of nudge also moves from an idea of a mainstream behavior. That rationality is, however, not applicable to real life, which is instead characterized by different situations that cannot be simply described in terms of calculable risks and probabilities.

As a result, as an alternative to nudges, other authors propose the boost approach, whose aim is to extend people's decision-making skills and refine the decision-making environment rather than focus on distortions and biases. In particular, Grüne-Yanoff and Hertwig (2016) refer to nudge and boost policies as two different research programs. The first program is based on the analysis of systematic cognitive heuristics and biases, which are considered to result in poor choices, whereas the second is a simple heuristics program where bounded rationality is considered in its capacity, rather than inability, to produce good inference and choice.

Conclusion

Nudge is receiving an increasing and strong attention for both the development of the behavioral approach and the possibility (which the nudge approach seems to offer) to realize low-cost public policies, whose effectiveness can be proven.

It is hard to believe that the political debate around the paternalistic approach can ultimately have a solution. Indeed, this is a long-term discussed issue and nudge tends to complicate, rather than simplify, its resolution.

The problem lies in the fact that different forms of nudge exist and each form is associated with different critical issues. At the same time, however, most discussions tend to simplify and take into account only specific examples.

Kahneman (2012) underlines how writing with a difficult-to-read font can focus attention and decrease the possibility of errors occurring in the solution of a problem. This is an example of nudge that increases the ability of people to influence for the better the relationship between individual well-being and choices.

Cross-References

- ▶ Behavioral Law and Economics
- ▶ Financial Education

References

Bovens L (2009) The ethics of nudge. In: Grüne-Yanoff T, Hansson SO (eds) Preference change: approaches from philosophy, economics and psychology. Theory and decision library A(42). Springer, New York, pp 207–220

Cialdini RB (2001a) The science of persuasion. Sci Am 284(2):76–81

Cialdini RB (2001b) Influence: science and practice. Allyn & Bacon, Boston

Colander D, Chong AQL (2010) The choice architecture of choice architecture: toward a Non-paternalistic nudge policy. Department of Economics Middlebury College Middlebury, Middlebury college economics discussion paper no. 10–36. Available at http://sandcat.middlebury.edu/econ/repec/mdl/ancoec/1036.pdf

Elster J (1979) Ulysses and the sirens. Cambridge University Press, Cambridge

Grüne-Yanoff T, Hertwig R (2016) Nudge versus boost: how coherent are policy and theory? Minds Mach 26(1):149–183

Kahneman D (2003) Maps of bounded rationality: a perspective on intuitive judgement and choice. Am Econ Rev 93(2):162–168

Kahneman D (2012) Thinking, fast and slow. Penguin Books, London

Kahneman D, Schkade D, Sunstein C (1998) Shared outrage and erratic awards: the psychology of punitive damages. J Risk Uncertain 16(1):49–86
Loewenstein G, Ubel P (2010) Economics behaving badly. N Y Times, July 14. Retrieved from http://www.nytimes.com/2010/07/15/opinion/15loewenstein.html
Loewenstein G, Bryce C, Hagmann D, Rajpal S (2015) Warning: you are about to be nudged. Behav Sci Policy 1(1):35–42
McCaffery EJ, Kahneman D, Spitzer ML (1995) Framing the jury: cognitive perspectives on pain and suffering awards. Virginia Law Rev 81(5):1341–1420
Nudge Blog (2010) The Nazis nudged. [Blog post], August 10. Retrieved from http://nudges.org/2010/08/11/the-nazis-nudged/
Parisi F, Smith VL (2005) The law and economics of irrational behavior. Stanford University Press, Stanford
Rayner G, Lang T (2011) Is nudge an effective public health strategy to tackle obesity? No. Br Med J 342: d2177
Rebonato R (2012) Taking liberties: a critical examination of libertarian paternalism. Palgrave Macmillan, New York
Sunstein CR (2014) Why nudge? The politics of libertarian paternalism. Yale University Press, New Haven
Thaler RH, Sunstein CR (2008) Nudge: improving decisions about health, wealth, and happiness. Yale University Press, New Haven
Vallgårda S (2012) Nudge: a new and better way to improve health? Health Policy 104(2):200–203
von Neumann J, Morgenstern O (1947) Theory of games and economic behaviour. Princeton University Press, Princeton

Nuisance

Keith N. Hylton
Boston University School of Law, Boston, MA, USA

Abstract

This entry sets out the law and the economic theory of nuisance. Nuisance law serves a regulatory function: it induces actors to choose the socially preferred level of an activity by imposing liability when the externalized costs of the activity are substantially greater than the externalized benefits or not reciprocal to other background external costs. Proximate cause doctrine plays a role in supplementing nuisance law.

