A

Absolute Priority

Alexander Dilger
Institute for Organisational Economics,
University of Münster, Münster, Germany

Synonyms

Liquidation Preference

Definition

Absolute priority is a rule by which different classes of creditors are paid in full one after another in a corporate insolvency and the owners are paid last if anything remains.

Explanation

Absolute priority means that in a corporate insolvency, different classes of creditors are paid one after another, while the (former) owners of an insolvent firm get the rest if anything remains after all creditors have been paid in full. Normally this means that the owners get nothing because otherwise the company would not be insolvent in the first place. Accordingly, absolute priority is not relevant for solvent companies because they pay everything they owe their creditors and the owners own the rest. If a company cannot pay everything it owes because its liabilities are higher than the value of all its assets, then it is insolvent and all its debts are due immediately. Following absolute priority the most senior creditors are paid first. If the value of the insolvent firm is high enough, the most senior creditors are paid in full and the second highest class of creditors is next. Otherwise, if the value of the insolvent firm is lower than all the claims of the most senior creditors, these claims are satisfied proportionally and all other creditors with lower priority as well as the owners get nothing. In general, one class of creditors is partially satisfied, while all more senior creditors are satisfied completely and all more junior creditors get nothing. Conversely, it is a violation of absolute priority if any junior creditor or an owner gets anything before the more senior claims have been paid in full.

It is quite easy to comply with absolute priority as long as the value of the insolvent company is known. This is especially the case after the liquidation of a firm when all its assets have been sold. However, in many cases, the value of a firm is higher as a going concern. When it is sold as a whole, the total revenue is known and can be distributed according to absolute priority. Yet the market for insolvent companies may be thin such that only a fire sale with a depressed price is possible (cf. Gale and Gottardi 2011) and it is better for all creditors together to continue the business by themselves at least for a while. In this case absolute priority requires an estimation of the company's value. In estimating this value, there is a conflict

© Springer Science+Business Media, LLC, part of Springer Nature 2019
A. Marciano, G. B. Ramello (eds.), *Encyclopedia of Law and Economics*,
https://doi.org/10.1007/978-1-4614-7753-2

of interest because the most senior creditors get a higher share of this company when the valuation is low, whereas the more junior creditors prefer a higher valuation to get more (than nothing). Bebchuk (1988) proposes the use of options as a fair solution to this problem. The most senior creditors get all new shares of the then debt-free company, while more junior creditors or even owners can buy shares by paying the corresponding ratio of more senior claims. Dilger (2006) shows that one auction of all shares is both simpler and fairer than Bebchuk's approach with options.

Besides the question on how to observe absolute priority, one may ask why it is important. That creditors should be more senior than owners in a bankruptcy is part of any meaningful debt contract. The differences between creditors (or also different categories of owners) such that some are more senior than others can also be arranged by contract or regulated by law. Violations of absolute priority are then a breach of these contracts or laws. Thus, they limit the freedom of contract and are inefficient ex ante (cf. Longhofer 1997; Bebchuk 2002). However, strictly maintaining absolute priority can be inefficient ex post if there are costly conflicts while determining the value of a company (cf. Baird and Bernstein 2006) or if nobody has sufficient incentives to begin or end a bankruptcy procedure (cf. Baird 1991). Empirically one can observe many deviations from absolute priority (cf. Weiss 1990). They seem to be fair and efficient as long as the more senior creditors agree. Thus, a veto power of them, individually or even by majority rule by creditor classes, against violations of absolute priority is better than a strict adherence to it in any case (cf. Warren 1991). Although there is much variation, the bankruptcy laws in many countries are in accordance with this (cf. Campbell 1992; Claessens and Klapper 2005).

References

Baird DG (1991) The initiation problem in bankruptcy. Int Rev Law Econ 11:223–232

Baird DG, Bernstein DS (2006) Absolute priority, valuation uncertainty, and the reorganization bargain. Yale Law J 115:1930–1970

Bebchuk LA (1988) A new approach to corporate reorganizations. Harv Law Rev 101:775–804

Bebchuk LA (2002) Ex ante costs of violating absolute priority in bankruptcy. J Financ 57:445–460

Campbell D (ed) (1992) International corporate insolvency law. Butterworths Law, London

Claessens S, Klapper LF (2005) Bankruptcy around the world: explanations of its relative use. Am Law Econ Rev 7:253–283

Dilger A (2006) Forced to make mistakes: reasons for complaining about Bebchuk's scheme and other market-oriented insolvency procedures. Eur J Law Econ 21:79–94

Gale D, Gottardi P (2011) Bankruptcy, finance constraints, and the value of the firm. Am Econ J Microecon 3:1–37

Longhofer SD (1997) Absolute priority rule violations, credit rationing, and efficiency. J Financ Intermed 6:249–267

Warren E (1991) A theory of absolute priority. N Y Univers Ann Surv Am Law 9:9–48

Weiss LA (1990) Bankruptcy resolution: direct costs and violation of priority of claims. J Financ Econ 27:285–314

Abuse of Dominance

Pinar Akman

School of Law, University of Leeds, Leeds, UK

Abstract

Article 102 of the Treaty on the Functioning of the European Union (TFEU) prohibits the abuse of a dominant position within the internal market or in a substantial part of it, as incompatible with the internal market, insofar as it may affect trade between Member States. This essay examines the constituent elements of the prohibition found in Article 102 TFEU. These elements are: (i) one or more undertakings of a dominant position within the internal market or in a substantial part of it; (ii) effect on trade between Member States; (iii) abuse.

Definition

Article 102 of the Treaty on the Functioning of the European Union (TFEU) prohibits any abuse by one or more undertakings of a dominant position

within the internal market or in a substantial part of it, as incompatible with the internal market, insofar as it may affect trade between Member States. Article 102 TFEU also provides examples of abusive practices. This entry examines the constituent elements of the prohibition found in Article 102 TFEU.

Introduction

Article 102 of the Treaty on the Functioning of the European Union (TFEU) prohibits the abuse of a dominant position within the internal market or in a substantial part of it, as incompatible with the internal market, insofar as it may affect trade between Member States. This entry examines the constituent elements of the prohibition found in Article 102 TFEU. These elements are (i) one or more undertakings of a dominant position within the internal market or in a substantial part of it, (ii) effect on trade between Member States, and (iii) abuse.

One or More Undertakings of a Dominant Position Within the Internal Market or in a Substantial Part of It

The Concept of an "Undertaking"

First and foremost, it is clear from Article 102 TFEU that a dominant position can only be held by "one or more undertakings" for the purposes of that provision. An "undertaking" for the purposes of EU competition law is "every entity engaged in an economic activity regardless of the legal status of the entity and the way in which it is financed" (Case C-41/90 *Hofner and Elser v Macrotron GmbH* [1991] ECR I-1979 [21]). In turn, "economic activity" is defined as "any activity consisting in offering goods or services on a given market" (Cases C-180/98 etc *Pavel Pavlov and others v Stichting Pensioenfonds Medische Specialisten* [2000] ECR I-6451 [75]). A "functional approach" is adopted when determining whether an entity is an "undertaking" for the purposes of competition law, meaning that the same legal entity may be acting as an undertaking when carrying on one activity but not when carrying on another activity (Whish and Bailey 2012, 84–85). Similarly, the fact that the entity in question does not have a profit motive or economic purpose is irrelevant in establishing whether it is engaged in "economic activity" (Case C-67/96 etc *Albany International BV v SBT* [1999] ECR I-5751 [85]; Case 155/73 *Italy v Sacchi* [1974] ECR 409 [13–14]). Activities that are *not* economic are those provided on the basis of "solidarity" and those that consist in the exercise of public power and procurement pursuant to a noneconomic activity (Whish and Bailey 2012, 87 et seq). "Solidarity" in turn is defined as "the inherently uncommercial act of involuntary subsidisation of one social group by another" (Opinion of Advocate General Fennelly in Case C-70/95 *Sodemare v Regione Lombardia* [1997] ECR I-3395 [29]). The case law on this has mostly been concerned with national health services, compulsory insurance schemes, pension schemes, etc., as well as provision of services by public authorities which fulfill essential functions of the State.

Dominant Position

Dominant position is not defined in the Treaty. As is clear from Article 102 TFEU, it can be held by one or more undertakings. Where the dominant position is held by more than one undertaking, they would have a position of "collective dominance." The first step in establishing whether an undertaking enjoys a dominant position for the purposes of Article 102 TFEU is to define the "relevant market." This is because a dominant position cannot be held in the abstract but can only be held over a relevant market.

Market Definition

Defining the relevant market delineates the products or services that are in competition and enables one to gauge how much power an undertaking has over its competitors and consumers (Rodger and MacCulloch 2015, 95). For the purposes of competition law, the relevant market has to be defined with regard to product, geography, and occasionally time. According to the Court of Justice of the European Union (CJEU), the

definition of the market is one of establishing interchangeability: if products or services are seen as interchangeable, then they are part of the same market (e.g., Case 6/72 *Europemballage Corporation and Continental Can Company Inc v EC Commission* [1973] ECR 215 [32]). The European Commission has published a Notice on the definition of the relevant market ([1997] OJ C 372/5). According to the Notice, the main purpose of market definition is to identify in a systematic way the competitive constraints that the relevant undertakings face (Notice [2]). The objective of defining a market in both its product and geographic dimension is to identify those actual competitors of the undertaking involved that are capable of constraining that undertaking's behavior and of preventing it from behaving independently of effective competitive pressure (Notice [2]). Consequently, market definition makes it possible inter alia to calculate market shares that would convey meaningful information regarding market power for the purposes of assessing dominance (Notice [2]).

There are three main sources of competitive constraints on an undertaking's conduct: demand substitutability, supply substitutability, and potential competition (Notice [13]). In the Notice, regarding demand substitutability, the Commission adopts the so-called "hypothetical monopolist" test. This test asks whether a hypothetical small but significant non-transitory increase in price (SSNIP) of the product produced by the undertaking under investigation would lead to customers switching to other products (Notice [17]). The range of such an increase is usually 5–10%. If customers would switch to other products following such an increase, then the product of the undertaking under investigation and those other products to which the customers would switch are considered to be in the same product market (Notice [18]). Supply substitutability seeks to establish if other suppliers would switch their production to the products of the undertaking under investigation and market them in the short term without incurring significant additional costs, if there was a small and permanent increase in the price of the product under investigation (Notice [20]). However, the Notice stipulates that the Commission would take into account supply substitutability only where its effects are equivalent to those of demand substitution in terms of effectiveness and immediacy (Notice [20]). Other than the relevant product market, the relevant geographical market has to be also defined. According to the CJEU, the geographical market is limited to an area where the objective conditions of competition applying to the product in question are sufficiently homogenous for all traders (Case 27/76 *United Brands Continentaal BV v EC Commission* [1978] ECR 207 [11]). In certain cases, it may be necessary to define the temporal market as well since competitive conditions may change depending on season, weather, time of the year, etc. It should be noted that the narrower that the market is defined, the more likely that the undertaking under investigation will be found to be dominant.

Establishing Dominance

Once the relevant market is defined, then it can be established whether a given undertaking is dominant. The CJEU has defined dominant position as "a position of economic strength enjoyed by an undertaking which enables it to prevent effective competition being maintained on the relevant market by affording it the power to behave to an appreciable extent independently of its competitors, customers and ultimately of its consumers" (*United Brands* [65]). Such a position does not exclude some competition – which it does where there is a monopoly or quasi-monopoly – but enables the undertaking which profits by the position, if not to determine, at least to have an appreciable influence on the conditions under which that competition will develop and in any case to act largely in disregard of it so long as such conduct does not operate to its detriment (Case 85/76 *Hoffmann-La Roche & Co AG v EC Commission* [1979] ECR 461 [39]). Although the definitions from case law appear to relate to only undertakings on the supply side, clearly, an undertaking on the buying side can also be a dominant undertaking for the purposes of competition law, as was the case in, for example, *British Airways* (see Case C-95/09 P *British Airways v EC Commission* [2007] ECR I-2331).

According to the European Commission, dominance entails that competitive constrains are ineffective, and hence, the undertaking in question enjoys substantial market power over a period of time (Guidance on the Commission's enforcement priorities in applying Article 82 of the EC Treaty to abusive exclusionary conduct by dominant undertakings [2009] OJ C45/7 [20]). An undertaking which is capable of profitably increasing prices above the competitive level for a significant period of time does not face sufficiently effective competitive constraints and can thus generally be regarded as dominant (Guidance [11]). The assessment of dominance will take into account the competitive structure of the market and in particular factors, such as (i) constraints imposed by existing supplies from, and the position on the market of, actual competitors, (ii) constraints imposed by the credible threat of future expansion by actual competitors or entry by potential competitors, and (iii) constraints imposed by the bargaining strength of the undertaking's customers (countervailing buyer power) (Guidance [12–18]). Regarding the first of these factors, market shares provide a useful first indication of the market structure and of the relative importance of various undertakings on the market (Guidance [13]). However, market shares will be interpreted in light of the relevant market conditions and, in particular, of the dynamics of the market and of the extent to which products are differentiated (Guidance [13]). According to the CJEU, "although the importance of the market share vary from one market to another the view may legitimately be taken that very large shares are in themselves, and save in exceptional circumstances, evidence of the existence of a dominant position" (Case 85/76 *Hoffmann-La Roche & Co v EC Commission* [1979] ECR 461 [41]). Interestingly, in *AKZO* the CJEU held that a market share of 50% could be considered to be very large, and in the absence of exceptional circumstances, an undertaking with such a market share would indeed be presumed to be dominant (Case C-62/85 *AKZO Chemie BV v Commission* [1991] ECR I-3359 [60]). Consequently, an undertaking with 50% of market share would have to rebut this presumption to prove that it is *not* dominant. Interestingly, in its Guidance, instead of referring to a presumption of dominance, the Commission refers to the fact that low market shares are generally a good proxy for the absence of substantial market power: dominance is not likely if the undertaking's market share is below 40% in the relevant market (Guidance [14]). It should be noted that the Guidance is not intended to constitute a statement of the law and is without prejudice to the interpretation of Article 102 TFEU by the CJEU or the General Court; the Guidance sets the enforcement priorities of the European Commission in applying Article 102 TFEU to certain types of abusive practices (see Akman 2010 on the Guidance and its legal position as a soft law instrument). It is noteworthy that the lowest market share that an undertaking has been held to be dominant by the Commission and confirmed on appeal by the General Court is 39.7% in *British Airways*. In any case, the Commission also acknowledges in the Guidance that there may be specific cases below the threshold of 40% market share where competitors are not in a position to constrain effectively the conduct of a dominant undertaking, for example, where they face serious capacity limitations, and such cases may also deserve the attention of the Commission (Guidance [14]).

Other than market shares, potential competition is also important for establishing dominance. Assessing potential competition entails identifying the potential threat of expansion by actual competitors, as well as potential entry by other undertakings into the relevant market. These are relevant factors since competition is a dynamic process and an assessment of the competitive constraints on an undertaking cannot be based merely on the existing market situation (Guidance [16]). An undertaking can be deterred from increasing prices if expansion or entry is likely, timely, and sufficient (Guidance [16]). According to the Commission, in this context, "likely" refers to expansion or entry being sufficiently profitable for the competitor or entrant, taking into account factors such as barriers to expansion or entry, the likely reactions of the allegedly dominant undertaking and other competitors, and the risks and costs of failure

(Guidance [16]). For expansion or entry to be considered "timely," it must be sufficiently swift to deter or defeat the exercise of substantial market power (Guidance [16]). Finally, to be "sufficient," expansion or entry has to be more than small-scale entry and be of such a magnitude to be able to deter any attempt to increase prices by the allegedly dominant undertaking (Guidance [16]).

The issue of potential competition brings to the fore the discussion of "barriers to entry or expansion." There is considerable debate over what should be included within the term "barriers to entry" (Rodger and MacCulloch 2015, 101): one school of thought (Chicago) would only accept as a barrier to entry a cost to new entrants which was not applicable to the existing operators when they entered the market, whereas another school of thought (including the European Commission) views barriers to entry to be much wider, including any factor that would tend to discourage new entrants from entering the market. According to the Commission, barriers to expansion or entry can take various forms, such as legal barriers, tariffs or quotas, and advantages enjoyed by the allegedly dominant undertaking (such as economies of scale and scope, privileged access to essential input or natural resources, important technologies, or an established distribution and sales network) (Guidance [17]). Barriers to expansion or entry can also include costs and other impediments faced by customers in switching to a different supplier (Guidance [17]). Furthermore, the allegedly dominant undertaking's own conduct may also create barriers to entry, for example, where it has made significant investments which entrants or competitors would have to match or where it has concluded long-term contracts with its customers that have appreciable foreclosing effects (Guidance [17]). Considering the conduct of the undertaking to be an entry barrier is clearly controversial since conduct is normally taken into account when assessing the "abuse" element of the provision rather than the element of dominance. Such an approach is circular in that conduct will not normally be considered abusive until dominance is established, but if conduct can also indicate dominance, then the likelihood of a finding of abuse clearly increases (Rodger and MacCulloch 2015, 102).

The final factor to be taken into account in establishing dominance is countervailing power held by the trading partners of the allegedly dominant undertaking. Competitive constraints may be exerted not only by actual or potential competitors but also by customers (or suppliers, if the dominant undertaking is on the buying side) of the allegedly dominant undertaking. According to the Commission, even an undertaking with a high market share may not be able to act to an appreciable extent independently of customers (or suppliers, if the dominant undertaking is on the buying side) with sufficient bargaining strength (Guidance [18]). Such bargaining power may result from the trading partner's size, commercial significance, ability to switch, ability to vertically integrate, etc. (Guidance [18]).

Dominant Position "Within the Internal Market or in a Substantial Part of It"

According to Article 102 TFEU, to be subject to the prohibition therein, the relevant undertaking has to have a dominant position "within the internal market or in a substantial part" of the internal market. This has been noted to be the equivalent of the de minimis doctrine under Article 101 TFEU, according to which agreements of minor importance are not caught by the prohibition found in Article 101 TFEU (Whish and Bailey 2012, 189). Clearly, where dominance is established to exist throughout the EU, there is no difficulty in deciding that the dominant position is held within the internal market or in a substantial part of the internal market. Where dominance is more localized than this, then it will have to be decided what "substantial" refers to. Substantiality does not simply refer to the physical size of the geographic market within the EU (Whish and Bailey 2012, 190). Rather, what matters is the relevance of the market in terms of volume and economic opportunities of sellers and buyers (Cases 40/73 etc *Suiker Unie and others v EC Commission* [1975] ECR 1663 [371]). Each Member State is likely to be considered to be a substantial part of the internal market, as well as parts of a Member State (Whish and Bailey 2012, 190).