Introduction

Nuisance law has suffered from the difficulty scholars have encountered in attempting to codify it in the form of simple rules and to understand the functions that the law serves. Prosser (1971) once described nuisance as an impenetrable jungle, and the dearth of efforts to state it in the form of black letter rules suggests that this opinion has been shared by many legal scholars. The process of scholarly codification, that is, of taking a mass of seemingly inconsistent court decisions and generating from them a set of clear legal propositions, has been slow in the area nuisance law.

Scholarly codification and an understanding of function are likely to occur contemporaneously. When courts and legal scholars have a firm understanding of the functions of a legal doctrine, it is relatively easy for them to summarize it in the form of simple rules. For example, the "Hand Formula" of *US v. Carroll Towing* is a summary of the negligence test that reflects a widely accepted understanding of the function of negligence law.

In recent years, research has focused on the function of nuisance law. I will set out the functional approach here, which uses economic analysis to understand nuisance doctrine at a high level of detail (Hylton 2011).

Earlier efforts have been made to provide an economic theory of nuisance law. Most of those early efforts, stemming from Coase (1960), have relied on the theory of transaction costs to explain the functional distinction between nuisance and trespass law (Calabresi and Melamed 1972; Merrill 1985; Smith 2004). But the core of nuisance law consists of balancing tests and limitations on the scope of liability that are not easily understood on the basis of transaction-cost theory. These tests and limitations are better understood using externality analysis.

Far from being an impenetrable jungle, nuisance law is a coherent body of rules that serves a socially desirable function. Nuisance law optimally regulates activity levels. The law induces actors to choose the socially optimal level of an activity by imposing liability when the externalized costs (of the activity) are substantially in

excess of externalized benefits or far in excess of background external costs. Proximate cause doctrine plays an important supplementary role to nuisance doctrine in regulating activity levels.

Nuisance Law

A nuisance is typically defined as an intentional, unreasonable, nontrespassory invasion of the quiet use and enjoyment of property. Each one of these terms has a special meaning in the law. Most of the terms are easily understood in terms of their general use in tort law. However, the term "unreasonable" is the most tricky concept, because there is equivalent use of the same term in other parts of tort law.

Take, for example, the word "intentional" in the definition of a nuisance. Intentional in nuisance law has a meaning that is not very different from its meaning in other areas of tort law. Typically the defendant is guilty of an intentional nuisance if he is aware of the invasion. The law does not require the defendant to have set out to harm the plaintiff.

Similarly, nontrespassory has a meaning that is readily ascertainable from the tort's case law. A trespassory invasion is one that displaces the plaintiff from all or some portion of his property. For example, a large rock that is thrown over to the plaintiff's property displaces the plaintiff from the space in which it travels and ultimately lands. This can be contrasted with a nontrespassory invasion, such as smoke or noise, which does not displace or oust the plaintiff from any space on his property.

The difficulty arises with the term "unreasonable." As a result, efforts to state nuisance law in the form of simple rules have been sparse and for the most part unsuccessful. The best known effort to codify nuisance doctrine is the Restatement Second of Torts § 826, 1977, which says:

> An intentional invasion of another's interest in the use and enjoyment of land is unreasonable if:
> (a) the gravity of the harm outweighs the utility of the actor's conduct, or
> (b) the harm caused by the conduct is serious and the financial burden of compensating for this and similar harm to others would not make the continuation of the conduct not feasible.

However, the Second Restatement's § 826 is of questionable value because it refers to the actor's *conduct* rather than his *activity*. One can draw an important distinction between these terms. Conduct often refers to an action or a series of actions within a short time span. Activity refers more broadly to an occupation or a significant pastime. For example, batting a baseball is a type of conduct, while playing professional baseball is an activity.

The reference to conduct could easily lead readers to believe that Restatement § 826 is equivalent to the balancing test observed in negligence law – i.e., the Hand Formula of *Carroll Towing*. The balancing test known as the Hand Formula says that the defendant is negligent if he fails to take care in a setting where his additional care would have been less costly than the additional injury costs that would have been avoided by that care. The language of Section 826 is easy to confuse with the analysis required by the Hand Formula.

If, instead, "conduct" in Restatement § 826 is understood to mean "activity," then it becomes difficult to understand how the balancing test announced in 826 should be conducted. How is it possible to compare the gravity of the victim's harm to the utility of the defendant's activity? Suppose, again, that we are talking about baseball. A ball is hit out of the baseball yard and injures a passerby on the street. How should one go about comparing the gravity of the victim's injury to the utility of playing baseball? Because this is so difficult to answer, Section 826 provides little guidance to lawyers and judges.

Moreover, part (b) of Restatement § 826 implies that strict liability should be applied to any activity that causes a "serious" interference with the plaintiff's use and enjoyment of property, provided that the activity would not be bankrupted by such liability. This implies, strangely, that a thinly capitalized activity has an advantage under nuisance law, because it appears to immunize activities that would be bankrupted by a claim for damages. The difficult question in nuisance law is how to balance the external risks and the external benefits of an activity, a question which Section 826 does not even begin to address.

The following test, based on Restatement Second § 520, provides a better summary of nuisance doctrine:

In order to determine if an invasion is unreasonable under nuisance law, the following factors should be examined:

(a) Existence of a high degree of interference with the quiet use and enjoyment of land of others
(b) Inability to eliminate the interference by the exercise of reasonable care
(c) Extent to which the activity is not a matter of common usage
(d) Inappropriateness of the activity to the place where it is carried on
(e) Extent to which its value to the community is outweighed by its obnoxious attributes

In the remainder of this entry, I will set out the basic economic theory of nuisance doctrine and explain why it is generally consistent with these factors.

Economics of Nuisances

The literature on the economics of nuisance law can be divided into two branches. One is the *transaction-cost framework*, which began with Coase's discussion of nuisance in his famous article on transaction costs and resource allocation. The transaction cost approach emphasizes the functional differences between nuisance and trespass law and provides a positive theory of the boundary between nuisance and trespass (Merrill 1985; Smith 2004). It has also been applied to explain the law on priority, often described as "coming to the nuisance" (Wittman 1980; Snyder and Pitchford 2003).

The other branch of work on the economics of nuisance law can be labeled the *externality model*, which focuses on the regulatory function of nuisance law (Hylton 1996, 2008, 2010). The notion that liability rules can be used to control externalities has been well understood for a long time in the law and economics literature (Polinsky 1979). The externality approach offers a model of the function of nuisance liability and a positive theory of the core doctrines of nuisance. The core doctrines examined under the externality model are those of intent, reasonableness, and proximate cause. I will focus on the externality model below and offer a few remarks reconciling the transaction cost and externality models at the end.

Activity Levels, Care Levels, and Externalities

The law and economics literature distinguishes care and activity levels (Shavell 1980). The care level refers to the level of instantaneous precaution that an actor takes when engaged in some activity. For example, an actor can take more care while in the activity of driving by moderating his speed or looking more frequently to both sides of the road. The activity level refers to the actor's decision with respect to the frequency or location of his activity. If, for example, the activity of concern is driving, it can be reduced by driving less frequently.

The invasions associated with nuisance law are external costs connected with activity level choices. Consider, for example, a manufacturer who dumps toxic chemicals into the water, as a byproduct of manufacturing. Suppose the manufacturer is taking the level of care required by negligence law (reasonable care), and, in spite of this, the manufacturing process leads to some discharge of toxic chemicals. In this case, the environmental harm is a negative externality associated with the manufacturer's activity level choice.

The framework below is of activities that impose external costs on society even when they are carried out with reasonable care (Hylton 2008). The question examined is how the law can regulate activity levels in a way that leads to socially optimal decisions.

The Economics of Activity Level Choices

Assume that there are two liability rules that can be applied to actors, strict liability and negligence. Under either rule, actors are assumed to take reasonable care.

For any activity, the actor engaged in it will set his privately optimal activity level at the point which maximizes his utility from that activity. That means the actor will consider the benefits