Effect on Trade Between Member States

The prohibition in Article 102 TFEU is only applicable to the extent that the conduct of the dominant undertaking "may affect trade between Member States." This is a *jurisdictional* criterion that establishes whether EU competition law is applicable, as well as demarcating the application of EU competition law from national competition laws of the Member States. The Commission has published Guidelines on the effect on trade concept contained in Articles 101 and 102 TFEU ([2004] OJ C 101/81). First and foremost, Articles 101 and 102 TFEU are only applicable where the effect on trade between Member States is *appreciable* (Guidelines [13]). Second, the concept of "trade" is not limited to exchange of goods/services but covers all cross-border economic activity, including the establishment of agencies, branches, subsidiaries, etc. in Member States (Guidelines [19, 30]). The concept of "trade" also includes cases where conduct results in a change in the *structure* of competition on the internal market (Cases 6 and 7/73 *Commercial Solvents v EC Commission* [1974] ECR 223 [33]; Guidelines [20]).

Regarding the notion of "may affect" trade between Member States, the CJEU has noted that this means that it must be possible to foresee, with a sufficient degree of probability on the basis of a set of objective factors of law or of fact, that the conduct may have an influence, direct or indirect, actual or potential, on pattern of trade between Member States (Guidelines [23]). It should be noted that this is a neutral test; it is not a condition that trade be restricted or reduced (O'Donoghue and Padilla 2013, 864).

Abuse

According to the CJEU, a finding of a dominant position is not in itself a recrimination, but means that irrespective of the reasons for which it has such a position, the dominant undertaking has a "special responsibility" not to allow its conduct to impair genuine undistorted competition on the internal market (Case 322/81 *Michelin v EC Commission* [1983] ECR 3461 [57]). Although the Court has created this concept of "special responsibility" for dominant undertakings, what exactly it entails – and if it entails anything above the parameters of the prohibition in Article 102 TFEU itself – is debatable. It has been suggested in the literature that dominant undertakings do not have any responsibility over and above complying with Article 102 TFEU itself (Akman 2012, 95).

Article 102 TFEU prohibits the abuse of a dominant position and lists examples of abuse in a non-exhaustive manner (*Continental Can* [26]). In general, it is considered that abusive conduct can be categorized as: (i) exploitative and (ii) exclusionary. Exploitative abuse relates to the dominant undertaking directly harming its customers (including consumers) as a result of, for example, the reduction in output and the increase in prices that the dominant undertaking can effect due to its market power. It has been defined as the dominant undertaking receiving advantages to the disadvantage of its customers that would not be possible *but for* its dominance (Akman 2012, 95, 303). In contrast, exclusionary abuse concerns the dominant undertaking's conduct that harms the competitive position of its competitors, mainly by foreclosing the market. Most of the decisional practice under Article 102 TFEU has concerned exclusionary conduct, despite the fact that the examples listed in Article 102 TFUE are mostly – if not only – concerned with exploitative abuse. Indeed, it has been argued in the literature that Article 102 TFEU itself is merely concerned with exploitation and not exclusion (see Joliet 1970; Akman 2009, 2012). In line with the decisional practice, the Commission's Guidance on enforcement priorities – the only official document on the application of Article 102 TFEU adopted by the Commission – is limited to exclusionary conduct. It must be noted that the Guidance was published at the end of a long period of debate on the role and application of Article 102 TFEU as the Commission's enforcement of this provision, as well as the European Courts' jurisprudence thereon had been criticized by many commentators for not being based on economic effects but on the form of conduct, for

failing to fall in line with the Commission's more modern approach to Article 101 TFEU and merger control, and "for protecting competitors, not competition" (see, e.g., O'Donoghue and Padilla 2013, 67 et seq. for an overview of the reform).

Exploitative Abuses

Unfair Pricing and Unfair Trading Conditions

Article 102(a) TFEU prohibits the imposition of unfair prices or unfair trading conditions. Although the prohibition is one of "unfair" pricing, it has been mostly interpreted as one of excessive pricing. In any case, there have been very few cases prohibiting prices as "excessive" or "unfair" since the prohibition poses many problems, such as the difficulty of defining what an "excessive" or "unfair" price is, the potential adverse effects on innovation and investment the prohibition could lead to, the lack of legal certainty resulting from a lack of a test for "excessive" or "unfair" prices, the inappropriateness of price regulation by competition authorities and courts, etc. As for the interpretation of "unfair pricing" by the CJEU, there is a two-staged test established in *United Brands*: first, it should be determined whether the price-cost margin is excessive, and if so, it should be determined whether a price has been imposed that is either "unfair in itself" or "when compared to competing products."

Regarding the abuse of imposing unfair trading conditions, there is similarly limited case law. Examples of unfair trading conditions include imposing obligations on trading partners which are not absolutely necessary and which encroach on the partners' freedom to exercise its rights, imposing commercial terms that fail to comply with the principle of proportionality, unilateral fixing of contractual terms by the dominant undertaking, etc. (Case 127/73 *Belgische Radio on Televisie v SV SABAM* [1974] ECR 313; *DSD* (Case COMP D3/34493) [2001] OJ L166/1; Case 247/86 *Alsatel v SA Novasam* [1988] ECR 5987 respectively).

Exclusionary Abuses

According to the Guidance paper, the Commission's enforcement activity in relation to exclusionary conduct aims to ensure that dominant undertakings do not impair effective competition by foreclosing their competitors in an anticompetitive way, thus having an adverse impact on consumer welfare, whether in the form of higher prices than would have otherwise prevailed or in some other form such as limiting quality or reducing consumer choice (Guidance [19]). In turn, "anticompetitive foreclosure" refers to a situation where effective access of actual or potential competitors to supplies or markets is hampered or eliminated as a result of the conduct of the dominant undertaking whereby the dominant undertaking is likely to be in a position to profitably increase prices (or to influence other parameters of competition, such as output, innovation, quality, variety, etc.) to the detriment of consumers (Guidance [19]). It must be noted that the Guidance has not received formal approval from the EU Courts as they have not yet had to deal with a case in which the Commission applied the principles of the Guidance and the EU Courts have not necessarily been keen to revise their case law in order to adopt a more economic effects-based approach. The rest of this section will consider some common types of exclusionary conduct that have been identified as priorities in the Commission's Guidance paper. It should be noted that the Guidance paper distinguishes between price-based and non-price-based exclusionary conduct. For price-based exclusionary conduct, the test promoted in the Guidance is that of the "as efficient competitor" test, according to which, the Commission will only intervene where the conduct concerned has already been or is capable of hampering competition from competitors which are considered to be as efficient as the dominant undertaking (Guidance [23]). This test was used in some earlier case law already, but it is noteworthy that in the recent appeal of *Intel*, the General Court has found the test to be a neither necessary nor sufficient test of abuse, at least for the particular conduct in question, namely, rebates (Case T-286/09 *Intel Corp v European Commission* (not yet published) [143–146]).

Exclusive Dealing

A dominant undertaking may foreclose its competitors by hindering them from selling to

customers through use of exclusive purchasing obligations or rebates, which the Commission together refers to as "exclusive dealing" (Guidance [32] et seq.). "Exclusive purchasing" refers to an obligation imposed by the dominant undertaking on a customer to purchase exclusively or to a large extent only from the dominant undertaking (Guidance [33]). According to the CJEU, it is irrelevant whether the request to deal exclusively comes from the customer, and it is also irrelevant whether the exclusivity obligation is stipulated without further qualification or undertaken in return for a rebate (*Hoffmann-La Roche* [89]).

Regarding rebates granted to customers to reward them for a particular form of purchasing behavior, particularly the European Courts have adopted a very formalistic approach. Specifically, rebates that create "loyalty" to the dominant undertaking are condemned to a degree which some might argue to be a per se prohibition. For example, recently in *Intel* the General Court held that fidelity/exclusivity rebates are abusive if there is no justification for granting them, and the Commission does not have to analyze the circumstances of the case to establish a potential foreclosure effect (*Intel* [80–81]). Quantity rebates, namely, rebates which are linked solely to the volume of purchases from the dominant undertaking which reflect the cost savings of supplying at higher levels, are generally deemed not to have foreclosure effects (*Intel* [75]).

Tying and Bundling

Tying refers to situations where customers that purchase one product (the "tying product") are required to also purchase another product from the dominant undertaking (the "tied product") and can be contractual or technical (Guidance [48]). Bundling refers to the ways the dominant undertaking offers and prices its products: pure bundling occurs where products are *only* sold together in fixed proportions, whereas mixed bundling occurs where products are available separately, but the price of the bundle is lower than the total price when they are sold separately (Guidance [48]). As well as having potential efficiency benefits for customers, tying and bundling can also be used by the dominant undertaking to foreclose the market for the tied product by leveraging its market power in the tying product market to the tied product market. For example, in the case of *Microsoft*, the tying of Microsoft Media Player to the Windows Operating System was found to be an abuse of Microsoft's dominant position (Case T-201/04 *Microsoft Corp v EC Commission* [2007] ECR II-3601).

Predatory Pricing

Predation involves the dominant undertaking selling its products at a price below cost. As such, the dominant undertaking deliberately incurs losses or foregoes profits in the short term, so as to foreclose or be likely to foreclose one or more of its actual or potential competitors, with a view to strengthening or maintaining its market power (Guidance [63]). The obvious difficulty with sanctioning predation is that an erroneous condemnation would imply the prohibition of low prices, and price competition leading to low prices is clearly part of legitimate competition. In *AKZO* the CJEU decided that prices below average variable costs (costs that vary according to the quantity produced) by means of which a dominant undertaking seeks to eliminate a competitor must be regarded as abusive since a dominant undertaking has no interest in applying such prices except that of eliminating competitors so as to enable it subsequently to raise its prices by taking advantage of its monopolistic position, since each sale generates a loss (*AKZO* [71]). As for prices between average variable costs and average total costs (variable costs plus fixed costs), the Court held that such prices will be held abusive if they are part of a plan for eliminating a competitor (*AKZO* [72]). In its Guidance, the Commission uses average avoidable cost instead of average variable cost ([64] et seq.). It is not necessary to prove that the dominant undertaking can possibly recoup its losses for predation to be abusive (see, e.g., Case C-202/07 P *France Télékom v Commission* [2009] ECR I-2369 [110]).

Refusal to Supply and Margin Squeeze

Despite the fact that generally any undertaking, dominant or not, should have the right to choose

its trading partners and to dispose freely of its property, there are occasions on which a dominant undertaking can abuse its position by refusing to deal with a certain trading partner (Guidance [75]). Such a finding entails the imposition of an obligation to supply on the dominant undertaking, and the Commission acknowledges that such obligations may undermine undertakings' incentives to invest and innovate and, thereby, possibly harm consumers (Guidance [75]). According to the Commission, typically competition problems arise when the dominant undertaking competes on the downstream market with the buyer whom it refuses to supply (Guidance [76]). The downstream market refers to the market for which the refused input is needed in order to manufacture a product or provide a service (Guidance [76]). It is irrelevant whether the customer to whom supply is refused is an existing customer or a new customer, but it is more likely that the termination of an existing relationship will be found to be abusive than a de novo refusal to supply (Guidance [79], [84]). Instead of refusing to supply, a dominant undertaking may engage in "margin squeeze": it may charge a price for the product on the upstream market which, compared to the price it charges on the downstream market, does not allow an equally efficient competitor to trade profitably in the downstream market on a lasting basis (Guidance [80]).

For refusal to supply and margin squeeze to constitute abuse, it must be the case that the refusal relates to a product or service that is objectively necessary ("indispensable") to be able to compete effectively on the downstream market, is likely to lead to elimination of effective competition on the downstream market, and is not objectively justified (Whish and Bailey 2012, 699). A particularly controversial application of this doctrine is the area of intellectual rights where a finding of abuse implies that these rights may be made subject to compulsory licensing. For example, in *Microsoft*, abuse was found in Microsoft's refusal to provide interoperability information to its competitors which would enable them to develop and distribute products that would compete with Microsoft on the market for servers. According to the General Court, Microsoft's conduct limited technical development under Article 102(b) TFEU (*Microsoft* [647]).

Objective Justification

Article 102 TFEU does not contain an exemption or exception clause like that found in Article 101(3) TFEU, which would "save" otherwise abusive conduct from breaching Article 102 TFEU due to any procompetitive gains the conduct might produce. However, the decisional practice and the case law have developed the concept of "objective justification" as a defense on the part of the dominant undertaking. The dominant undertaking may demonstrate that its conduct is objectively necessary or that the anticompetitive effect produced by the conduct is counterbalanced or outweighed by advantages in terms of efficiencies that also benefit consumers (Case C-209/10 *Post Danmark A/S v Konkurrenceradet*, not yet reported [41]; Guidance [28–31]). According to some commentators, the defense of objective justification is somewhat a tautology (O'Donoghue and Padilla 2013, 283). This is because, as noted by Advocate General Jacobs in *Syfait*, the very fact that conduct is characterized as abusive suggests that a negative conclusion has already been reached, and therefore, a more accurate conception would be to accept that certain types of conduct do not fall within the category of abuse at all (Case C-53/03 *Syfait and others v GlaxoSmithKline plc and another* [2005] ECR I-4609 [53]).

References

Akman P (2009) Searching for the long-lost soul of Article 82EC. Oxf J Leg Stud 29(2):267–303

Akman P (2010) The European commission's guidance on Article 102TFEU: from *Inferno* to *Paradiso*? Mod Law Rev 73(4):605–630

Akman P (2012) The concept of abuse in EU competition law: law and economic approaches. Hart Publishing, Oxford

Joliet R (1970) Monopolization and abuse of dominant position. Martinus Nijhoff, La Haye

O'Donoghue R, Padilla J (2013) The law and economics of Article 102 TFEU, 2nd edn. Hart Publishing, Oxford

Rodger BJ, MacCulloch A (2015) Competition law and policy in the EU and UK, 5th edn. Routledge, Oxford

Whish R, Bailey D (2012) Competition law, 7th edn. OUP, Oxford

Access to Justice

David Allen Larson
Hamline University School of Law, Saint Paul, MN, USA

Abstract

Access to justice has both procedural and substantive components. Context matters and a society that views its members as part of a collective, for example, may perceive access to justice differently than a more individualistic society. International law documents ensuring access to justice generally take either a general human rights approach or provide specific protections for disadvantaged populations. Although substantive access to justice appears to have improved over time, procedural access to justice may not have kept pace. Money and time are very real limitations. Physical barriers have a severe impact on persons with disabilities and individuals living in poverty. Institutional barriers also limit access to justice for reasons that include ponderous or bias court systems, discriminatory police conduct, expense, and political interference. Additionally, limited education and social custom impair access to justice. When public trust is lacking, individuals may not rely on justice institutions to settle disputes and resolve problems. Challenges remain concerning which substantive rights we need to protect and what efficient and effective procedures are available.

Definition

Access to justice, frequently abbreviated ATJ or A2J, refers to two different, but closely related, concerns. On the one hand, procedural access to justice focuses on both the processes that are available to help people enforce their rights and privileges under the law and the effectiveness of those processes. When one explores procedural access to justice, for example, one might examine access to the physical locations of justice administration (such as courthouses and police stations), individuals' ability to understand and participate in proceedings (such as court hearings, police interviews, and conversations with one's attorney), and due process of law. On the other hand, the notion of substantive access to justice focuses on the nature and extent of the rights to be protected. Questions about substantive access to justice can focus on specific rights or address broad questions such as whether a society ensures equitably equal access to opportunities and benefits and whether all individuals have the ability to live a just life. It is difficult to imagine a just life that does not require the existence and enforceability of substantive rights, such as the right to travel freely or the right to freedom of conscience. But any discussion of access to justice must acknowledge that context matters. A society or culture that views its members as part of a collective may have a different understanding of substantive access to justice than a society that is more individualistic.

Access to Justice

Procedural access to justice is central to the enforcement of legal rights and privileges (Ortoleva 2011). When individuals and groups are excluded from society, talents and value are lost. Improving procedural access to justice may be particularly helpful for groups that traditionally and historically have been subject to discrimination (Ortoleva 2011). When a law or practice prevents or discourages individuals from participating in the mechanisms for enforcing laws (which include policing, civil and criminal court claims, alternative dispute resolution processes, appeals, and final judgments), it denies those individuals procedural access to justice (Ortoleva 2011).

The phrase access to justice also may be used when one focuses on the question of whether individuals truly are experiencing a just existence. Substantive access to justice is generally the goal, if not always the result, of procedural access to justice. Procedural access to justice provides a route to substantive access to justice. But even

the most carefully designed procedures may provide little benefit in the absence of recognized or cognizable substantive rights.

Many international law documents, as well as the constitutions and laws of most nations, ensure the right of people to access the courts. International law documents ensuring access to justice can be divided into two types: (1) those that take a general human rights approach and (2) those that make specific provisions for populations that historically have lacked access to justice for one reason or another. As promising as they sound, however, it may be difficult to enforce the rights described in these documents.

Documents belonging to the first category include the United Nations (UN) Universal Declaration of Human Rights and the Hague Convention on International Access to Justice, along with various regional declarations and conventions, such as the American Declaration of the Rights and Duties of Man and the European Convention for the Protection of Human Rights and Fundamental Freedoms. These documents are primarily aspirational, though some (such as the Hague Conventions and the International Covenant on Civil and Political Rights) potentially are binding on states. Aspirational documents tend to contain more substantive provisions than procedural ones, but they do address both types of access to justice.

Many general human rights documents express the same rights and freedoms as appear in the Universal Declaration of Human Rights. Rights believed important for access to justice include the right to recognition as a person before the law; the right to equal protection of the laws; the right to an effective, enforceable remedy for a violation of one's rights; and the right to a full, fair, and public hearing by an independent and neutral tribunal. Although the line is not always bright, these primarily procedural rights are central to the enforcement of substantive rights, such as the rights of freedom of thought, conscience, and religion; freedom to travel; and freedom from arbitrary arrest, detention, and exile.

The second category, consisting of population-focused documents, includes many UN conventions, such as the Convention to Eliminate All Forms of Discrimination Against Women, the UN Convention for the Elimination of Racial Discrimination, and the UN Convention on the Rights of Persons with Disabilities. These documents reaffirm the principles of general human rights documents but also are sensitive to the special needs and circumstances of the populations they protect. For example, the Convention on the Rights of Persons with Disabilities contains many of the same rights as general human rights documents. But it also provides, for example, that states should ensure that persons with disabilities have equal access to information and that those working in fields of justice administration receive appropriate training to advance and protect the rights of persons with disabilities.

International bodies within the United Nations, such as the Human Rights Committee, also bear some of the responsibility for monitoring the impact of the international community's efforts to improve access to justice. The United Nations Development Programme, the primary UN body responsible for the UN's Millennium Development Goals, provides support to legal aid providers in several countries, with particular emphasis on providing services and education to the poor and other marginalized groups.

Many national constitutions secure rights that are necessary to access to justice. Such rights include the right to counsel, the right to equal treatment by the government, and the right to seek redress for wrongs through the courts. Although these rights may mirror those contained in general human rights documents, they can be more easily enforced because they are the laws of the state rather than aspirations of the international community.

In addition to including provisions in their constitutions, nations also promote access to justice through legislation. Countries may enact legislation to fulfill obligations under international conventions but also may be responding to the need in their own country for heightened protection for identifiable populations. For example, one can find legislation in many countries that promises protection for individuals based on their race, religion, gender, health, or country of origin. Antidiscrimination legislation at the national level, public educational initiatives, and programs

directed toward empowering and informing marginalized populations of their rights can improve access to justice by eliminating gaps in political, economic, and social power. Additionally, some nations have chosen not only to prohibit discrimination against marginalized populations but also encourage policy makers to continue to review court decisions and other governmental actions that may have a disproportionate effect on those identifiable populations. These efforts are most effective when they are a result of community participation and consensus.

When a person or group is denied access to justice, that denial may be either direct or indirect. Access to justice is denied in a direct manner when a specific person or group is explicitly prevented from attaining procedural or substantive justice. Historical examples include restrictions on personhood based on race and the shedding of a woman's legal personality upon marriage (coverture). Indirect denial of access to justice, in contrast, consists of restrictions that appear neutral but have a disparate impact on a specific group. This type of discrimination includes filing fees that may be affordable for most people but prevent the poor from accessing the courts; height, weight, or strength requirements that exclude qualified women from a job even though those qualities are not necessary to perform the actual duties of the job; or diploma, certificate, or skill requirements that again may not actually be needed to perform a well-paying job. Employment can improve access to justice by providing the resources (time, money, knowledge) necessary for a person to access the justice system.

Access to Justice in Historical Context

Before the modern era of human rights, the common view was that only states could invoke the protections of international law (Francioni 2007). Individuals had limited access to justice. They could seek redress only in the state where the wrong occurred, which meant that their rights and privileges were defined by that state's own laws (Francioni 2007). When individuals sought redress in foreign states that may have had more favorable laws, they often faced local prejudices, language barriers, wide variances in substantive law, and other challenges that hindered access to justice (Francioni 2007).

In order for a person to enforce his or her rights in the international realm, a state would need to intercede on the person's behalf (Francioni 2007). Claims made by a state on behalf of one of its citizens were not considered individual claims for remedy, but rather diplomatic issues to be resolved between the states themselves (Francioni 2007). These claims often were addressed in ways that did not involve judicial mechanisms and which were not transparent (Francioni 2007). Even today, some forums are only available to states, such as the International Court of Justice.

Eventually, multiple international forums for enforcing individual rights came into existence (Francioni 2007). Regional and nongovernmental organizations have created forums for the vindication of individual rights, such as the European Court of Justice and the Inter-American Court of Human Rights. In addition to these international bodies, individuals today retain their historical access to domestic courts (Francioni 2007). Agreements between states which ensure that individuals have access to justice often include the individual's right to access the courts of a state regardless of nationality. While procedural and substantive barriers may still arise, discrimination based solely on nationality is generally forbidden, along with discrimination based on characteristics such as race, gender, and religion.

Access to justice has improved in some respects while lagging in others. In most societies, substantive access to justice appears to have been strengthened. When one examines modern constitutional and legislative language, it is not uncommon to find language guaranteeing access to justice for all persons. Bold and impressive language often promises access to justice for all persons regardless of factors such as age, race, and sex. Populations with other shared characteristics, such as persons with disabilities and those with HIV/AIDS, also have better access to justice as countries work to address histories of stigmatization. But even though substantive access to

justice may have improved, procedural access to justice may not have kept pace. Court systems and administrative agencies often are terribly backlogged, which results in significant delays and in some cases a denial of access to justice. Money and time can be very real limitations affecting access to justice in modern societies. The availability of alternative forums for justice, such as restorative justice processes, mediation, and arbitration, may increase access to justice, but the potential for unfair outcomes still looms when these processes are not held to the same standards of fairness, equality, and transparency as court systems.

Although research generally supports the commonly held assumption that early referral to mediation increases the chances that parties will reduce their expenses, it does not support the claim that mediation by itself reduces party costs (McEwen 2014). Analyses of civil mediation programs do not find any consistent differences in attorneys' fees, hours, or other costs between mediated and other cases (McEwen 2014). Litigation costs and litigation activity (depositions, interrogatories, and motions) do not automatically decline whenever parties choose mediation (McEwen 2014). Research focusing on business-to-business disputes does suggest, however, that if parties engage in mediation early in the litigation process, then it can significantly reduce costs (McEwen 2014). Cost savings appear most likely when mediation (and case management) alters normal litigation practices, particularly by reducing the amount of discovery and motions and by initiating serious settlement efforts early in the case (McEwen 2014).

Physical Access to Justice

One of the most essential aspects of access to justice is the ability to reach the locations where justice is administered. Physical barriers such as distance can have a severe impact on persons with disabilities and the poor (Ortoleva 2011; Carmona 2012). Distance makes it more difficult for victims to report crimes, for police to respond to reports of crime, for those seeking justice to secure legal representation, and for persons who are disabled or poor to do something as seemingly simple as physically appear at court proceedings (Ortoleva 2011; Carmona 2012). For the poor, securing transportation to the locations where justice is carried out can be prohibitively expensive (Carmona 2012). Even when transportation is affordable, lengthy travel to physical locations often means losing valuable time at work or at home (Carmona 2012). The relative wealth or poverty of a country obviously contributes to these difficulties, particularly in those parts of the world where transportation and communications infrastructures may be underdeveloped.

Certain disabilities can make physical access to justice especially difficult. In areas without comprehensive legal requirements for building accessibility, people who cannot walk are either unable to enter courthouses and police stations or find it very difficult or humiliating (Ortoleva 2011). Those who are blind or deaf may be physically present, but often do not enjoy the full benefit of that presence without interpreters or Braille materials (Ortoleva 2011).

Institutional Access to Justice

The enforcement of the rights that underlie access to justice requires institutions that work for everyone. Such institutions must provide timely, fair, and effective resolution of questions about people's rights. Institutional barriers that hinder access to justice include ponderous or biased court systems, the expense of using legal institutions, and political interference with judicial processes (Agrast 2014).

While the justice systems in the world's most developed nations promise fairness and accessibility, they do not always fulfill those promises. Despite guarantees of fairness, usually included in these countries' laws and constitutions, marginalized groups still face problems when trying to access justice that can include physical access challenges, discriminatory police conduct, or judicial bias (Agrast 2014).

The right to legal counsel, generally regarded as a valuable guarantee, in fact creates some of the tension that exists when we think about access to

justice, especially with respect to the poor. This right often applies only for those accused of a crime, for example, not to those involved in a civil damages dispute (Agrast 2014). But a civil damages lawsuit potentially may be financially devastating, and excellent legal representation may be required to avoid that devastation. That level of civil legal representation, however, can be prohibitively expensive (Bloch 2008). And even though one may be guaranteed legal counsel in criminal cases, the quality of that legal counsel may again depend upon one's financial resources (Ogletree and Sapir 2004).

The right to counsel in criminal proceedings is certainly an important consideration in defense of one's rights, but the primary mechanisms for enforcing rights that impact access to justice often are civil proceedings. Therefore, enforcing rights attendant to access to justice often is a pay-to-play proposition, and the poor may find it difficult or impossible to challenge policies and laws that restrain their access to justice when they do not have legal counsel (Ortoleva 2011). This problem is compounded in nations where the government is either unable or unwilling to provide adequate funding to support the legal needs of the poor (Agrast 2014). These problems have been partially addressed by the European Convention on Human Rights, which has been interpreted to provide a limited right to counsel in civil cases, but enforcement of that right remains problematic. Alternative routes to remedies, such as legislation, are also costly, often prohibitively so for individuals.

Class action mechanisms alleviate some of the economic strain of civil litigation, but class actions are not universally available (Watson 2001). Furthermore, the procedural barriers to initiating a class action lawsuit require even more expenses at the beginning of the lawsuit, which may prove insurmountable even if the cost is shared by the class (Watson 2001). Nonetheless, class actions can be a powerful tool once the procedural requirements are overcome. Aggregating the claims of many individuals into a single lawsuit can make viable claims for amounts that otherwise would have been less than the cost of obtaining them individually (Watson 2001).

Class actions can promote judicial economy by consolidating what otherwise would be a multitude of separate lawsuits heard by many judges into a single lawsuit (Watson 2001). In both class actions and smaller-scale litigation, however, individuals living in poverty may have their remedies limited by fee shifting provisions. So-called "loser pays" systems of fee shifting, where the party who loses the lawsuit must pay the winning party's fees, increase the potential cost of litigation (Watson 2001; Hodges 2001). While this has the commonly desired effect of discouraging frivolous or weak claims, it may also discourage strong claims by those who simply cannot afford to pay the other party's legal fees (Hodges 2001). Some jurisdictions, recognizing that simple consumer claims need not be subject to the entire trial process, have attempted to simplify the process for these small claims by eliminating the need for legal representation and removing some procedural requirements from the claims process itself. Both class actions and simplified processes may reduce the expense of pursuing a claim, making redress more accessible to those who may otherwise lack the financial means to vindicate their rights.

But cost is not the only reason why class actions may not be very helpful. There are significant procedural hurdles. In order to proceed with a class action, for example, the complaining parties must share a common claim. Change a few facts, change the alleged wrongful actor, or change the location and it may be impossible to proceed as a class. This requirement can prove especially problematic for persons claiming disability discrimination because, given the seemingly infinite range of possible disabilities, claimants may not be able to demonstrate that they are similarly situated.

At a more basic level, it is obvious that in order to enforce substantive rights, people must first be aware of those rights. This is of special concern when it comes to persons with intellectual, visual, or auditory disabilities (Ortoleva 2011). Justice institutions such as courts and police stations often lack interpreters for the deaf, materials available in a format accessible to the blind, and guardianship for those with intellectual disabilities (Ortoleva 2011). Groups who have limited

education (often the poor and women) and who are not allowed by social custom or law to access justice institutions for themselves (women and children) share similar challenges (Turquet et al. 2015). Even when there are no court proceedings in the foreseeable future, the right to counsel clearly is valuable. Counsel can inform and advise individuals regarding their rights (Ortoleva 2011). Even something as simple as programs that provide informative materials for those with certain disabilities, or that encourage public education about rights and the law, can help bridge the knowledge gap.

The public also must be able to trust that justice institutions will serve their purpose. Minority groups often believe that these institutions serve the majority at the expense of the minority or that they even act in direct opposition to the minority. The public may not only believe that procedural access to justice is compromised, they may also believe that substantive access to justice is lacking. Both claimants and respondents may have this belief. For example, after the passage of the Americans with Disabilities Act (ADA) in the United States, news coverage and public discourse revealed a concern that courts responsible for handling ADA cases awarded excessive amounts to plaintiffs at the expense of their employers (Krieger 2003). The fact is that a significant number of reasonable accommodations for persons with disabilities cost very little or nothing at all, however, and the average cost of reasonable accommodation is only in the hundreds of dollars (Krieger 2003). Public perception does not always reflect what actually occurs in justice institutions, but a lack of public trust may prevent people from relying on justice institutions to settle disputes and resolve problems. Thus even when institutional processes are available, individuals may decide that it is not worth their time or money to pursue those processes.

Although generally available, the institutions central to access to justice may be particularly vulnerable during periods of crisis, such as war, natural disaster, or civil unrest (Francioni 2007). Financial resources may be diverted to address the crisis and the institutions themselves may be so damaged or compromised that they are ineffective.

Legal Rights and Access to Justice

We still may be at a point in time where access to justice requires the existence of enforceable legal rights that protect both procedural and substantive access to justice. As explained earlier, procedural rights govern the ability to access and use justice institutions. Procedural rights include formal rules that explicitly protect one's ability to testify before a court (competency), for example, or rules that protect the ability to bring a claim or respond to a criminal charge or rules of evidence and formal procedure.

Substantive rights identify the types of rights that we hope to protect, such as rights to life, liberty, or prosperity. These rights can adapt to specific circumstances and populations. They can be broad pronouncements or more specific to fit a particular situation, such as the right to physically access courthouses that have a wide variety of architectural design. Majority groups are often the most protected, while minorities may find that they lack some of the substantive rights and privileges that the majority possesses. Women, for example, may lack the right to own or inherit property, restraints that are not usually experienced by men (Turquet et al. 2015). Some rights, such as freedom of expression, may be reserved to a privileged social or cultural group, or they may be enforced unevenly. In that latter circumstance, the rights would not be protected procedurally, resulting in de facto censorship or chilling of speech. Thus the challenge remains twofold: what are the substantive rights that we need to protect and what procedures are available that are effective and efficient? The answers to these two questions will help guide local, national, and international efforts to improve and secure access to justice.

References

Agrast MD (2014) World Justice Project Rule of Law Index 2014. Available at http://worldjusticeproject.org/sites/default/files/files/wjp_rule_of_law_index_2014_report.pdf

Bloch FS (2008) New directions in clinical legal education: access to justice and the global clinical movement. Wash Univ J Law Policy 28(1):111–139

Carmona MS (2012) Report of the Special Rapporteur on extreme poverty and human rights. U.N. Doc. No. A/HRC/23/36 (11 Mar 2012). Available at http://daccess-dds-ny.un.org/doc/UNDOC/GEN/G13/117/94/PDF/G1311794.pdf?OpenElement
Francioni F (2007) Access to justice as a human right. Oxford University Press, New York
Hodges C (2001) Multi-party actions: a European approach. Duke J Comp Int Law 11(2):321–354
Krieger LH (2003) Backlash against the ADA. University of Michigan Press, Ann Arbor
McEwen C (2014) Determining whether to initiate a mediation program and how its structure affects results – costs for parties, 1 Mediation: Law, Policy and Practice § 14:5, Cole S, McEwen C, Rogers N, Coben J, Thompson P, Thomson Reuters
Ogletree CJ Jr, Sapir Y (2004) Keeping Gideon's promise: a comparison of the American and Israeli public defender experiences. N Y Univ Rev Law Soc Change 29(2):203–235
Ortoleva S (2011) Inaccessible justice. ILSA J Int Comp Law 17(2):281–320
Turquet L et al (2015) Progress of the world's women: in pursuit of justice. U.N. Women. Available at http://progress.unwomen.org/pdfs/EN-Report-Progress.pdf
Watson GD (2001) Class actions: the Canadian experience. Duke J Comp Int Law 11(2):269–288

Further Reading

Friedman LM (2009) Access to justice: some historical comments. Fordham Urban Law J 37(1):3–15
Johnson E (1985) The right to counsel in civil cases: an international perspective. Loyola Law Rev 19(2): 341–436

Accountability

► Codes of Conduct

Act-Based Sanctions

Marie Obidzinski
CRED, Université Panthéon-Assas Paris II, Paris, France

Abstract
An act-based sanction punishes actions independently of the actual occurrence of harm. On the opposite, harm-based sanctions are imposed when harm occurs. When harm is certain, the distinction between act-based and harm-based sanctions is not relevant. On the opposite, when harm is probable, it is worth distinguishing act and harm based sanctions. In a basic public law enforcement framework, optimal deterrence can be achieved by both types of sanctions. This statement does not hold when potential offenders either have limited assets or are risk adverse or can invest in avoidance activities. Again, this assertion is not true when punishment is costly.

Definition

An act-based sanction punishes actions directly. It does not depend on the actual occurrence of harm.

Act-Based Sanctions Versus Harm-Based Sanctions

Act-based sanctions are a tool to control undesirable acts. Undesirable acts are those whose private gains are lower than the expected social harm (Shavell 1993). Act-based sanctions are generally opposed to harm-based sanctions. While the latter are imposed on the basis of the actual occurrence of harm, act-based sanctions directly punish actions, irrespective of the actual occurrence of harm (Polinsky and Shavell 2007). The distinction between harm- and act-based sanctions refers to the timing of legal intervention, before any harm or after. This distinction between act- and harm-based sanctions is relevant when harm is probabilistic, rather than certain. In the case where an action creates harm with certainty, there is no difference between setting the sanction on the basis of action or harm. Examples where harm is probabilistic are numerous in safety regulation, in environmental law (Rousseau and Blondiau 2014), or in traffic law. In crime law, attempts are a typical case where the act did not result in harm.

Shavell (1993) analyzes the choice between prevention, act-based sanctions and harm-based sanctions are in relation with the structure of law

enforcement. Prevention refers to the use of force to forbid the action. This is the case when a firm has no right to sell a good, when a person is put in jail, or when fences are built. No choice is left to the person; law is directly enforced. On the other hand, act-based and harm-based sanctions tend to influence the choice of the actors. It is worth noting that in criminal law, all three methods are used by law enforcers. A policeman can prevent the crime and arrest on the basis of an attempt or on the basis of harm. On the other hand, safety regulation is mostly characterized by act-based sanctions (Shavell 1993). The timing of legal intervention has also been analyzed by the ex post liability versus ex ante regulation literature since Wittman (1977). Act-based sanctions refer to input monitoring and harm-based sanctions to output monitoring.

The aim of the threat of sanctions is to provide the right incentives in order to separate desirable and undesirable acts, as actors differ in the private benefits they receive from acts. The threat of sanctions induces decision-makers to internalize the expected costs of their acts. Law enforcement theory therefore balances the advantages and benefits of act- versus harm-based sanctions according to circumstances.

Theoretically, in a basic law enforcement framework, optimal deterrence can be achieved by both types of sanctions. Assume that the probability of detection and conviction is exogenously set and equal under act- and harm-based sanctions. The monetary fine should be higher under harm-based sanctions in order to reach the same deterrence level as under act-based sanctions, given that the harm is probabilistic. This balance can be tilted in favor of act-based sanctions when individuals or firms have limited assets or are risk adverse (Polinsky and Shavell 2007). The reverse is true when punishment is costly. Harm-based sanctions can also incite people to engage in harm avoidance activities, where possible.