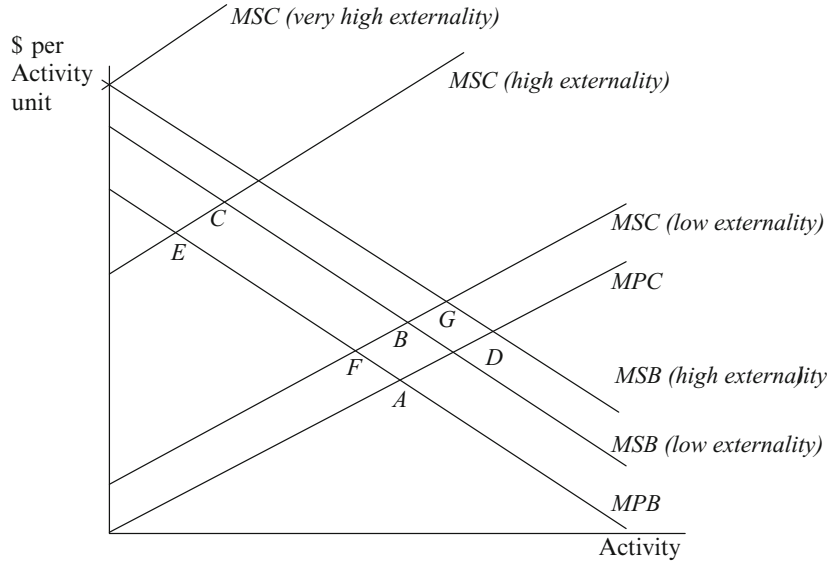
he derives from the activity as well as the costs and choose a level at which the excess of private benefits over private costs is at its maximum. If we let *MPB* represent the incremental or marginal private benefits to the actor from his activity, and *MPC* represent the incremental private costs to the actor from increasing the scale of activity, the actor will increase his activity level as long as the marginal private benefit of an additional unit of activity exceeds the marginal private cost ($MPB > MPC$). The privately optimal level of activity is the level at which the marginal private benefit to the actor is just equal to the marginal private cost ($MPB = MPC$).

The diagram labeled Fig. 1 can be used to illustrate this argument. Assuming marginal benefits diminish as the actor increases his activity level, the marginal private benefit schedule can be represented by a downward sloping line, as shown in Fig. 1. The marginal private cost schedule is assumed to increase as the actor increases his level of activity (see *MPC* in Fig. 1). The reason for this is that the incremental cost of the activity goes up as the actor increases his scale. For example, if the activity is driving, the upward-sloping *MPC* schedule assumes that it is more costly to go from 50 miles per week to 51 than to go from 10 miles per week to 11. (Of course, this assumption may not be valid in some cases. The incremental cost of going from 50 to 51 miles per week may be the same, in some cases, as the incremental costs of going from 10 to 11, but the results of this analysis are not dependent on this assumption of increasing marginal cost.)

The actor's privately optimal activity level choice is given by the intersection of the marginal private benefit and marginal private cost schedules, shown by point *A* in Fig. 1. At the intersection point, the net benefits (excess of private benefits over private costs) are at its maximum.

Now I introduce externalities. On the cost side, there are negative externalities (or external costs) associated with many activities. Consider, for example, driving. With each mile driven, the actor imposes some risk of harm from an accident or from pollution on the public in general. Or, if the activity is manufacturing, with each widget produced, a manufacturer who discharges chemicals in the water imposes cleanup costs on others. The marginal social cost of the actor's activity is simply the sum of the marginal private cost and the marginal external cost imposed on society.

On the benefit side, it is possible that there are benefits to society generated by the actor's activity. In the manufacturing case, suppose that instead of producing widgets, the manufacturer is



Nuisance, Fig. 1 Activity levels and externalities

producing a vaccine for some communicable disease. Vaccines cast off substantial external benefits by reducing the risk of disease even to the unvaccinated. The marginal social benefit is the sum of the marginal private benefit and the marginal external benefit of an additional unit of activity.

The final step of this introduction to the economics of activity level choices is to consider the differences between private and social incentives. Consider the case of low externalities on both the cost and benefit sides. Suppose there are external costs and external benefits connected to the activity, but they are relatively modest. They are shown in Fig. 1 by *MSC* (*low externality*) and *MSB* (*low externality*). The socially optimal level of activity, which equates marginal social benefit and marginal social cost, is found at the point *B* in Fig. 1. The socially optimal level of activity (*B*) is roughly the same as the privately optimal level of activity (*A*). The reason is that the modest positive and negative externalities cancel each other out. Given this, there is no reason for the law to intervene to try to reduce the level of activity.

The case just considered is similar to that of an "irrelevant externality" (Buchanan and Stubblebine 1962; Haddock 2007). Although there is an external cost, society should not try to correct it because there is an offsetting external benefit. Buchanan and Stubblebine emphasized the case of offsetting internal benefit, but the concept of an irrelevant externality is equally valid if there is an offsetting external benefit.