Another significant determinant of the choice between act- and harm-based sanctions lies in the information possessed either by the government or by the potential offenders on the expected harm (Garoupa and Obidzinski 2011). The level of deterrence of act-based sanctions depends on the belief of the agents making the law (the authority). The level of deterrence of harm-based sanctions is more dependent to the beliefs of the people to whom the law applies (Friedman 2000). Beliefs regarding the probability of harm may considerably differ in the population. This might justify the use of act-based sanctions.

Cross-References

▶ Criminal Sanctions and Deterrence
▶ Public Enforcement

References

Friedman DD (2000) Law's order: what economics has to do with law and why does it matter? Princeton University Press, Princeton

Garoupa N, Obidzinski M (2011) The scope of punishment: an economic theory. Eur J Law Econ 31(3):237–247

Polinsky AM, Shavell S (2007) Handbook of law and economics, vol 1. Elsevier, Amsterdam

Rousseau S, Blondiau T (2014) Act-based Versus Harm-based Sanctions for Environmental Offenders. Env Pol Gov 24:439–454

Shavell S (1993) The optimal structure of law enforcement. J Law Econ 36:255–287

Wittman D (1977) Prior regulation versus post liability: the choice between input and output monitoring. J Leg Stud 6:193–211

Administrative Corruption

Maria De Benedetto
Department of Political Science, Roma Tre University, Rome, Italy

Abstract

The widespread interest at national and international level in combating administrative corruption is strictly connected with the idea that it produces many negative effects, distorts incentives and weakens institutions. On the other hand, administrative corruption has been also considered as an extra-legal institution which –

under certain conditions – could even produce positive effects.

Anticorruption strategies have been developed with reference to a Principal-Agent-Client model or using an incentive-disincentive approach as well as an ethical perspective.

However, preventing corruption needs a toolbox: good quality regulation, also when regulation determines sanctions; controls, which should be sustainable and informed to deterrence and planning; administrative reforms, in order to reduce monopoly and discretionary powers, to strengthen the Civil Service and to ensure transparency and information.

Definition

Abuse of public power for private gain.

Administrative Corruption: The Definition Debate

Defining administrative corruption is not a simple task. There is large agreement about the idea that corruption crosses legal systems, history, and cultures and that it is "as old as government itself" (Klitgaard 1988, p. 7) and a "persistent and practically ubiquitous aspect of political society" (Gardiner 1970, p. 93). At the same time, there is an agreement about the separate idea that corruption is a relative concept, that it should be understood only inside a specific cultural context, and that a behavior which is considered to be corrupt in one country (or at one time) could be considered not to be corrupt elsewhere (or at different times): in other words, "corruption is the name we apply to some reciprocities by some people in some context at some times" (Anechiarico and Jacobs 1996, p. 3).

Despite this difficulty, scholars have provided a number of definitions starting from various points of view.

A first approach has focused on the moral stance of corruption (Banfield 1958) and "tends to see corruption as evil" (Nye 1967, p. 417), requiring changes in "values and norms of honesty in public life," because without "active moral reform campaigns, no big dent in the corrosive effects of corruption is likely to be achieved" (Bardhan 1997, p. 1335).

The second approach, however, has considered corruption also from an economic point of view, highlighting that under certain conditions, it might produce positive effects. As a consequence, corruption should be considered more objectively because it represents an "extra-legal institution used by individuals or groups to gain influence over the action of the bureaucracy" and, moreover, because "the existence of corruption *per se* indicates only that these groups participate in the decision-making process" (Leff 1964, p. 8).

Important contributions to the definition debate (Klitgaard 1988; Rose-Ackerman 1999; Ogus 2004) recognized that – in any case – "economics is a powerful tool for the analysis of corruption" (Rose-Ackerman 1999, p. xi).

Furthermore, corruption "involves questions of degree" (Klitgaard 1988, p. 7): in this light, "petty" administrative corruption has been distinguished from political (or "grand") corruption which "occurs at the highest level of government and involves major government projects and programs" (Rose-Ackerman 1999, p. 27). There is also systemic corruption when it is "brought about, encouraged, or promoted by the system itself. It occurs where bribery on a large scale is routine" (Nicholls et al. 2006, p. 4).

One of the most important definitions of corruption is "behavior which deviates from the formal duties of a public role because of private-regarding (personal, close family, private clique) pecuniary or status gains; or violates rules against the exercise of certain types of private-regarding influence" (Nye 1967, p. 419).

However, the most used definition has been provided by transnational actors, such as the World Bank, which refers to a concept of corruption vague enough to be used in every national context and to include all kinds of corruption: "the abuse of public power for private benefit" (World Bank, Writing an effective anticorruption law, October 2001, Washington, 1; World Bank. Helping countries combat corruption: progress at the World Bank since 1997, Washington, 2000).

Corruption reveals a rent-seeking activity (Lambsdorff 2002), an effort to achieve an extra income (J. Van Klaveren, The Concept of Corruption, in Heidenheimer et al. 1993, 25) by circumventing (as in the case of creative compliance, R. Baldwin et al., Understanding Regulation: Theory, Strategy and Practice, Oxford University Press, 2012, 232) or directly by breaking the law.

Effects of Administrative Corruption on Administrative Performance

Economic effects of administrative corruption are controversial, so are the effects of corruption on administrative performance (D.J. Gould and J.A. Amaro-Reyes, The Effects of Corruption on Administrative Performance. Illustrations from Developing Countries, World Bank Staff Working Papers, number 580, Management and Development Series, number 7, 1983; see also Nye 1967).

On one side, there is a point of view which tends to overestimate the positive effects of corruption. Many aspects have been mentioned in this regard: positive effects have been recognized especially when corruption is "functional" to the agency's mission (Gardiner 1986, p. 35) or when it secures, in some cases, economic development (Leff 1964) or when it corrects "bad" (inefficient) regulation (Ogus 2004, pp. 330–331). In other words, corruption "may introduce an element of competition into what is otherwise a comfortably monopolistic industry" (Leff 1964, p. 10).

On the other side, there is a different point of view which recognizes that corruption "can determine who obtains the benefits and bears the costs of government action" (Rose-Ackerman 1999, p. 9) and, in so doing, that "distorts incentives, undermines institutions, and redistributes wealth and power to the undeserving. When corruption undermines property rights, the rule of law, and incentives to invest, economic and political development are crippled" (Klitgaard 2000, p. 2).

However, even though some economic and bureaucratic benefits of corruption have been recognized, "in the large majority of cases intuition suggests that there will be significantly outweighed by the costs" (Ogus 2004, p. 333).

In particular, corruption has a "destructive effect [. . .] on the fabric of society [. . .] where agents and public officers break the confidence entrusted to them" (Nicholls et al. 2006, p. 1; see also O.E. Williamson, Transaction-Cost Economics: The Governance of Contractual Relations, in "Journal of Law and Economics", Vol. 22, No. 2, 1979, 242: "Governance structures which attenuate opportunism and otherwise infuse confidence are evidently needed").

Furthermore, corruption has been regarded as a "sister activity" of taxation, but it has been considered to be more costly: in fact, it presupposes secrecy which "makes bribes more distortionary than taxes" (Shleifer and Vishny 1993, p. 600) and which represents the greatest threat to the integrity of public officials.

Combating Corruption: Why?

High levels of corruption have been econometrically associated with lower levels of investments as a share of Gross Domestic Product (GDP) (Mauro 1995) even if "the connection between corruption and the lack of growth is more often assumed than demonstrated" (Ogus 2004, p. 229). This is one of the reasons which justifies the increasing national interest in combating corruption.

There is, also, a wider interest in anticorruption policies characterized by an international effort which presents important practical consequences, e.g., the case for the anticorruption prerequisites in World Bank loans to developing countries (Guidelines on "Preventing and Combating Fraud and Corruption in Projects Financed by IBRD Loans and IDA Credits and Grants", 2006-revised 2011) which require the putting in place of regulatory and institutional mechanisms to fight corruption (Ogus 2004, p. 329). Anticorruption policies are, in such cases, a sort of condition for obtaining a loan.

In order to clarify some aspects relevant to combating corruption, it would be important to make note of the lack of data in this area (Gardiner 1986, p. 40) as well as of the difficulty in

measuring it (Anechiarico and Jacobs 1996, p. 14). Furthermore, it is also important to remember that a large part of the institutional debate is focused on corruption perception rather than on corruption reality and (finally) that the only sustainable institutional goal is reducing corruption because corruption is considered impossible to eradicate (Ogus 2004, p. 342): "anti-corruption policy should never aim to achieve complete rectitude" (Rose-Ackerman 1999, p. 68).

As a consequence, "the optimal level of corruption is not zero" (Klitgaard 1988, p. 24). Anticorruption controls, in fact, are expensive (Anechiarico and Jacobs 1996), so it could be necessary to decide the extent to which we should combat corruption: the point of intersection of the curves – which describe the quantity of corruption and the marginal social cost of reducing corruption – identifies "the optimal amount of corruption" (Klitgaard 1988, p. 27). In other words, when we say "why combat corruption?", in some way we simply mean "why keep corruption under control?"

Looking closely at the question, one of the most important reasons for which corruption should be controlled (a reason characterized at the same time by a moral and an economic stance) is that corruption represents a "form of coercion, namely economic coercion" (C.J. Friedrich, Corruption Concepts in Historical Perspective, in Heidenheimer et al. 1993, 16) which produces a fundamental distortion in the economic process, artificially separating economic activity and its result into two abstract concepts (M. De Benedetto, *Ni ange, ni bête*. Qualche appunto sui rapporti tra morale, economia e diritto in una prospettiva giuspubblicistica, in "Nuove autonomie", no. 3, 2010, 657, quoting the Italian legal philosopher Giuseppe Capograssi). The result benefits someone else even though it belongs to others (see A.G. Anderson, Conflicts of Interest: Efficiency, Fairness and Corporate Structure, in "UCLA Law Review", vol. 25, 1978, 794). In so doing, corruption changes the value (i.e., the result of the economic activity, at the same time moral and economic value) into mere advantage and the economic process in itself is corrupted.

Preventing Corruption: How?

Anticorruption could be considered as a "comprehensive strategy" which should be built with a number of tools (Rose-Ackerman 1999, p. 6; J.A. Gardiner and T.R. Lyman, The Logic of Corruption Control, in Heidenheimer et al. 1993, 827).

The first point in preventing corruption is to recognize that there are different kinds of transaction which directly produce (or indirectly stimulate) corrupt behavior and that these should be considered as separate battlefields which need specific tools.

In this regard, some contributions have used the "Principal-Agent-(Client)" model (Banfield 1975; Rose-Ackerman 1978; Klitgaard 1988; Della Porta and Vannucci 2012). In ordinary cases of corruption, there is a bribe giver and a bribe taker, but corrupt transactions involve three actors: the Principal (the State and its citizens, always considered the victims), the Agent (the civil servant in charge of administrative tasks), and the Client (the enterprise or the citizen, e.g., as tax payer).

Corrupt transactions in strict sense can be performed in the Agent-Client relationship (e.g., bribery, extortion). The Principal-Agent relationship as well as the Principal-Client relationship could present behavior oriented toward illicit rent seeking (e.g., in the first case, internal fraud and theft of government properties; in the second case, tax frauds or illegal capital transfers), very often facilitated by widespread conflicts of interest (Auby et al. 2014), but "this is not considered to be corruption since it does not include the active (or passive) collusion of an agent of the state" (Robinson 2004, p. 110).

The institutional response to prevent direct or indirect corruption in these different kinds of transaction involves a tool-kit: internal or external controls over administrative action (P/A), inspections on private economic activities (P/C), and criminal investigations into specific cases of corruption (A/C).

The second point in order to prevent corruption regards the opportunity to adopt an incentive/disincentive approach, changing the system of

rewards and penalties (Klitgaard 1988, p. 77; Gardiner 1986, p. 42) in a mix of "carrots and sticks" (Rose-Ackerman 1999, p. 78). There is large agreement about the opinion that "corrupt incentives exist because state officials have the power to allocate scarce benefits and impose onerous costs" (Rose-Ackerman 1999, p. 39). Reducing incentives to corruption and increasing its costs could involve structural reform (see infra 4.3) as "the first line of attack in an anticorruption campaign" (Rose-Ackerman 1999, p. 68).

The third point implies a problematic analysis of the large recourse – in institutions as well as in business, at national level as well at the international one – to ethical codes and other similar tools in order to ensure ethical responses to corruption and to strengthen anticorruption policies by stimulating individual morality. However, this tendency to regulate ethics is more effective in some cultural contexts than in others but could produce side effects. Ethics is typically free, while law is characterized by coercion (and by sanctions). If we use legal provisions (or any kinds of sanction) to induce ethical behavior, we are transforming a free behavior into a legal obligation, reducing the moral involvement of the individual, even if the legal provision is established by soft laws (such as ethical codes): this is a real paradox in regulating ethics (M. De Benedetto, *Ni ange, ni bête* cit., 656; see also Anechiarico and Jacobs, 1996, p. 202). In other words, tools should be consistent with the objective: law can establish incentives for moral behavior but cannot either impose or produce a moral (free) behavior by coercive means.

Regulation

As we have seen, the problem of corruption has in part a moral and a social stance, so a first step in preventing corruption would be for regulation to accept its own limits and to recognize that not only legal but also extralegal norms operate (on this point see R. Cooter, Expressive Law and Economics, in "The Journal of Legal Studies", vol. XXVII, June 1998, 585).

Furthermore, anticorruption policies have better chances of success if legal provisions and public opinion converge. The problem is particularly relevant in cases of "gray" corruption (A.J. Heidenheimer, Perspectives on the Perception of Corruption, in Heidenheimer et al. 1993, 161) in which there is a mismatch between what is considered corruption by public opinion and what is corruption by law. So, if regulation wants to achieve anticorruption objectives, it should take into account the social and moral context in which it will be applied and – in this way – it will strengthen enforcement and increase compliance: "the majority of government employees are honest, not because of rules, monitoring or threats but because of value and personal morality" (Anechiarico and Jacobs 1996, p. 202).

Moreover, it has long been clear that the functioning of market economy needs "a firm moral, political and institutional framework," which implies "a minimum standard of business ethics". The market economy, indeed, is not capable of increasing the "moral stock" by itself because "competition reduces the moral stamina and therefore requires moral reserves outside the market economy" (W. Röpke, The Social Crisis of our Time, Transaction Publishers, 1952, 52). This seems to be even more true when the individual choice on ethical rules takes place inside large groups (J.M. Buchanan, Ethical Rules, Expected Values, and Large Numbers, in "Ethics", vol. 76, 1965, 1).

Secondly, far from being a solution, regulation is recognized as a direct factor that promotes corruption (Tanzi 1998, p. 566). Overregulation could increase bureaucratic power and multiply the opportunities for creative compliance, nurturing a corruptible social environment and allowing more and more corruption: "the possibility of its transgression or perversion is always already inscribed into the law as hidden possibility. This, then, is the secret of law" (Nuijten and Anders 2007, p. 12).

Thirdly, since regulatory processes are fragile, special attention should be paid to sensible steps in the procedures: consultations, for example, could "increase the opportunity for corrupt transactions" (Ogus 2004, p. 341).

Starting from these premises, the problem seems to be not anticorruption regulation but good regulation in itself, regulation capable of

making rules effective, of ensuring enforcement, and of increasing compliance (in general on this point, Becker and Stigler 1974). Regarding the content of such good regulation, anticorruption objectives could be achieved thanks to reducing monopoly and discretion (Rose-Ackerman 1999) as we will see later (par. 4.3).

Another important regulatory matter in preventing corruption concerns sanctions (Klitgaard 1988, p. 78). They should be well calibrated because they respond to an intrinsically economic logic – indispensible to making laws effective – and because they can even influence the amount of the bribe: "penalizing the official for corruption changes the level of the bribe he demands, but does not change the essence of the problem" (Shleifer and Vishny 1993, p. 603; Ogus 2004, p. 336; see also Svensson 2003). On the other hand, it should be clear that "in the presence of corruption, it is optimal to impose (or at least threaten to impose) nonmonetary sanctions more often" (Garoupa and Klerman 2004, p. 220) as well as to reward enforcement (Becker and Stigler 1974, p. 13) and compliant groups (Gardiner and Lyman, The Logic of Corruption Control cit., in Heidenheimer et al. 1993, 837; Ogus 2004, p. 337): this idea is not new and was, in a similar form, already proposed in 1766 by Giacinto Dragonetti in his "Treatise on Virtues and Awards" (first English translation 1769).

Controls

In order to prevent corruption, controls are needed because human behavior is fallible and corruptible, because human behavior changes when subject to controls, and because corruption lives in the dark and controls may constitute the most important tool in rebalancing the asymmetric information between corrupt people and institutions (in this regard, compensating whistleblowers has been considered critically by Anechiarico and Jacobs 1996, p. 199 and Ogus 2004, p. 338). Among the different kinds of controls, inspections constitute the strongest tool, because they are characterized by coercive power.

On the other hand, traditional corruption controls have been considered inadequate and even "outdated and counterproductive" (Anechiarico and Jacobs 1996, p. 193).

Furthermore, it should be taken into account that controls (e.g., inspections) have a hybrid nature: not only are they a way to combat or prevent corruption but also they are real occasions for corrupt transactions because they represent a concrete contact between the Agent and the Client, particularly sensitive and dangerous when the Client has an interest to maintain (in any case and at any condition) the extra income which comes from illicit activities or when the Agent (who want to achieve an illicit extra income) has the opportunity to extort the Client.

The system of controls should be, indeed, sustainable from an administrative point of view, also in the field of anticorruption. In particular, it would be important to reduce the number of controls, because they represent a cost (not only for public administration but also for enterprises and citizens; see in this regard the Hampton Report, H.M. Treasury, Reducing administrative burdens: effective inspection and enforcement, March 2005) and because they are (as we have seen) occasions for corruption.

At the same time, it is important to increase their effectiveness in preventing corruption cases: for this purpose, anticorruption controls as a system should be informed by deterrence. The most relevant general contribution on this topic (Becker 1968) suggests that the individual decision about compliance is a result of an economic reasoning which connects the cost of compliance, the size of the penalty, and the risk of incurring the penalty. This reasoning, on the same grounds, contributes to establishing the eventual size of the bribe (Ogus 2004, p. 336). Furthermore, planning controls in anticorruption policies should be guided by a risk-based approach (in general, R. Baldwin et al., Understanding Regulation cit., 281), capable of mapping the most dangerous areas of administrative activity (in terms of probability of corruption) and capable of focusing – between the possible objects of control – on cases in which it is more probable to find evidence of corruption.