Now consider the case of high externality on the cost side and low externality on the benefit side. This is shown by the intersection of the *MSC* (*high externality*) and *MSB* (*low externality*), point *C* in Fig. 1. In this case there is a wide divergence between the privately optimal level of activity (point *A*) and the socially optimal level of activity (point *C*). This case is one in which it appears desirable for the law to intervene to reduce the level of activity. Indeed, in the case of very high externality on the cost side (see *MSC* (*very high externality*)), it may be desirable to shut down the activity completely.

Consider lastly the case of low externality on the cost side and high externality on the benefit side. The intersection of the marginal social cost and marginal social benefit schedules occurs at point *D* in Fig. 1. In this case, the privately optimal level of activity (*A*) is substantially below the socially optimal level (*D*). The law should intervene to increase the actor's level of activity.

Law

I have considered external costs and external benefits associated with activities conducted with reasonable care. Since the actors are assumed to be exercising reasonable care, the negligence rule cannot influence their activity level choices. The negligence rule holds the actor liable only when he fails to take reasonable care.

Strict liability has the property that it imposes liability on actors even when they have taken reasonable care. The legal system can regulate activity levels through imposing strict liability. This part examines the conditions under which strict liability leads to optimal activity levels.

First, consider the case in which externality is high on the cost side and low on the benefit side. The socially optimal scale is point *C* in Fig. 1. In the absence of strict liability, the privately optimal scale is point *A*. Imposing strict liability on the actor is desirable in this case. When strict liability is imposed on the actor, his marginal private cost schedule becomes equivalent to the marginal social cost schedule.

In the case of high externality on the cost side coupled with low externality on the benefit side, the actor's privately optimal activity level under strict liability will be point *E*. It is not exactly the optimal level, which is at point *C*, but it is close. Social welfare will most likely be improved by using liability to lead the actor to produce at scale *E* rather than at the socially excessive scale *A*.

Consider the case in which externality is low both on the cost and on the benefit side. The socially optimal scale of activity is associated with point *B*. The privately optimal level of activity is associated with point *A*. These are the same activity levels. If strict liability is imposed on the actor, it will reduce his activity level below the socially optimal scale and therefore reduce social welfare. Strict liability will cause the actor to

choose the scale *F*, which is below the socially optimal scale.

This analysis implies that *strict liability is desirable only when the external cost of the actor's activity substantially exceeds the external benefit associated with the actor's activity*. In this case imposing strict liability reduces activity levels to a point that is closer to the socially optimal scale than would be observed under the negligence rule. When the external benefits are roughly equal to or greater than the social costs associated with the actor's activity, strict liability is not socially desirable.

Another case in which strict liability is not socially desirable is observed when two actors cross-externalize equivalent costs. Put another way, *when the costs externalized by two actors to each other are reciprocal*, *strict liability is not socially preferable to negligence*. The reason is that under strict liability, you will pay for harms to others, while under negligence (when everyone is complying with the negligence standard), you will pay only for the harms you suffer. Since those harms are the same, activity levels will not differ under the two regimes (Hylton 2008, 2011).

Application to Law: Nuisance and Abnormally Dangerous Activities

To this point, I have presented a model of the economics of externalities and considered its implications for law. In this part, I will examine the extent to which the law conforms to the predictions of the model.

Abnormally Dangerous Activities

The most straightforward application of this model is to the law of abnormally dangerous activities (e.g., blasting). To determine whether an activity is abnormally dangerous, Section 520 of the Restatement (Second) of Torts provides the following factors:

(a) Existence of a high degree of risk of some harm to the person, land, or chattels of others
(b) Likelihood that the harm that results from it will be great
(c) Inability to eliminate the risk by the exercise of reasonable care
(d) Extent to which the activity is not a matter of common usage
(e) Inappropriateness of the activity to the place where it is carried on
(f) Extent to which its value to the community is outweighed by its dangerous attributes

The provisions of Section 520 are in line with the theory set out in the previous part of this entry. First, note that Section 520 can be divided into two parts, the first three provisions and the last three provisions. The first three provisions govern the degree of residual risk and imply that strict liability for operating an abnormally dangerous activity is appropriate only when the residual risk – the risk that remains after the actor takes reasonable care – is high. If the residual risk of the actor's activity is high, strict liability may be appropriate. On the other hand, if the residual risk is relatively low, strict liability would be inappropriate under Section 520. Judge Richard Posner famously applied this component of Section 520 to hold that strict liability would be inappropriate in *Indiana Harbor Belt Railroad Co. v. American Cyanamid Co*., 916 F.2d 1174 (7th Cir. 1990).

The final three provisions of Section 520 line up with the language in *Rylands v. Fletcher*, L.R. 3 H.L. 330 (1868), which provides the foundation for the law on abnormally dangerous activities. The third factor, common usage, helps us identify activities for which the risks are reciprocal to those of other common activities. If an activity is one of common usage, then actors engaged in those activities will impose reciprocal risks on each other, and there is therefore no basis for adopting strict liability over negligence. The fourth factor, inappropriateness, is another way of determining whether the activity imposes a reciprocated risk. The last provision, comparing benefits and risks, guides courts to compare the external benefits thrown off by the activity with the external costs. If the external costs are great relative to the external benefits, strict liability is appropriate under this provision.

Consider an example. If the actor holds a lion as a pet in his backyard, he will inevitably impose a great risk on his neighbors. Moreover, it is a risk

that remains great even after the actor has taken reasonable care. For this reason, holding a lion as a pet satisfies the first three elements of the Section 520 strict liability test. The last three elements are also satisfied. Holding a lion as a pet is not a common activity – the risk the lion holder externalizes to his neighbors is not equivalent to the risk they externalize to him. The benefits externalized to neighbors from holding a lion as a pet are likely to be far less than the risks externalized to them. For these reasons, it is appropriate under the theory presented here and under Section 520 to apply strict liability to the activity of holding lions as pets.

Nuisance

The law on abnormally dangerous activities is the most obvious application of the theory of this entry. However, the theory applies equally well to nuisance law, the subject of this entry. Most of the standard environmental interferences, such as air or water pollution, have been treated as nuisances under tort law.

The theory of this entry suggests a clear interpretation for the rules governing nuisance law. First, consider the basic legal definition of a nuisance: an intentional, nontrespassory, and unreasonable invasion into the quiet use and enjoyment of property. Intentional, in nuisance law, has had a meaning very similar to its use in the context of trespass law: it is enough if the defendant was aware of the nuisance. There is no need, on the part of the plaintiff, to prove that the defendant aimed to harm the plaintiff. The term nontrespassory has always had the effect of distinguishing between invasions that interfere with exclusive possession of property or a portion of it (e.g., a boulder hurled onto the plaintiff's property) and invasions that merely make it less desirable to remain in possession of property (e.g., smoke).

Perhaps the most important term in the definition of nuisance is unreasonable. The theory of this entry suggests that the factors of Section 520 are equally applicable to nuisance disputes. Paraphrasing Section 520, the appropriate test for unreasonableness under nuisance law can be articulated as follows:

(a) Existence of a high degree of interference with the quiet use and enjoyment of land of others
(b) Likelihood that the harm resulting from that interference will be substantial to the typical member of the community
(c) Inability to eliminate the interference by the exercise of reasonable care
(d) Extent to which the activity is not a matter of common usage
(e) Inappropriateness of the activity to the place where it is carried on
(f) Extent to which its value to the community is outweighed by its dangerous attributes

The first three factors of this test require that the interference be substantial even when the actor is taking reasonable care. As in the case of abnormally dangerous activities, the first three factors should be treated as minimal requirements for nuisance liability.

The last factor asks the court to compare the benefits externalized by the activity and the costs externalized. When the benefits are substantial, the last factor suggests that the court should be reluctant to impose liability on a nuisance theory. Consider, for example, the noise generated by a busy fire station. The noise generated by fire trucks constantly moving in and out of the station with their alarms running could be deemed to substantially interfere with the quiet use and enjoyment of land by neighbors. However, the neighbors also benefit by being located close to the fire station. Since those benefits are substantial and widely dispersed, the neighbors should not be allowed to impose strict liability on a nuisance theory against the fire station, the conclusion reached in *Malhame v. Borough of Demarest*, 392 A. 2d 652 (Law Div. 1978). There is no economic basis for using liability as an incentive to force the fire station to reconsider its location decision.