Administrative Reforms

There is a sort of conflict – observed by some scholars – between anticorruption policies and administrative reform. When anticorruption prevails, administrative reforms seem to be reduced in importance or to become marginal: "the logic of antibureaucratic reform leads to a model of public administration that ignores corruption, while the logic of anticorruption reform ignores public administration" (Anechiarico and Jacobs 1996, p. 204). It could even be possible that governments' anticorruption effort produce further costs "[. . .] not only in terms of the funds spent to control corruption, but in the deflection of attention and organizational competence away from other important matters" (Klitgaard 1988, p. 27).

At the same time, it could be useful to approach the problem of administrative reforms (in the perspective of anticorruption) in a practical manner, because "in terms of economic growth the only thing worse than a society with a rigid, overcentralized, dishonest bureaucracy is one with a rigid, overcentralized, honest bureaucracy" (Huntington 1968).

The first idea, in this regard, is to reduce the public sector: "the only way to reduce corruption permanently is to drastically cut back government's role in the economy" (Becker 1997).

The second relevant aspect is to reduce monopoly and discretionary powers (Ogus 2004, p. 331): "insofar as government officials have discretion over the provision of these goods [licenses, permits, passports and visas], they can collect bribes from private agents" (Shleifer and Vishny 1993, p. 599).

The third aspect regards the civil service (Rose-Ackerman 1999, p. 69) also because there are bureaucracies which seem to be less corruptible than others (S. Rose-Ackerman, Which Bureaucracies are Less Corruptible?, in Heidenheimer et al. 1993, 803). The question should be analyzed both from the point of view of civil service independence (Anechiarico and Jacobs 1996, p. 203) and from the point of view of civil service incentive payments ("often cited as one of the most effective ways of fighting corruption", Bardhan 1997, p. 1339).

The fourth aspect involves procedural and organizational design (Ogus 2004, p. 338) as well as administrative cooperation, crucial in order to enforce regulation and to effectively prevent corruption cases: in fact, "internal organisation of institutions influences their members' propensity to corruption" (Carbonara 2000, p. 2). Among other aspects, increasing international cooperation in combating corruption is indispensible: this is clear at the EU level, where it was recently affirmed that "anti-corruption rules are not always vigorously enforced, systemic problems are not tackled effectively enough, and the relevant institutions do not always have sufficient capacity to enforce the rules" (EU Anticorruption Report, COM 2014, 38 final, 2). Furthermore, this is confirmed at further levels, as in the case of GRECO, Group of States against Corruption, established in order "to improve the capacity of its members to fight corruption" (Statute of the GRECO, Appendix to Resolution (99) 5, art. 1) or as in the case of the OECD Convention on Combating Bribery of Foreign Public Officials in International Business Transactions.

Overall, indeed, there is a general problem of transparency and information. Transparency reduces the opportunities for corruption "through procedures which make the process and content of decision-making more visible" (Gardiner and Lyman, The Logic of Corruption Control cit., in Heidenheimer et al. 1993, 830). In the same way, information on what the Agent and the Client are doing allows the principal "to deter corruption by raising the chances that corruption will be detected and punished" (Klitgaard 1988, p. 82).

However, every anticorruption project needs a fine-tuning (Anechiarico and Jacobs 1996, p. 198) which implies both a legal and an economic approach, but which should by now be open to the contributions of other disciplines (e.g., behavioral sciences). In any case, "scholars of law and economics" will continue developing studies in the area of corruption also "because it raises fascinating issues about the enforcement of law in the broadest sense" (Bowles 2000, p. 480).

A

References

Anechiarico F, Jacobs JB (1996) The pursuit of absolute integrity. How corruption control makes government ineffective. The University of Chicago Press, Chicago
Auby JB, Breen E, Perroud T (eds) (2014) Corruption and conflicts of interest. A comparative law approach. Edward Elgar, Cheltenham
Banfield EC (1958) The moral basis of backward society. Free Press, New York
Banfield EC (1975) Corruption as a feature of government organization. J Law Econ 18:587–605
Bardhan P (1997) Corruption and development: a review of issues. J Econ Lit 35:1320–1346
Becker GS (1968) Crime and punishment: an economic approach. J Polit Econ 76:169
Becker GS (1997) Want to squelch corruption? Try passing out raises. Business Week, no. 3551, 3 Nov 1997, p 26
Becker GS, Stigler GJ (1974) Law enforcement, malfeasance, and compensation of enforcers. J Leg Stud 3:1–20
Bowles R (2000) Corruption. In: Bouckaert B, De Geest G (eds) Encyclopedia of law and economics, volume V. The economics of crime and litigation. Edward Elgar, Cheltenham, pp 460–491
Carbonara E (2000) Corruption and decentralisation, dipartimento scienze economiche, Università di Bologna, Working papers no. 342/83
Della Porta D, Vannucci A (2012) The hidden order of corruption. An institutional approach. Ashgate Publishing, Farnham, Survey
Gardiner JA (1970) The politics of corruption. Russel Sage Foundation, New York
Gardiner JA (1986) Controlling official corruption and fraud: bureaucratic incentives and disincentives. Corrupt Reform 1:33–50
Garoupa N, Klerman D (2004) Corruption and the optimal use of nonmonetary sanctions. Int Rev Law Econ 24:219–225
Heidenheimer AJ, Johnston M, Levine VT (eds) (1993) Political corruption. A handbook. Transaction Publishers, New Brunswick (third printing, first printing 1989)
Huntington SP (1968) Political order in changing societies. Yale University Press, New Haven, p 69
Klitgaard R (1988) Controlling corruption. University of California Press, Berkeley and Los Angels
Klitgaard R (2000) Subverting corruption. Financ Dev 37:2–5
Lambsdorff JG (2002) Corruption and rent-seeking. Public Choice 113:97–125
Leff N (1964) Economic development through bureaucratic corruption. Am Behav Sci 8:8–14
Mauro P (1995) Corruption and growth. Q J Econ 110(3):681–712
Nicholls C, Daniel T, Polaine M, Hatchard J (2006) Corruption and misuse of public office. Oxford University Press, Oxford
Nuijten M, Anders G (eds) (2007) Corruption and the secret of law. A legal anthropological perspective. Ashgate Publishing, Aldershot, Hampshire
Nye JS (1967) Corruption and political development: a cost-benefit analysis. Am Polit Sci Rev 61(2): 417–427
Ogus A (2004) Corruption and regulatory structures. Law Policy 26:329–346
Robinson M (ed) (2004) Corruption and development. Routledge, London
Rose-Ackerman S (1978) Corruption. A study in political economy. Academic press, New York
Rose-Ackerman S (1999) Corruption and government: causes, consequences and reform. Cambridge University Press, Cambridge
Shleifer A, Vishny RW (1993) Corruption. Q J Econ 108:599–617
Svensson J (2003) Who must pay bribes and how much? Evidence from a cross section of firms. Q J Econ 118(1):207–230
Tanzi V (1998) Corruption around the world: causes, consequences, scope, and cures. IMF Staff Pap 45(4): 559–594

Administrative Courts

Sofia Amaral-Garcia
Center for Law and Economics, ETH Zurich, Zürich, Switzerland

Definition

Administrative courts are courts that specialize in and apply administrative law, a branch of public law that focuses on public administration. In other words, administrative courts adjudicate cases of litigation involving the state and private citizens (both individuals and corporations) and apply the law that governs the activities of administrative agencies of government. There is no homogeneous model for administrative courts across jurisdictions. Even if one considers a broad distinction of legal families, it is possible to find many distinct models. Every jurisdiction has a way of solving litigation with the state, independently of having a specialized administrative court.

Introduction

Administrative courts are courts that specialize in administrative law, a branch of public law that

focuses on public administration. In other words, administrative courts adjudicate mainly cases of litigation involving the state (both institutions and public officials) and private citizens (both individuals and corporations) and apply the law that governs the activities of administrative agencies of government. The role of these types of courts is directly related to the existence and functions of a state or government: in the one hand, it would make no sense for these courts to exist without a state; on the other hand, these courts may have distinct roles depending on the specific functions of the state, which might vary across jurisdictions and time.

Constitutional courts (if existent) also adjudicate cases related to the government. For this reason, one might argue that it is hard to draw a clear-cut line dividing the types of cases that each court hears. Assigning the tripartite division of government branches to administrative courts and constitutional courts might help somehow drawing a division, even though it is never an exclusive one. Generally speaking, administrative courts adjudicate primarily cases related to the executive function of the state, and constitutional courts adjudicate cases related to the legislative, judicial and executive functions.

Institutions in general, and courts in particular, matter for economic growth and development (e.g., Mahoney 2001). Scholars from law and economics, institutional economics and legal studies have acknowledged their role and importance, especially in recent decades. Many answers are still unanswered, in particular if we narrow the analysis down to administrative courts as extensive empirical analysis in administrative courts is virtually nonexistent. The emerging field of comparative administrative law has comprehensively examined and explained substantial differences of administrative law across countries. Nevertheless, the collection of administrative courts' models (even within the same legal families) embedded in highly complex legal, political, and economic systems create a challenge to comparative analysis. In fact, it is difficult to assess which model is the best or to quantify how much one model is better than the other because an extensive number of factors play a role as well. The transplant of other models does not necessarily result in better outcomes precisely for this reason.

"The core idea of government under law requires an independent institution willing to enforce that law against the administration – both the political and the bureaucratic administration – whenever the law is broken" (Bishop 1990, pp. 492–493) – this is where administrative courts come in. Administrative courts hear a variety of cases similar to those involving individuals, such as contracts, torts or property, as long as the state is one of the parties involved. Nevertheless, the state has particular functions that private citizens do not have. Declaring war, issuing passports, collecting taxes, having the monopoly of legitimate coercion (which also imposes the duty of procedural fairness), or having a monopoly over some activities and provision of certain goods or services are a few examples of state's specific functions (e.g., Cane 2011). Generally speaking, some of the most common areas of administrative law are taxes, immigration, and licensing which tend to make up for a significant proportion of cases that administrative courts hear. Moreover, these areas are more relevant in civil law countries given their inclination for *ex ante* control mechanisms rather than *ex post* control mechanisms as it tends to be the case in common law systems. To sum up, administrative courts also resolve disputes that are specific to the role of the state and protect citizens from state overreaching.

Historical Origins and Institutional Differences

Although it is not the purpose of this work to make an extensive analysis of all administrative courts in Europe or elsewhere, some basic notions of the models adopted in England and France are useful at this stage. France and England deeply influenced many legal systems across the world, even though these countries opted for two different ways of dealing with administrative law. In England, ordinary courts were in charge of hearing administrative cases, whereas this role was assigned to specialized state officials connected

to the executive branch in France. This is generally perceived to be the major difference between administrative law systems (Bignami 2012, p. 148). There were important historical differences in those two countries with respect to the rise of bureaucracy and administrative law that explain the option for "ordinary" or specialized courts.

In France, the seventeenth and eighteenth centuries were marked by intense conflicts between *intendants* (officers responsible to the Crown in charge of the provinces' administration) and *parlements* (powerful courts controlled by local elites). In fact, the "paralysis of the royal administrative of the ancient régime, caused by the *Parlements* (which were, in fact, judicial bodies), is often considered one of the causes of the French revolution and was a prime impetus for the French conception of separation of powers, in which the judicial courts lack jurisdiction over administrative acts of the State" (Massot 2010, p. 415). Napoleon created the *Conseil d'État* (Council of State), which became the successor of the *Conseil du Roi* (King's Council, a specialized review body) and the first specialized administrative court. The Council of State still maintains nowadays the dual function of adjudicating cases against the French administration and drafting government laws and rules. At the time of its creation, judicial review made by ordinary courts (i.e., the judicial power of the government) represented an intrusion on the executive power and these courts were not allowed to adjudicate claims against the government. This was the main rationale for assigning that task to a specialized body: the executive, not the judiciary (Cane 2011, p. 43; Bignami 2012, p. 149; see also Mahoney 2001).

In England, the distinction between public and private law has been historically less sharp not because England lacks public-law courts but because ordinary courts have jurisdiction over all types of disputes (Cane 2011). A constitutional struggle developed in the seventeenth century between the Stuarts and judges with respect to the judges' right to decide cases related to the royal power or to decide cases in which the king had an interest (Page and Robson 2014). At that time, the Stuart kings attempted to create separate courts to deal with cases related to the government in order to expand royal control. However, the victory of the Parliament established the independence of judges and everyone should obey the law. Courts developed judicial review, i.e., the revision of administrative decisions, as a way of supervising inferior government bodies (Cane 2011). Local elites with little central involvement would administer the business of government and appeals against government officers would be taken to courts of general jurisdiction (Bignami 2012, p. 149). Administrative tribunals developed during the twentieth century (Shapiro 1981, p. 111 & *seqs*.) which, in a very simplistic way, can be thought of as an adjudicatory body that is not a court (Cane 2010).

The German model of judicial review is an alternative to the French and English models. During the nineteenth century, German liberals endeavored to implement legal structures that would limit state power by the monarchs of German states. At the time, the ideas of the jurist Rudolf von Gneist strongly influenced German administrative law. Gneist was a strong advocator of an independent judicial review system that would ensure the protection of citizens' rights and contended that there should be a generalist administrative court independent of the executive. Moreover, the composition of the court should allow an independent judicial control of administrative power, so it would be staffed by professional administrators and respected citizens (Feld 1962, p. 496; Nolte 1994, p. 199). In 1872 Gneist's ideas were implemented with the Prussian Supreme Administrative Court, which was the highest judicial body of a three-tier system of administrative courts. This court exerted a great influence on the development of German administrative law (Feld 1962). According to the German model, currently the most widespread model in Europe, a specialized branch of the judiciary specializes in administrative law (Fromont 2006, p. 128). Civil judges and administrative judges follow the same recruitment process and guarantee of independence. The main difference is that administrative judges specialize in administrative law.

The spirit of the French, English, and German administrative models has been implemented elsewhere. Several countries adopted the French administrative model, having a Council of State separated from the judiciary. In continental Europe, some of these countries are the Netherlands, Belgium, Italy, and Greece. However, whereas cases of government liability are adjudicated by the Council of State in France, these cases are adjudicated by courts in Italy, Belgium, and the Netherlands. The British model of a generalist court has been implemented, among others, in Ireland, the USA, Australia, and New Zealand. Austria, Finland, Poland, Portugal, Spain, Sweden, and Switzerland have implemented a model of the German type.

The Evolution of the State and the Role of Administrative Courts

The current role of administrative courts is necessarily intertwined with the functions of the modern state, which are complex and diverse across jurisdictions. Moreover, these functions are not static and evolve continuously, together with economic and political development (the development of the welfare state has been particularly relevant). For instance, the provision of health, housing, education, electricity and transport, among others, went through diverse degrees of public ownership and control (Cane 2011). During the 1980s and the 1990s several Western countries experienced a shift in the boundary between the public and private sectors. With the aim of reaching a single market, the European Commission liberalized numerous network industries, such as gas, air transport, electricity, postal and railroad services (Custos 2010, p. 279). In some countries, there was a shift from public to private in many economic sectors due to privatizations (for example, in the UK, Portugal, Italy, France, and Spain). Meanwhile, a deregulation process took place in the USA, where government regulation was reduced in a number of sectors. The government opted for contracting-out some services to private companies and public-private partnerships emerged for performing functions that have previously been carried out by governmental bodies. A few examples are road repairs, highways constructions, and garbage collection.

In particular, nowadays the state is a constant presence in many spheres of daily life, its powers are vast, and its functions are considerably complex. An enormous machinery must be in place so that the state can perform its complex functions. The legislature approves laws and statutes that will be executed by government agencies, run by nonelected civil servants. A democratic government running under the law must provide a way of monitoring and supervising the performance of government agencies and bureaucrats. Therefore, administrative courts can be asked to perform judicial review of administrative decisions (e.g., if a public body acted beyond its powers or if a public body failed to act or perform a duty statutorily imposed on it).

The actions of state officials might also impose harm to citizens, very similarly to what happens in contract or tort. Some examples include medical liability of a doctor practicing in a public hospital or accidents caused by state-owned cars and driven by public employees. The difference in these cases is that the state is one of the parties, most likely the party supposedly causing the harm. Administrative courts might be called to adjudicate state liability and award damages in case the state is found liable, but the reliance on administrative courts to perform this task depends on each specific jurisdiction. In reality, there is no unique model for administrative courts or for adjudicating litigation involving the state. In terms of procedure, administrative courts in civil law tradition countries have exclusive jurisdiction over tort litigation against the state as sovereign (Dari-Mattiacci et al. 2010). Moreover, administrative courts follow rules of administrative procedure that tend to treat the state as a nonordinary defendant (differences might be found with respect to statute of limitation, liability standards or the possibility of having out-of-court settlements, to name only a few). Cases in which the state acts as a private entity might lead to jurisdictional ambiguity, and depending on the country these cases might end up in ordinary courts. In some countries, cases of jurisdictional conflict

might be addressed to a court of conflict (e.g., France, Italy, and Portugal).

Some Examples of Institutional Differences

It is nevertheless worth mentioning that a jurist trained in the USA might face some difficulties when trying to understand the functioning of administrative courts and the profession of administrative judges in Continental European civil law tradition countries. In the USA, much of administrative review is vested in public authorities and independent agencies, often described as administrative tribunals. The agency officials, whose task is to adjudicate cases according to the Administrative Procedure Act (APA) of 1946, are administrative law judges and administrative judges (Cane 2010, p. 427). These administrative officials are typically recruited by the Office of Personnel Management to become adjudicators in agencies that are part of the executive branch of government. The APA did not introduce specialized courts to deal with judicial review of administrative acts but it effectively created "a special form of jurisdiction that governs the review of agency decisions in ordinary courts. Especially to continental jurists, then, it is often worth emphasizing that the US legal system, in some sense, also distinguishes between administrative law questions and other legal disputes" (Halberstam 2010, p. 187). Moreover, in the US, the Court of Appeals for the District of Columbia (DC) Circuit has specialized in administrative law and is frequently the final court for administrative matters.

Recently in the UK there were important judicial reforms, namely the Constitutional Reform Act 2005 and the Tribunals, Courts and Enforcement Act 2007. According to this later reform, administrative courts are effectively being created. Additionally, the jurisdiction of some subject-specific tribunals (such as social security, education, taxation, pensions, and emigration) has been transferred to the new First-tier Tribunal and Upper Tribunal. The First-tier Tribunal comprises seven chambers has as main function deciding general appeals against a decision made by a Government agency or department. The Upper Tribunal comprises four chambers (one of which being the Administrative Appeals Chamber) and hears appeals from the First-tier Tribunal in points of law. In a way, recent times have introduced a shaper distinction between administrative and private English law.

Significant transformations also took place in French administrative law, namely with the creation of the administrative courts of appeal in 1987 and various reforms that extended the powers of administrative judges (in particular with respect to injunctions and the possibility of issuing urgent judgments to prevent illegal administrative behavior) (see Auby and Cluzel-Métayer 2012, p. 36). Currently, there are 42 administrative tribunals, 8 administrative courts of appeal and the highest level is still the *Conseil d'État*. French administrative courts do not adjudicate all administrative cases. The view is that the administration is only partly subjected to special rules and to public law, which translates in some cases ending up in ordinary courts even if they involve the administration. Delimiting the precise jurisdiction of French administrative courts is extremely complex (for more on this see Auby and Cluzel-Métayer 2012, p. 25).

With respect to Germany, administrative courts are one of the five branches of the judiciary, which besides administrative courts includes ordinary courts (civil and criminal), labor courts, fiscal courts, and social courts. The last three courts comprise the so-called special courts (e.g., Schröder 2012, p. 77). There are three levels of administrative courts: lower administrative courts, higher administrative courts, and the Federal Administrative Court. The total number of lower administrative courts depends on the population and size of each *Land* (state), there is at least one administrative court in each State but no more than one higher administrative court (Singh 2001, p. 187 & seqs.; Schröder 2012, p. 78 & seqs.).

Generally speaking, civil law systems tend to hear cases of judicial review within one administrative court (with a few exceptions, such as social security cases in some jurisdictions or taxes in Germany) whereas common law systems tend to have a different appeal tribunal for each agency (Bishop 1990, p. 525). However, the few examples presented above illustrate the complexity of

the organizational setting of administrative courts. Homogeneity cannot be found, not even at the highest court level of Continental European jurisdictions where some systems have one Administrative Supreme Court (e.g., Portugal), others have one Administrative Section within the Supreme Court (e.g., Spain) and others have none of these. Independently of institutional and legal differences, every jurisdiction has a way of solving litigation between the state and citizens and of appealing these decisions. The main differences are generally with respect to whether there is a specialized administrative court, whether there are particular procedural rules, and whether there are specialized courts to deal with conflicts of jurisdiction.

Do We Need Specialized Administrative Courts?

The debate on specialized administrative courts has raised several contentious questions that have been asked for decades (see, e.g., Nutting 1955; Revesz 1990). For instance, should the adjudicative function of administrative cases be vested in separate bodies and, if so, are courts the most appropriate institutions? Should administrative courts have original and appellate jurisdiction? Is it more beneficial to have specialized administrative courts to adjudicate cases involving the state or leave it to generalist courts? The answers to these questions remain controversial, with several arguments being found in favor and against specialized courts (see, among others, Dreyfuss 1990; Revesz 1990; Baum 2011).

Some of the potential benefits of specialized courts are extensive to administrative courts as well. The "neutral virtues" of judicial specialization are the quality of decisions, efficiency, and uniformity in the law (Baum 2011, pp. 4 & 32). Specialization should result in more correct decisions in complex areas of the law, precisely because the adjudicator will become an expert in a certain type of decisions. This can be particularly relevant in fields of law that involve complex technical skills. Moreover, specialization might allow decisions to become more uniform and coherent. One potential advantage of specialized courts on administrative matters might be that the judges feel more confident given their expertise, which may make them more willing to go against administrative decisions. If judges have the incentives and training to become experts in administrative law, it is also possible to have tailored procedures in court to deal with the particular features of the state as defendant (Dari-Mattiacci et al. 2010, p. 28). A major advantage of separate administrative courts is the possibility to develop a set of principles that accepts the specific nature of the state as defendant (such as access to information, evidence produced by the administration and control over administrative discretion) balancing the interests of citizens and the ability of the administration to pursue the public interest (Bell 2007, pp. 291–293 and p. 299). There are possible "nonneutral effects" of specialized courts on the substance, namely a change in the ideological content of judicial policy or the support for competing interests in a given field (Baum 2011, p. 4 & seqs.).

As for the arguments against specialized courts, there are also many. Courts and judges that become specialized in some particular area of the law might apply the law in a narrower way, and have fewer skills in applying concepts from other areas of law when necessary. Moreover, some arguments against specialized administrative courts are particularly relevant. Specialized administrative courts might be more easily captured by the state, which becomes more likely with the separation of jurisdictions. The marginal cost for the judge in deciding against the state is higher in administrative than in ordinary judicial courts (e.g., Mahoney 2001). Specialization makes accountability more difficult, because the knowledge of administrative law becomes a specific asset on human capital for administrative judges, which might become more dependent on the government precisely for this reason (see Dari-Mattiacci et al. 2010, p. 28). Hence, it might be more difficult for administrative judges to realize that the state has overreached either because judges exhibit systemic biases or because judges want to maintain their place as state officials. This

would be an undesirable outcome in case it would result in administrative judges issuing biased decisions in favor of the state. However, and in spite of the debate on this issue, the lack of empirical evidence makes it impossible to draw rigorous and extensive conclusions on this claim.

The recruitment of administrative judges becomes naturally a critical piece of the well-functioning of administrative courts. Traditionally, commonwealth systems have recognition judiciaries which, together with ordinary courts for administrative review, tends to result in a higher degree of autonomy in comparison with continental systems (Garoupa and Mathews 2014, p. 12). In the majority of civil law countries in Continental Europe, administrative judges follow a career as generalist judge and specialize in administrative law when serving in administrative courts. An exception is France, where most of the members of the *Conseil d'Etat* are civil servants, recruited from the *École Nationale d'Administration*, the elite school for training French government executives. In both civil and common law systems judges are appointed for life: in civil law, judges are appointed for life as judges (i.e., career tenure) and in common law judges are appointed for life in a specific court (i.e., court tenure).

It is not possible to properly assess the pros and cons of specialized administrative courts without considering the legal, administrative lawmaking and political systems in which a particular court is located in. In the end, judicial specialization may affect courts output in a very complex way. Whether the effects are good or bad depend on the particular system where the court is located. There are still limited empirical findings which makes it difficult to assess the effects of specialized administrative courts. Moreover, even if there were strong evidence that the benefits of specialized administrative courts outweigh the disadvantages, it would still be unclear what the best institutional model for administrative courts would be. Indeed, there are many different ways to structure specialized courts, namely specialized courts with generalized judges, generalized courts with exclusive special jurisdiction, specialization at the trial and/or appellate level, specialized trial courts with general appellate courts or general trial courts reviewed by special appellate courts (Dreyfuss 1990, p. 428). Furthermore, as the brief comparison among different jurisdictions has shown, it is possible to have distinct formats even at the highest court levels. Additionally, the interactions between judges and the state may have important implications which might be translated in more or less politicized courts (see Garoupa et al. 2012).

Concluding Remarks

Administrative courts are part of the system of checks and balances of the government system. Therefore, administrative courts have an important institutional role in modern societies, given that they still keep the task of protecting citizens from the powerful state. In the same way that modern societies have institutions and mechanisms to enforce the law and solve conflicts among private parties, it is fundamental to allow citizens to review government decisions and to be compensated if they have been harmed by government action or inaction. The function of administrative courts and the quality of their decisions can also have relevant economic impacts. All in all, it is not only important to have a way of making a claim against the state: it is also important that the decisions being held by the institutions in charge of the adjudication are not prostate biased.

Recent decades have introduced changes in the functions of the state, with the development of the welfare state being eventually the most relevant. Naturally, the role of administrative courts has also been affected by these changes. Concomitantly, important developments took place with repercussions in the administrative sphere, such as the creation of the European Union and the establishment of the European Convention on Human Rights. It is still unclear what the implications of global administrative on national administrative courts will be. Contributions from political science and law and economics would be fruitful and would bring important insights to the debate.

References

Auby J-B, Cluzel-Métayer L (2012) Administrative law in France. In: Seerden R (ed) Administrative law of the European Union, its Member States and the United States, 3rd edn. Intersentia, Antwerpen

Baum L (2011) Specializing the courts. University of Chicago Press, Chicago

Bell J (2007) Administrative law in a comparative perspective. In: Örücü E, Nelken D (eds) Comparative law: a handbook. Hart Publishing, Oxford

Bignami F (2012) Comparative administrative law. In: Bussani M, Mattei U (eds) The Cambridge companion to comparative law. Cambridge University Press, Cambridge, pp 145–170

Bishop W (1990) A theory of administrative law. J Leg Stud XIX:489–530

Cane P (2010) Judicial review and merits review: comparing administrative adjudication by courts and tribunals. In: Rose-Ackerman S, Lindseth PL (eds) Comparative administrative law. Edward Elgar, Cheltenham/Northampton, pp 426–448

Cane P (2011) Administrative law, 5th edn. Oxford University Press, Oxford

Custos D (2010) Independent administrative authorities in France: structural and procedural change at the intersection of Americanization, Europeanization and Gallicization. In: Rose-Ackerman S, Lindseth PL (eds) Comparative administrative law. Edward Elgar, Cheltenham/Northampton, pp 277–292

Dari-Mattiacci G, Garoupa N, Gómez-Pomar F (2010) State liability. Eur Rev Private Law 18(4):773–811

Dreyfuss RC (1990) Specialized adjudication. Brigham Young Univ Law Rev 1990:377–441

Feld W (1962) The German administrative courts. Tulane Law Rev XXXVI:495–506

Fromont M (2006) Droit administrative des États européens. Thémis droit, Paris

Garoupa N, Mathews J (2014) Strategic delegation, discretion, and deferrence: explaining the comparative law of administrative review. Am J Comp Law 62:1–33

Garoupa N, Gili M, Gómez-Pomar F (2012) Political influence and career judges: an empirical analysis of administrative review by the Spanish Supreme Court. J Empir Leg Stud 9(4):795–826

Halberstam D (2010) The promise of comparative administrative law: a constitutional perspective on independent agencies. In: Rose-Ackerman S, Lindseth PL (eds) Comparative administrative law. Edward Elgar, Cheltenham/Northampton, pp 185–204

Mahoney PG (2001) The common law and economic growth: Hayek might be right. J Leg Stud XXX:503–525

Massot J (2010) The powers and duties of the French administrative judge. In: Rose-Ackerman S, Lindseth PL (eds) Comparative administrative law. Edward Elgar, Cheltenham/Northampton, pp 415–425

Nolte G (1994) General principles of German and European administrative law – a comparison in historical perspective. Mod Law Rev 57(2):191–212

Nutting CB (1955) The administrative court. N Y Univ Law Rev 30:1384–1389

Page EC, Robson WA (main contributors) (2014) Administrative law. In: Encyclopædia Britannica Online. Retrieved 20 Aug 2014, from http://www.britannica.com/EBchecked/topic/6108/administrative-law

Revesz RL (1990) Specialized courts. Univ PA Law Review 138:1111–1174

Schröder M (2012) Administrative law in Germany. In: Seerden R (ed) Administrative law of the European Union, its Member States and the United States, 3rd edn. Intersentia, Antwerpen

Shapiro M (1981) Courts, a comparative and political analysis. The University of Chicago Press, Chicago

Singh MP (2001) German administrative law in common law perspective, Max-Planck-Institut für ausländisches öffentliches Recht und Völkerrecht. Springer, Berlin

Administrative Law

Giulio Napolitano
Law Department, Roma Tre University, Rome, Italy

Abstract

The law and economics literature has traditionally paid very little attention to administrative law history and rules.Economic analysis of law, however, can provide a useful explanation of the logic of administrative law, beyond the purely legal top-down approach. On one side, administrative law provides public bodies with all the needed powers and prerogatives to face and overcome different types of market failures. On the other side, administrative law is a typical regulatory device aiming to face some structural and functional distortions of bureaucracy, as a multi-principal agent. From this economic (and political) point of view, administrative law is much less stable than what it is usually thought to be. Frequent changes in both substantive and procedural rules can be explained as the outcome of repeated interactions among the legislator, the bureaucrats and the private stakeholders.

Definition

A set of rules governing public bodies and their interactions with private parties.

Logic and Ratios of Administrative Law

The law and economics literature has traditionally paid very little attention to administrative law history and rules. Notwithstanding its great expansion also in nonmarket fields, it is still an "unexpected guest" in the public law context (Ulen 2004). Major contributions come from other scientific approaches, like public choice (Farber and Frickey 1991) and political economy of law (McCubbins et al. 2007). They are different from the law and economics movement, of course, but they all share individualistic methodology and the rational interpretation of human behaviors. Unfortunately, on one side, economic and political literature has little confidence with highly technical and detailed provisions of administrative law. On the other side, administrative law scholars represent a close community, usually unwilling to open their minds to methodologies different from legal positivism. That's why, at least until today, the capacity of law and economics to shed new light into the field of administrative law has been very limited: a missed opportunity both for the law and economics movement and for the traditional scholarship of administrative law (Rose-Ackerman 2007).

According to the still dominant legal scholarship, administrative law is a coherent set of rules, ordered by some general principles, like the rule of law, impartiality, transparency, and proportionality, and characterized by its own specific features, such as the existence of public law entities, the special prerogatives of the Executive and its related branches, the decision-making procedure, and the judicial review. From this perspective, administrative law is represented as a quite stable institution, notwithstanding the frequent change of its specific rules.

Economic analysis, however, can play a very important role in explaining the intimate essence and the distinctive functions of administrative law. Two ratios explain the logic of administrative law, outside the purely legal top-down approach. On one side, administrative law provides public bodies with all the needed powers and prerogatives to face and overcome different types of market failures. On the other side, administrative law is a typical regulatory device aiming to face some structural and functional distortions of bureaucracy, as a multi-principal agent. From this economic (and political) point of view, administrative law is much less stable than what it is usually thought to be. Frequent changes in both substantive and procedural rules can be explained as the outcome of repeated interactions among the legislator, the bureaucrats, and the private stakeholders (Napolitano 2014).

Administrative Powers as a Mean to Correct Market Failures

The market failures theory offers a powerful explanation of the tasks that modern bureaucracies accomplish in modern societies and of the powers they exercise. Roles and prerogatives of public bodies, of course, are the outcome of complex social and political processes. But the market failures theory shows the existence of a rational basis underlying those historical developments and provides a useful test to verify the persisting need for the public intervention (Barzel 2002).

Many of the so-called sovereign functions of modern governments (like the protection of public security or the defense against external attacks) represent a solution to the problem of public goods. Non-excludability and non-rivalry make the private provision inefficient. Only governments, through the compulsory power of taxing, can solve the free-riding problem, which impedes the proper working of markets.

The public regulation of many economic activities aims to face other market failures (Posner 1997). Entry controls, technical requirements, and pollution standards address the problem of negative spillovers generated especially by industrial and noxious facilities. The regulation of prices and quality standards (in many countries enacted by independent authorities) limits the

market power of network operators in utilities like electronic communications and electricity. Obligations of disclosure in financial markets balance the informational asymmetry between financial institutions, investors, and savers.

Moreover, the government can make compulsory the consumption of merit goods, like education (at least for early stages) and health care (e.g., to prevent contagion, in case of epidemic diseases). The public provision of those services on a larger scale, however, is coherent also with the protection of constitutional rights and the design of the welfare state existing especially in Europe.

In all these cases, administrative law provides public authorities with the prerogative powers that are necessary to make citizens comply with regulations, orders, and duties to pay. This way, public authorities can overcome the high transaction costs that they would otherwise face in order to reach an (often impossible) agreement with private parties. From this point of view, administrative law plays an empowerment function in favor of public bodies, giving them all the prerogatives that are necessary to pursue the public interest. According to the rule of law, those powers must be assigned by the legislator through specific provisions. This is coherent from an economic perspective, because unilateral decisions are not Pareto efficient, reducing the welfare of affected parties. Their sacrifice, then, is justified only in the name of collective preferences aggregated by the legislator, as representative of the community.

Bureaucracy as a Multi-principal Agent

In all the contemporary societies and democracies, public policies and regulations that are needed to solve market failures and to satisfy the collective preferences must be implemented at the administrative level. Bureaucrats do not simply execute the law, like an automatic machine. On the contrary, they exercise a wide discretionary power, making choices among different possible alternatives, all coherent with the law. Delegation from elected bodies to expert agents is a world strategy, which responds to a principle of efficient division of labor (Mashaw 1985). The benefits of such a strategy, however, can be reduced by the emergence of opportunistic behavior, as it happens in every principal-agent relation (Horn 1995). Many provisions of administrative law pursue exactly the purpose of keeping under control the bureaucratic behavior in the attempt of preventing and punishing drift, capture, and corruption.

More precisely, bureaucracy is a multi-principal agent. At the national level, different political actors compete in order to assume the guidance of public administrations. Even if citizens are supposed to be the original patrons of public policies and of their administrative implementation, they have no direct voice. All the relevant powers are in the hands of elected representatives. But they can embody different political visions, trying to make them prevail in the administrative arena, through specific administrative law devices. This happens particularly in a system of divided government (Mueller 1996). In the USA, the President, especially through the cost-benefit analysis review (Posner 2001; Kagan 2001), and the Congress, enabling "fire alarm" by citizens and through special committees monitoring (McCubbins and Schwartz 1984; Ferejohn and Shipan 1990; Bawn 1997), adopt different legal strategies in order to assume the control over bureaucracy.

Other principals of public administration emerge at supranational level. A growing number of public policies are designed at global or macro-regional level by supranational institutions. They require national administrations to implement them in a coherent way, even though that can create a conflict with elected bodies at state level. Regulation and directives on the administrative implementation and enforcement of common policies represent a fundamental tool in order to achieve compliance against national drifts.

Public administrations also have multiple stakeholders. Individual citizens (and future generations) cannot easily act for the satisfaction and protection of their interests, unless they are specifically struck by an administrative decision. On the contrary, economic actors of different nature and dimension usually organize themselves to influence the exercise of administrative powers,

especially by regulatory agencies. But they compete and fight, one against the other, to obtain *ex ante* from legislators legal tools to play future games before agencies starting from an advantageous position (Macey 1992; Holburn and Vanden Bergh 2006).

All these repeated interactions among competing (and often in opposition) actors explain the existence of different strategies and conflicts in shaping administrative law and in using it as a powerful tool to keep bureaucracy under control.