Nuisance law does not provide for compensation to the extra-sensitive plaintiff (*Rogers v. Elliott*, 15 N.E. 768 (Mass. 1888)). The justification for this well-settled piece of the law is best understood in terms of the model of this entry. A nuisance exists, under the model here, when the externalized costs associated with an activity are

substantially in excess of externalized benefits. The comparison of externalized costs and benefits is made with respect to statistical averages, not to any particular plaintiff. If, on the basis of statistical averages, the externalized costs associated with an activity are not substantially greater than the externalized benefits, then the activity is not a nuisance under the theory here. If a particular plaintiff suffers a severe injury under these conditions, that harm may be actionable under some other legal theory, such as negligence, but it is not actionable under nuisance law.

Local conditions play an important role in nuisance law. In particular, the last three factors (d, e, and f) of the test proposed here depend on local conditions. Most environmental pollutants are regulated because of the risk of harm they impose on people located near the source. In most cases, the risk of harm declines as people move further from the source. Thus, externalized costs are likely to be substantial near the source and declining to zero as one moves further away. Strict liability provides incentives for the pollution generator to locate in regions in which externalized costs are insignificant.

Under the proximate cause rule, courts have limited the scope of nuisance liability to injuries that are connected in a predictable way to the externalized risk. Injuries that are not predictably related to the externalized risk are not within the scope of strict nuisance liability. The externality model suggests a reason for this: to focus liability on the cost-externalizing features of the defendant's activity rather than the activity per se. Suppose the victim drives his car into the defendant's malarial pond. To permit a strict liability action would fail to tax the defendant's activity for the specific risk creation – i.e., the risk of malaria – that nuisance law aims to discourage.

A clearer justification for the proximate cause rule in nuisance law can be based on the model of the previous section. Let the externalized risk component be separated into two subcomponents, where one is the normal risk externalized by activities of the defendant's type and the other is the extraordinary risk that makes the defendant's activity a nuisance. For example, in the case of a malarial pond, the normal part is the risk externalized by any water storage, and the extraordinary part is the malaria risk. The proximate cause rule excludes liability for the normal-risk component. If, as nuisance law implicitly assumes, normal risks are balanced off by (normal) positive externalities, then excluding liability for normal risk leads to optimal activity levels.

The proximate cause rule leads to the social optimum in activity by excluding the normal-risk component as a source of liability. In terms of Fig. 1, the "low externality" cost increment (*MSC* (*low externality*)) is representative of the normal risk. If normal positive externalities are present, so that *MSB* (*low externality*) measures the marginal social benefit of the activity, the socially optimal activity level (assuming the risk consists of both the normal and extraordinary components) is that associated with point *C*. However, strict liability applied without any offset based on the proximate cause rule would lead the actor to choose the activity level associated with point *E*. Applying the proximate cause rule of nuisance law, which limits application of strict liability to those injuries attributable to the extraordinary risk, leads the actor to choose the socially optimal activity level (point *C*). Thus, the proximate cause rule improves on strict liability by leading to an optimal imposition of liability.

Coming to the Nuisance

Sometimes defendants argue that plaintiffs should not be able to recover because they "came to the nuisance." The coming-to-the-nuisance defense is valid in some cases but not in all. The theory of this entry provides a justification for the ambiguous treatment of the coming-to-the-nuisance defense.

Since the goal of nuisance liability is to optimally regulate activity levels, a victim's decision to come to the nuisance is certainly a relevant piece of information. The victim's decision to move is no different from the case of the buyer who contracts with a seller to purchase some item with a latent and dangerous defect. If the buyers are aware of the negative feature of the product, then the resulting market equilibrium would be

socially optimal. Similarly, if a smoke-belching factory sits alone in an area, and the victim moves next door to it, there would be no reason to view the factory's activity as socially excessive. In this case, the coming-to-the-nuisance defense applies.

There are two reasons that the coming-to-the-nuisance defense might not be desirable in this model. First, the victim may not have been aware of the offender's activity when purchasing his property. In *Ensign v. Walls*, 34 N.W.2d 549 (Mich. 1948), the defendant maintained dog-breeding business in the residential area of Detroit. The invasions (odors, noise, occasional escapes, filth) caused by the defendant's activity may not have been obvious to prospective residents; most probably became aware of the nuisance only after moving in.

The second reason the coming-to-the-nuisance defense may not be socially desirable is that the market for real property can be distinguished from most other markets for goods or services. Suppose the community consists of one smoke-belching factory and 99 residents. It is clear in this case that the reciprocal harm condition would not be satisfied; the background risks externalized by the residents would be trivial in comparison to the cost externalized by the factory. If the coming-to-the-nuisance defense were allowed, there would be no mechanism to control the activity level of the factory. The factory could double its level of activity without meeting any liability.