The Industrial Organization of Bureaucracy

Public administrations, in the exercise of their tasks, produce goods and services in the interest of the entire community. Those goods and services cannot be delivered simply on the basis of a case-by-case bilateral agreement between a public agent and a private recipient. The production of those goods and services, on the contrary, requires the establishment of a stable organization, governed by a set of well-defined rules and codes of conduct.

The institutional design of bureaucracy responds to some relevant regularity in time and space. Everywhere, the Executive is at the center of the stage and exercises, directly or through dedicated units, the core functions of government (the protection of public security, the defense, the administration of justice, the collection of tax). According to a principle of efficient division of labor, agencies and independent authorities accomplish specialized, technical, and regulatory tasks, being separated and somehow insulated from the Executive. State-owned corporations are engaged in business, when considered to be the best way to manage market failures. In federal or decentralized countries, states, regions, and municipalities are autonomous entities and provide directly many goods and services to citizens, being closer and more accountable to them.

The legislator is usually free to change the institutional design of bureaucracy. Name and tasks of ministries can be changed. Agencies, independent authorities, and state-owned corporations can be established, merged, broken up, and abolished at any time. Tasks and autonomy degree of local governments can be modified or fine-tuned. However, all these operations are limited by path dependence. Legal and bureaucratic structures tend to be stable. Their existence is often protected by different stakeholders (employees, clients, pressure groups). That's why adaptation to changing needs and to institutional reforms may be slow or ineffective.

Bureaucracies work very differently from enterprises. Many of the goods and services they produce have no market value. Competition, which plays a very important role in pushing enterprises toward efficiency, doesn't exert any pressure in raising the quality of public services and in reducing the costs of government. Customer satisfaction, which is fundamental for every enterprise, is not so relevant to top managers in the public sector, whose main concern is to comply with regulations and political instructions. This happens also because public bodies have no market revenues. On the contrary, they are financed through the tax system and the subsequent budgetary decisions, which are assumed by the Parliament and the Executive. Unfortunately, the electoral cycle and instability in government make it difficult for politicians to know in depth the workings of bureaucracy and how to develop long-term cooperative strategies with public sector managers. Finally, bureaucracies must satisfy collective preferences that are dispersed, changing, and sometimes conflicting, (which is different from earning high profit aspirations of companies' stakeholders). That's why it becomes very difficult to assess the bureaucratic performance, while the private sector's managers are kept under control, even if in a very partial and imperfect way, looking at the value of stocks and at the market share (Tirole 1994).

All that explains why bureaucracies suffer from different forms (productive, economic, and industrial) of inefficiency. Many constitutional and administrative law provisions aim to reduce such inefficiency. Political directives are separated from every day management. The legal and economic treatment of civil servants is progressively becoming equal to that of private

employees. Artificial indicators of bureaucratic performance are created in order to keep under control bureaucratic behavior and to introduce performance-related pay systems also in the public sector (Rose-Ackerman 1986).

Strategies of Administrative Action

In order to allow administrative agencies and other public bodies to perform their tasks, administrative law provides them with different legal means of action. From this point of view, especially in the continental European tradition, the main distinction is between unilateral and bilateral means of action. In the first case, the law gives public bodies the power to adopt decisions, which produce binding effect on third parties, even if they weren't consented. Those unilateral decisions are considered an expression of the supremacy of public authorities over the citizen, due to their linkage to the State. In the second case, public bodies conclude contracts on the basis of their general capacity of private law. Given this general distinction, the legal scholarship describes the typology of the most important administrative decisions, according to their proper effect on the recipient, some of them enlarging the private sphere (grants, licensing, other permissions), others restricting it (orders, takings, fines, and administrative sanctions). A similar approach is followed in relation to contracts. Public bodies can make use of different procurement schemes and conclude other kinds of transactions.

The economic analysis of law offers the opportunity to go further. Means of action of public bodies are analyzed not only in a static way but in a dynamic way too, in order to catch the different strategies that rational (at least to a certain extent) public bodies and officials pursue. From this point of view, administrative law empowers the public bodies, leaving them the choice between different legal instruments and the proper design of their content. First of all (and from a very general point of view), in many cases, public bodies evaluate whether it is more convenient to adopt unilateral decisions (e.g., a taking) or to find an agreement with a private party (a sale). Such an evaluation can be made on the basis of the transaction costs' theory. If it is not too costly to find such an agreement, the consensual solution will satisfy both the public and the private interest, therefore avoiding any case for judicial review. On the contrary, when transaction costs are too high (e.g., because the private party holds a monopolistic position), the unilateral decision might be the most efficient solution.

Economic analysis provides a useful insight also in understanding strategies of acting through law making ("wholesale procedures") or through adjudicatory decisions ("retail procedures") (Cooter 2000, pp. 164–165). It gives precious suggestions about the proper design of licenses and concessions (time, price, contractual conditions). It's helpful in the fixing of the amount of compensation in the case of a taking and in the determination of the optimal level of administrative sanctions. Economic analysis reminds us the importance of the proper design of public contracts. The general interest can be satisfied only if the transaction is convenient for the private party too; otherwise the procurement procedure will fail or the private contractor will be unable to execute the contract.

Finally, game theory provides a useful insight in the evaluation of the interactions between public bodies and private parties, explaining success and failure of public policies (as in the case of the so-called simplification of the administrative procedures of licensing, often vanished by the absence of trust between agencies and private parties (Von Wangenheim 2004)).

The Regulation of Administrative Action

Administrative action is characterized by two distinctive features. On one side, public bodies, in executing the law, act as agents of different principals. On the other side, public bodies exercise a legal and economic power against private parties. The regulation of administrative action, then, pursues two purposes. The first is to ensure the fulfillment of the public interest, in coherence with the guidelines issued by the principals. The second is to protect the firms and the citizens

interacting with the public bodies in a subordinate position. From this double perspective, it becomes possible to understand the whole set of the legal provisions related to the administrative procedure, the exercise of discretionary powers, and contractual activity.

Many countries have enacted a legislation to regulate the administrative procedure in general (the USA in 1946 with the Administrative Procedure Act, Germany in 1976, Italy in 1990). In other countries, like France and the UK, relevant rules can be drawn from the principles stated by the courts and from sector by sector legal provisions. The regulation of administrative procedures, on one side, structures the decision-making process in the public interest. On the other side, it recognizes the rights of participation in favor of affected parties. Such participation is not only in the private interest. It's also a powerful tool of decentered "fire-alarm" control, in order to reduce informational asymmetry of political principals and to prevent bureaucratic drift (McCubbins et al. 1987).

The limitation of administrative discretion is very important too in order to avoid decisions contrary to the public interest or unduly negative for the private recipients. Legal provisions select the protected interests, fix the prerogatives of the acting bodies, and determine the criteria which must be followed in the exercise of power. However, a certain degree of flexibility is necessary to allow the administrative agencies to manage the specificity of every case and to face unpredictable circumstances (Cooter 2000, p. 91).

A detailed regulation is needed also when public bodies conclude contracts. On one side, the waste of public money must be prevented. On the other side, there is the risk of discrimination in the adjudication of contracts. That's why, international, European, and national regulations require fair and open tendering procedures to select the best offer. From this point of view, the way in which public bodies select their contractual partners is very similar to a marriage by mail. This explains the importance of a formal and detailed contract in which all future circumstances are covered by an *ex ante* agreement.

The efficiency of the rules devoted to the regulation of administrative action, anyhow, must be put under scrutiny. There are good reasons which explain the need for those rules and their growth in recent times. However, those regulations don't always prove to be really useful or to ever serve their intended purpose upon enactment. The final outcome is often an over-regulation of administrative action, which makes it excessively rigid and unfit to meet the quest for flexibility required to satisfy the public interest.

The Judicial Review and the Non-expert Control Paradox

The regulatory system of administrative law in every country is completed by the existence of a judicial review against the decisions adopted by public bodies. Private parties, which are affected by an administrative decision, can go before a court to obtain the declaration of invalidity of that decision, if retained as illegitimate. In this way, while protecting their private interests, they launch "fire-alarm" signals to allow a third party independent oversight of the bureaucratic behavior.

Courts play a very important role in controlling the respect of the legal provisions that limit administrative discretion. They also check the exercise of effective investigations on facts and interests and the disclosure of all the information at the basis of the administrative decision. Finally, they enforce the respect of all the procedural requirements that allow the participation and the external control of administrative action by private parties (Bishop 1990).

Different models of judicial review and of control on the proper behavior of bureaucracy exist in the world (Josselin and Marciano 2005). Common law countries give to ordinary courts the competence to review administrative decisions. Many continental European countries, on the contrary, established special administrative courts, following the French model of the *Conseil d'Etat*. Administrative courts are supposed to be more sensible to the public interest and more generous

in recognizing and protecting power and prerogatives of the bureaucracy. However, differences among ordinary and administrative courts are greatly reduced over time. Special benches of ordinary courts were established, e.g., in the UK, in order to address administrative law cases. Administrative courts, at the same time, became much more effective in protecting private interests.

The judicial review of administrative action by courts can appear a paradox. Judges, who have only a legal training, are asked to check the decisions adopted by expert bureaucrats. This way, the system of judicial review runs the risk of frustrating the very reason of delegation to public administration, which is made exactly in name of an efficient division of labor to highly specialized and expert bodies. The existence of specific administrative courts, different form ordinary courts, however, reduces the danger of scarce preparation. By focusing their attention on disputes between agencies and citizens, they can develop a specific knowledge of the different fields of administrative action. Moreover, administrative judges, especially in France and Italy, act as consultant of public bodies and are often appointed as heads of staff in the Cabinet. This way they can reach a deeper comprehension of the way in which bureaucracies work.

In any case, the judicial review can ameliorate the quality of administrative decisions for at least three reasons. First, procedural rules before court would more likely guarantee the due process than administrative procedure rules in isolation would do. This way, courts may have access to a greater deal of information than that available to bureaucracy at the moment of the administrative decision. Second, due to self-selection made by private parties, the judicial review can be focused on really suspicious cases rather than having to check every administrative decision. And third, public agencies have an incentive to make *ex ante* investments in the preparation of the administrative decisions, through a more detailed evaluation of facts and interests, and in legal protection, in order to prevent the judicial review made by courts (Von Wangenheim 2005).

References

Barzel Y (2002) A theory of the state. Economic rights, legal rights, and the scope of the state. Cambridge University Press, Cambridge

Bawn K (1997) Choosing strategies to control the bureaucracy: statutory constraints, oversight, and the committee system. J Law Econ Organ 13(1):101–126

Bishop W (1990) A theory of administrative law. J Legal Stud XIX(2):489–530

Cooter RD (2000) The strategic constitution. Princeton University Press, Princeton

Farber DA, e Frickey PP (1991) Law and public choice. A critical introduction. Chicago University Press, Chicago

Ferejohn J, Shipan C (1990) Congressional influence on bureaucracy. J Law Econ Organ 6:1–20

Holburn GLF, Vanden Bergh R (2006) Consumer capture of regulatory institutions: the creation of public utility consumer advocates in the United States. Public Choice 126:45–73

Horn MJ (1995) The political economy of public administration. Cambridge University Press, Cambridge

Josselin JM, Marciano A (2005) Administrative law and economics. In: Backhaus JG (ed) The Elgar companion to law and economics, IIth edn. Edward Elgar, Cheltenham, pp 239–245

Kagan E (2001) Presidential administration. Harv Law Rev 114:2245–2385

Macey J (1992) Organizational design and the political control of administrative agencies. J Law Econ Organ 8(1):93–110

Mashaw JL (1985) Prodelegation: why administrators should make political decisions. J Law Econ Organ 1(1):81–100

McCubbins M, Schwartz T (1984) Congressional oversight overlooked: police patrols versus fire alarms. Am J Polit Sci 28(1):165–179

McCubbins M, Noll R, Weingast B (1987) Administrative procedures as instruments of political control. J Law Econ Organ 3(2):243–277

McCubbins M, Noll R, Weingast B (2007) The political economy of law. In: Polinsky AM, Shavell S (eds) Handbook of law and economics, IIth edn. Elsevier, Amsterdam, pp 1651–1738

Mueller DC (1996) Constitutional democracy. Oxford University Press, New York/Oxford

Napolitano G (2014) Conflicts and strategies in administrative law. Int J Constit Law 12(2):357–369

Posner RA (1997) The rise and fall of administrative law. Chicago-Kent Law Rev 72:953–963

Posner EA (2001) Controlling agencies with cost-benefit analysis: a positive political theory perspective. Univ Chicago Law Rev 68(4):1137–1199

Rose-Ackerman S (1986) Reforming public bureaucracy through economic incentives? J Law Econ Organ 2:131–161

Rose-Ackerman S (2007) Introduction. In: Ead (a cura di) The economics of administrative law. Edward Elgar, Cheltenham, pp XIII–XXVIII

Tirole J (1994) The internal organization of government. Oxf Econ Pap 46(1):1–29
Ulen TS (2004) The unexpected guest: law and economics, law and other cognate disciplines, and the future of legal scholarship. Chicago-Kent Law Rev 79(2):403–429
von Wangenheim G (2004) Games and public administration. The law and economics of regulation and licensing. Edward Elgar, Cheltenham
von Wangenheim G (2005) Should non-expert courts control expert administrations? In: Josselin J-M, e Marciano A (eds) Law and the state. A political economy approach. Elgar, Cheltenham/Northampton, pp 310–332

Administrative Procedure

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

Abstract
This chapter examines the economics of administrative procedure and shows how law and economics insights can be used to understand fundamental features of administrative procedure. Moreover, it offers a survey of law and economics literature on administrative procedure and addresses three general aspects of administrative procedure. The first is the administrative decisions and access to information. The second general aspect is administrative procedure and agency capture problem. The third general aspect addressed in this chapter is the issue of administrative procedure and international trade.

Definition

Administrative procedure is a set or system of rules that govern the procedures of public bodies for managing organizing bureaucratic actions inside bureaucracies and in their interactions with private parties. These procedures are generally meant to establish efficiency, consistency, responsibility, and accountability.

Introduction

In all industrialized societies, there is a tension between market and collectivist system of economics organization. The latter one is the system where state seeks to direct or encourage behavior which would not occur without such an intervention. The aim of such an intervention is to correct perceived or real market failures in meeting collective or public interest goals. Such goals are generally pursued by a plethora of executive and legislative branches of governments. In this respect, modern law and economics analysis of public law divides into two strings of work. One dealing with causes and consequences of bureaucratic action inside bureaucracies and the other focusing on external interaction, such as between legislators and the bureaucratic institutions comprising executive branch, between the latter institutions and the citizens and enterprises (Weigel 2006). The latter field of research is in law and economics known as the issue of regulation (Ogus 2004). Ogus (2004) actually argues that civil legal cultures have specific concepts and instruments to govern those "external relationships," for example, German law operates with "*Wirtschaftsverwaltungrecht*" (literally: "Administrative Law of the Economy") and the French with "*Droit Public Economique*," whereas English law employs imprecise expressions, such as "regulation" and "regulatory law." Although perplexing problems and issues of regulation have triggered some attention of law and economics scholars (Weigel 2003; Ogus 2004; Rose-Ackerman 1995; Schuck 1994; Johnson and Libcap 1994; Rose-Ackerman and Lindseth 2010; Lewish and Parker 2017) and despite its great expansion to nonmarket fields, the proper interpretation of administrative procedure has received relatively little law and economics analysis in recent years. One may even argue that law and economics are still an unexpected guest in the field of administrative law (Ulen 2004: Napolitano 2014). Napolitano (2014) shows that indeed the law and economics literature has traditionally paid very little attention to administrative law, history, and rules. While civil and criminal procedures have been thoroughly addressed and produced a striking amount of

literature, the assessment of public/administrative law actually reveals surprising gaps in the law and economics literature of administrative procedure.

Moreover, the law and economics literature has generally paid very little interest toward procedural issues, and the bulk of work done in that field contains hardly any economics (Schuck 1994). Weigel (2006), for example, stresses that general administrative procedure acts (which play central roles in civil law systems) have actually rarely been economically thoroughly addressed or even investigated whether those essential systems of procedural rules are efficient tools for the promotion of effective, wealth-maximizing outcomes.

Identified lack might indeed be contributed to the limited confidence that the law and economics scholars might have with highly technical and vastly detailed provisions of administrative procedure. Furthermore, Rose-Ackerman (2007) points out that administrative law scholars represent a close community, usually unwilling to open their minds to methodologies different from legal positivism. This isolation might also explain why the capacity of law and economics to shed new light into the field of administrative procedure has been very limited. Hence, this lack of literature represents a missed opportunity both for the law and economics and for the traditional legal scholarship of administrative procedural law (Ogus 1998; Rose-Ackerman 2007).

However, the limited amount of literature might be contrasted with a Bishop's (1990) theory that the essential function of administrative law and of administrative procedure is thought to come to grips with economic slack created by bureaucrats. Moreover, there is, for example, a vast law and economics literature on production of legal rules by agencies and bureaucracies (von Wangenheim 2004).

Administrative Decisions and Access to Information

Classical law and economics literature conceives legal procedure as an institution designed to minimize the sum of "error costs" (the social costs generated when a judicial system fails to carry out the allocative or other social functions assigned to it) and "direct costs" (such as lawyers', judges', and litigants' time) of operating legal dispute resolution machinery (Posner 1973). According to this classical law and economics framework, the rules and other features of any procedural system can be analyzed as efforts to maximize efficiency (Posner 1973). In line with this observation, Posner (1973) actually already, as part of his case studies, addresses the deterrence issues of administrative sanctions, argues against any judicial review of agency fact-finding, and suggests that the characteristic combination of prosecution and adjudication in the same agency may be a source of inefficiency. Stewart (1975) supports this information disclosure function by arguing that administrative decision should be made transparent through requirements for public comment and open meetings and that administrative procedures should give citizens and organized interests the ability to represent their views in the administrative process. This proposition is in line with Stigler's (1971) path-breaking insights that open procedures are thought to foster pluralist politics that protect against regulatory capture, the danger that an industry will come to control an agency's decision-making to secure private benefits.

Posner (2011) also argues that administrative procedure will be less consistent over time than judicial one and also stresses that precedents will play a smaller role in administrative than in judicial decision-making. He shows that a heralded innovation of the administrative procedure in the USA was the administrative agency's looseness structure where agency is able to issue rules, bring cases, decide cases, conduct studies, issue advice, and even propose a legislation (Posner 2011).