The justifications for the law on priority based on the externality model do not diminish the more traditional transaction-cost-based understanding. A rule favoring priority would encourage socially wasteful races and expropriation (Wittman 1980; Snyder and Pitchford 2003). Snyder and Pitchford (2003) distinguish their analysis from the seminal analysis of Wittman (1980) by noting that their model allows for the court to have limited information and for low transaction costs between the parties after the first move invests. Smith (2004) addresses nuisance law generally from the perspective of information costs, arguing that exclusion rules are favored by the law because they facilitate the production and disclosure of information.

Transaction-Cost Model Versus Externality Model

A complete economic model of nuisance law would consist of the transaction-cost model and the externality model, with the transaction-cost model used to explain the boundaries of nuisance law and the externality model used to explain its regulatory function. The foregoing analysis focuses less on the boundary question that has been the focus of transaction-cost analysis and more on the regulatory function of nuisance.

I have already noted some of the boundary questions examined under the transaction-cost model, specifically the choice between trespass and nuisance and the rule on priority. The transaction-cost model appears to be superior to the externality model as a theory of the boundary between nuisance and trespass law. However, both the transaction-cost and externality models provide justifications for the law's treatment of priority.

One other boundary question, unexamined so far, is the exclusion of protection under nuisance law for aesthetic interests, such as the right to sunlight or to a view of the mountains, *Fontainebleau Hotel Corp. v. Forty-Five Twenty-Five, Inc.*, 114 So. 2d 357 (Fla. App. 1959). The exclusion of aesthetic interests appears to be better explained by the transaction-cost model than by the externality model. It is obviously an externality, in the technical sense, when a landowner erects a fence that blocks the sunlight to another adjacent landowner. There is no reason suggested by the externality model for not treating the harm to the adjacent landowner as potentially a nuisance.

Under the transaction-cost model, there is a clear economic case for excluding liability for aesthetic harms. If aesthetic interests were protected by nuisance law, there would immediately be questions of information and proof. If one adjacent landowner can sue the owner of a hotel for blocking sunlight, why not allow other adjacent landowners? The transaction costs of resolving these disputes in the bargaining process would be enormous. On the other hand, if the law refuses to protect aesthetic interests, then the transaction

costs of resolving disputes would be much more manageable.

Conclusion

Nuisance law is complicated and covers a wide array of land use disputes. However, at its core, nuisance is simple. The law generates optimal activity levels by imposing strict liability when externalized risks are far in excess of externalized benefits or far in excess of background risks. Nuisance doctrine is consistent with this theory.

References

Buchanan JM, William C (1962) Stubblebine, externality. Economica 21:371
Calabresi G, Douglas Melamed A (1972) Property rules, liability rules, and inalienability: one view of the cathedral. Harv L Rev 85:1089
Coase RH (1960) The problem of social cost. J Law Econ 3:1
Haddock DD (2007) Haddock, irrelevant externality angst. J Interdiscip Econ 19:3–18
Hylton KN (1996) A missing markets theory of tort law. Nw U L Rev 90:977–1008
Hylton KN (2008) A positive theory of strict liability. Rev Law Econ 4:153–181
Hylton KN (2010) The economics of public nuisance and the new enforcement actions. Supreme Court Econ Rev 43(1):43–76
Hylton KN (2011) The economics of nuisance law. In: Ayotte K, Smith HE (ed) Research handbook on the economics of property law, Edward Elgar Publishing, pp 323–343
Merrill TW (1985) Trespass, nuisance, and the costs of determining property rights. J Leg Stud 14(1):13
Polinsky AM (1979) Controlling externalities and protecting entitlements: property right, liability rule, and tax-subsidy approaches. J Leg Stud 8(1):1–48
Prosser WL (1971) Handbook of the law of torts, 4th ed. p 571
Restatement (Second) of Torts: unreasonableness of intentional invasion § 826 (1977)
Shavell S (1980) Strict liability versus negligence. J Leg Stud 9(1):1–25
Smith HE (2004) Exclusion and Property Rules in the Law of Nuisance. Va L Rev 90:965
Snyder CM, Pitchford R (2003) Coming to the Nuisance: An Economic Analysis from an Incomplete Contracts Perspective. J Law Econ Org 19(2):491–516
Wittman DA (1980) First Come, First Served: An Economic Analysis of "Coming to the Nuisance". J Leg Stud 9:557–568

Nuisance Lawsuits

▶ Frivolous Suits