However, one of the first field-specific law and economics contributions in the field of administrative procedure was Rose-Ackerman's seminal article on "The Progressive Law and Economics-And the New Administrative" on the lowering information cost function of modern administrative procedural laws. Rose-Ackerman (1988) suggests that if administrative procedures lower the cost of access to information about administrative

decisions, then interested parties will be more likely to be able to find out about an upcoming decision. The more time an interested party has to mobilize financial and political resources against a particular decision, the more likely it will be able to force the administrative decision-maker, through his or her more politically accountable superiors, to influence that decision. The impact on substantive policy comes from the greater benefit to less organized interest groups of lowering this cost of access to information (Rose-Ackerman 1988).

Cooter (2002) employs game theoretical tools and shows, while assessing the US Administrative Procedure Act discusses the court-agency relationships. He provides evidence that the relatively high transaction costs of formal decisions increase the agency's discretionary power over a class of decisions, whereas the relatively low transaction costs of informal decisions reduce the agency's discretionary power (Cooter 2002). Furthermore the court will have less concern over the agency's discretionary power when agency preferences are close to court preferences. Thus, Cooter's model predicts the courts will favor formal decision-making when agency preferences are close to court preferences (Cooter 2002).

In addition, Verkuil (1978) analyzed the emerging concept of the US administrative procedure and argued that administrative procedure should be concerned with the overall fairness and accuracy of decisions, with their efficient and low-cost resolution and, in a democratic society, with participant satisfaction with the process.

Administrative Procedure and Agency Capture Problem

One of the central problems of the representative democracy is how to ensure that decisions are responsive to the interests or preferences of citizens. In the 1980s scholars began to utilize law and economics insights to study public law and examined how rules and administrative procedures shape policy and evaluated these outcomes from the perspective of economics efficiency (McNollgast 2007). This stream of research conceptualizes the decision-making procedures of government as rationally designed by elected official to shape policies arising from decisions by executive agencies and to align incentives of agencies with the ones of citizens. McCubbins et al. (1987, 1989) study argue that administrative procedure is actually designed to solve the agency problem between citizens and agencies and that the mechanics of fire-alarm oversight are imbedded in the administrative procedures of agencies that are within the jurisdiction of the US Administrative Procedure Act. They actually see administrative procedures as another mechanism to control bureaucracy as kind of a check-and-balance system that prevents the opportunism and moral hazard of bureaucratic system and its institutions. For example, all of the procedural features of the US Administrative Procedure Act economically speaking actually reduce agency's information advantage (McNollgast 2007). McNollgast also points out that congress employs administrative procedure to delegate some monitoring responsibility to those who have standing before an agency and who have a sufficient stake in its decisions to participate in its decision-making process and to trigger oversight by pulling the fire alarm (McNollgast 2007). In addition, administrative procedures create a basis for judicial review that can restore the status quo without legislative correction, and as a result administrative procedures may cope with the first-mover advantage of agencies (McNollgast 2007). He also states that administrative procedures might facilitate the political control of agencies in five different ways:

(a) administrative procedures ensure that agencies cannot secretly conspire against elected official;
(b) agencies must solicit valuable information;
(c) the administrative proceedings are public, thereby enabling political principals to identify not just potential winners or losers of the policy but also their views;
(d) the sequence of procedure (notice, comment, deliberation, collections of evidence, and construction of records) enables elected representatives to respond when administrative

agencies seek to move toward a goal that is not aligned with citizens ones; and

(e) it serves to indicate the stakes of a group in an administrative proceeding (McNollgast 2007).

Moreover, McCubbins et al. (1987) argue that procedural requirements also affect the institutional environment in which agencies make decisions and thereby limit an agency's range of feasible policy actions. In recognition of this, elected officials may design procedures to solve two prototypical problems of political control. First, procedures can be used to mitigate informational disadvantages faced by politicians in dealing with agencies (McCubbins et al. 1987). Second, procedures can be used to enfranchise important constituents in agency decision-making processes, thereby assuring that agencies are responsive to their interests (McCubbins et al. 1987). The most subtle and the most interesting aspect of procedural controls is according to McCubbins et al. (1987) the possibility to assure administrative agency's compliance without specifying, or even necessarily knowing, what substantive outcome is most in their interest. Namely, by controlling processes, political leaders assign relative degrees of importance to the constituents whose interests are at stake in an administrative proceeding and thereby channel an agency's decisions toward the substantive outcomes that are in the interest of those who are intended to be benefited by the policy. Hence, political leaders can be responsive to their constituencies without knowing the details of the policy outcomes that these constituents want (McCubbins et al. 1987, 1989). Rose-Ackerman (2007) supports this "fire-arm" proposition and adds that even though the constraints are nominally procedural, they have also substantive effects. Namely, instead of requiring close legislative control, the required procedures assure the enacting coalition that the agency will respond to changes in the preferences of those interests served by the legislation (Rose-Ackerman 2007).

Contradicting McCubbins et al. (1987), Moe argues that the role administrative procedures play must be understood in a dynamic context (Moe 1989). According to Moe's argument, these procedures create a "lock-in" effect – constraining both future politicians and their agents – and therefore act as a mechanism whereby current majorities can ensure, at a cost, that future majorities will not upset their policy intentions (Moe 1989).

Recently, De Figueiredo and Vanden Bergh (2004) provide the first empirical investigation of previously discussed theories. They empirically, by employing an administrative procedures enactment dataset for the period 1941–1981, test whether organizational procedures that are enacted by public officials have any impact on the nature of both bureaucratic control and performance. They have also examined the set of conditions under which the US federal states adopt administrative procedures and found that the likelihood of adopting administrative procedural acts are increased when (a) democratic legislative supermajorities (veto-proof majority) face a Republican governor and (b) when democratic control is perceived to be temporary and when they fear the future loss of power (De Figueiredo and Vanden Bergh 2004). Their results indicate that existing theories emphasizing agency and dynamic effects are empirically valid, albeit with an important qualification: there is a distinctive partisan bias in the usefulness of administrative procedures for these purposes (De Figueiredo and Vanden Bergh 2004). De Figueiredo and Vanden Bergh (2004) also show that despite a dual set of conditions under which they might be adopted, administrative procedure acts seem to have a markedly Democratic bias.

Administrative Procedure and International Trade

Another stream of law and economics research focuses on the importance and on the impact of administrative procedural law on trade policy. Rosenbaum (1998), for example, argues that administrative procedures that lower the cost of obtaining information about and notice of agency decision-making and that therefore increase the amount of time between an interest group's

learning of a potential decision and the final administrative decision shift the balance of power between more and less organized groups. Namely, by giving less organized interest groups a greater opportunity to mobilize political resources to influence bureaucrats through the politicians who oversee them, administrative procedures might shift the balance of power between constituencies (Rosenbaum 1998). While more organized groups may have the political resources, connections, money, and staff – to learn of agency decision-making without procedural constraints – less organized groups will be able to benefit from the increased possibility of notice and opportunity to mobilize. Therefore, by lowering the cost of access to information about administrative decisions, and by thus shifting the balance of power among constituencies from more to less organized groups, administrative procedural law gives according to Rosenbaum (1998) consumers greater power in determining trade policy. Hence, since consumers of a given good are interested in lower trade barriers while producers are interested in greater protection, this shift of power will lower trade barriers (Rosenbaum 1998).

Concluding Remarks

While civil and criminal procedures have been thoroughly addressed and produced a vast amount of law and economics literature, the assessment of public/administrative law actually reveals surprising gaps. Identified gap in the law and economics of administrative procedure calls for further empirical, behavioral, experimental, and theoretical research that would shed additional light upon the most triggering questions and would provide additional insights, normative policy implications for legislators in improved daily decision-making and lawmaking. Coglianese (2002) actually suggests that empirical research on administrative law has the potential for evaluating and ultimately improving prescriptive efforts to design administrative procedures in ways that contribute to more effective and legitimate governance. Hence, identified missed opportunity for the law and economics scholarship of administrative procedural law opens doors for an extensive law and economics treatment that would offer sets of legal and economics arguments for an improved regulatory response.

References

Bishop W (1990) A theory of administrative law. J Leg Stud 19:489–530

Coglianese C (2002) Empirical analysis and administrative law. University of Illinois Law Review. University of Illinois, Urbana Champaign, pp 1111–1137

Cooter R (2002) The strategic constitution. Princeton University Press, Princeton

De Figueiredo RJP Jr, Vanden Bergh RG (2004) The political economy of state-level administrative procedure acts. 47. J Law Econ 2:569–588

Johnson RN, Libcap G (1994) The federal civil service system and the problem of bureaucracy. University of Chicago Press, Chicago

Lewish Parker JS (forthcoming 2017) Procedure in American and European law: a general economic analysis. In: Hatzis AN (ed) Economic analysis of law: a European perspective. Edward Elgar, Cheltenham

McCubbins MD, Noll RG, Weingast BR (1987) Administrative procedures as instruments of political control. J Law Econ Org 3:243–277

McCubbins MD, Noll RG, Weingast BR (1989) Structure and process, politics and policy: administrative arrangements and the political control of agencies. Virginia Law Rev 75:431–482

McNollgast (2007) The political economy of law. In: Polinsky AM, Shavell S (eds) Handbook of law and economics. North-Holland, Amsterdam, pp 1707–1715

Moe TM (1989) The politics of bureaucratic structure. In: Chubb JE, Peterson PE (eds) Can the Government Govern? Brookings Institution Press, Washington, DC, pp 267–330

Napolitano G (2014) Administrative law. In: Marciano A, Ramello GB (eds) Encyclopedia of law and economics. Springer, New York, pp 1–8

Ogus AI (1998) Regulatory appraisal: a neglected opportunity for law and economics. Eur J Law Econ 6:53–68

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

Posner RA (1973) An economic approach to legal procedure and judicial administration. J Leg Stud 2:399–458

Posner AR (2011) Economic analysis of law. Kluwer, Austin

Rose-Ackerman S (1988) The progressive law and economics-and the new administrative law. 98. Yale Law J 98:341–368

Rose-Ackerman S (1995) Controlling environmental policy. Yale University Press, New Haven

Rose-Ackerman (2007) Public choice, public law and public policy. Keynote address, First World Meeting of the Public Choice Society Amsterdam, Mimeo

Rose-Ackerman S, Lindseth PL (2010) Comparative administrative law. Edward Elgar, Cheltenham
Rosenbaum RD (1998) Domestic bureaucracies and the international trade regime: the law and economics of administrative law and administratively imposed trade barriers. Harvard Law School Discussion Paper.
Schuck PH (1994) Foundations of administrative law. Foundation Press, New York
Stewart RB (1975) The reformation of administrative law. Harv Law Rev 88:1669–1813
Stigler GJ (1971) The theory of economic regulation. Bell J Econ Mgt Sci 2:3–21
Ulen TS (2004) The unexpected guest: law and economics, law and other cognate disciplines, and the future of legal scholarship. Chicago-Kent Law Rev 79(2):403–429
Verkuil PR (1978) The emerging concept of administrative procedure. Columbia Law Rev 78(2):258–329
Wangenheim GV (2004) Procedural rules and administrative decisions. In: Encyclopedia of law and economics. Edward Elgar, Cheltenham
Weigel W (2003) Rechtsökonomik. Vahlen Verlag, München
Weigel W (2006) Why promote the economic analysis of public law? Homo Econ 23(2):195–216

Adoption

Martha M. Ertman
Francis King Carey School of Law, University of Maryland, Baltimore, MD, USA

Abstract

Historical and recent developments in legal economic analysis of adoption the United States reveal changing supply and demand of children and the emergence of submarkets for adoption through agencies (public or private) and independent adoptions. Curent legal rules against baby-selling and adoption agency practices mask the existence of adoption markets by banning payments to birth parents yet exempting payments to agencies and other adoption professionals. Economic proposals seek to narrow gaps between supply and demand by creating incentives (or removing disincentives) through substitutes, subsidies, and reduced transactions costs. These solutions could prevent a good number of would-be parents from remaining childless, provide homes for many children who currently languish in foster care or group homes, and better recognize the costs that adoption imposes on birth parents.

Synonyms

Children; Family; Parenthood

Introduction

This entry tracks historical and recent developments in legal economic analysis of adoption. Focusing on the United States, it first traces historical market changes as evidenced by changing supply and demand of children and the emergence of submarkets for adoption through agencies (public or private) and independent adoptions. It then identifies current legal rules against baby-selling and adoption agency practices that mask the existence of adoption markets by banning payments to birth parents yet exempting from that ban payments to agencies and other adoption professionals. It concludes by describing economic proposals to narrow current gaps between supply and demand by creating incentives (or removing disincentives) through substitutes, subsidies, and reduced transactions costs. These solutions, if implemented, could prevent a good number of would-be parents from remaining childless, provide homes for many children who currently languish in foster care or group homes, and better recognize the costs that adoption imposes on birth parents.

Adoption Markets

History

The laws of supply and demand have long governed adoption. In the nineteenth century, norms against unwed motherhood and the lack of a social safety net made adoption a buyer's market. The "price" of children whose parents could not raise them was so low that many birth mothers had to pay people known as "baby

farmers" to take the children off their hands. In the custody of the baby farmers, that low value translated to low survival rates. Children lucky enough to survive were put to work, getting hired out for farm or domestic work. Gradually, adoptive parents became willing to pay to adopt. In the late nineteenth and early twentieth century, newspaper classified advertisements commonly listed children for adoption along with a price (Zelizer 1985; Herman 2008).

Gradually, adoption law and practice came to reject market rhetoric as society and legal rules began to view childhood as a vulnerable period spent in school instead of laboring in fields or factories. In 1851, Massachusetts passed the first adoption statute, which required that a child's interests be accounted for in an adoption. Other states followed suit, eventually establishing standards for determining a child's adoptability and adoptive parents' suitability. Gradually, states completely replaced the birth family with the adoptive family courtesy of new birth certificates and sealed records. Along the way, agencies and social workers acquired a nearly exclusive gatekeeping function over both the adoption and information passing between birth and adoptive families. Limits on child labor and increased life spans due to industrialization and medical advances all contributed to these social and economic changes that justified higher investments in children's human capital. Technologically, the advent of safe, affordable infant formula in the 1920s and 1930s further encouraged couples to adopt. By 1937, a magazine article could exclaim, "The baby market is booming . . . the clamor is for babies, more babies . . . We behold an amazing phenomenon: a country-wide scramble on the part of childless couples to adopt a child" (Zelizer 1985).

Adoption rates continued to rise until the early 1970s, when several phenomena contributed to both a rapid drop in the supply of infants available for adoption and increase in demand from new sources. On the supply side, more unmarried mothers could collect child support from the children's fathers once the Supreme Court struck down as unconstitutional rules that relieved nonmarital fathers of financial responsibilities. Poverty was less likely to force a woman to relinquish her child for adoption as federal welfare programs expanded to cover more poor, often unmarried, women and their children. Rises in nonmarital births and single parenthood after divorce decreased the stigma of single motherhood.

Some changes in legal rules and cultural practices influenced both supply and demand. As women could get effective birth control and legal abortions, they became less likely to conceive or carry an unplanned pregnancy to term. That increased control made motherhood more of a choice than it had ever been, allowing some women to take full advantage of newly expanded opportunities to develop their human capital by pursuing a career. On the supply side, that increased access to well-remunerated work meant that more single women who did bear children could keep them. On the demand side, many women who had delayed conception while they established careers found their fertility diminished when they got around to having children.

Consequently, in 1972, nearly 20% of white single mothers placed their babies in adoptive homes, but by 1995, that rate decreased to 1.5% and has since dipped to about 1%. Black mothers, in contrast, have never relinquished in high numbers. Even before cases like *Griswold v. Connecticut* and *Roe v. Wade* expanded reproductive choices by decriminalizing birth control and abortion, only about two percent of black single mothers surrendered their infants for adoption. That low rate was due to many agencies' unwillingness to place African-American babies and the willingness of extended kinship networks in the black community to take in those children. Since the 1970s black women's relinquishment rate has dipped to nearly zero. By the late 1980s, this low supply of healthy infants translated to a hundred would-be adoptive parents competing to adopt each healthy infant available for adoption, according to the National Council for Adoption (Joyce 2013).

Agencies and adoptive parents continue to express preferences for some children over others. Prior to World War II, most agencies considered children of color and those with disabilities unfit

for adoption, relegating them to institutions. While more agencies began to accept these children in the 1950s – increasing the supply of adoptable infants – many would-be adoptive parents retained their preference for white children. In the early twenty-first century, most would-be adoptive parents – a majority of whom are white – continue to prefer to adopt white over African-American children, though adoptive parents who are single or gay are more willing to create an interracial family through adoption (Baccara and Collard-Wexler 2012).

Adoption Markets Today

Today, adoption continues to function as a market, though rhetoric of both law and adoption professionals persists in masking its market characteristics. About 2.5% of American children are adopted, 1.6 million in 2000. Adoption professionals – agencies, attorneys, and social workers – generate fees of $2–3 billion annually (Baccara 2010). Agencies refer to adoption as a "gift" instead of an exchange, and statutes criminalize or otherwise prohibit payment of fees in connection with an adoption. But those statutes also exempt fees paid to professionals, and states enforce the statutes largely to penalize birth parents receiving any payment. Some states permit adoptive parents to pay for a birth mother's maternity clothes, psychotherapy, and living expenses, while others prohibit those expenditures or impose strict time or monetary limits. A black or gray market persists in the shadow of these prohibitions, with adoptive parents or agencies paying for birth mothers' expenses (Ertman 2003, 2015).

The legal adoption market is fragmented. A child could be placed through a private agency, through a public child welfare agency ("foster care"), or without agency involvement ("independent adoption"). An adoption may also be subject to rules imposed by a religiously affiliated agency, may involve relatives or nonrelatives, and could be domestic or international. While accurate data on adoption is scarce and rates vary over time, as of 2002, about 84% were domestic – evenly divided among familial and nonfamilial – with 16% international (which are nearly always nonfamilial). That same year, over half of the children adopted by nonrelatives were adopted out of foster care. Many of the children adopted out of foster care are older; suffer from physical, mental, and emotional disabilities; and/or are children of color (Bernal et al. 2007; Moriguchi 2012). This entry focuses on three categories that exhibit different market patterns: unrelated domestic adoption, adoption out of foster care, and independent adoption.

Within the first category – unrelated domestic adoptions – demand for healthy white infants far outstrips supply (Landes and Posner 1978). Consequently, adoptive parents commonly pay agencies higher fees to adopt those babies than to adopt children whom adoption professionals describe as "hard to place." (Ertman 2015). Another common fee structure charges adoptive parents on a sliding scale, with higher-income adopters paying higher fees. Wealth and income gaps between whites and African-Americans, coupled with most adoptive parents' preference for a child of their own race, combine to make it more expensive in general to adopt a healthy white infant than to adopt children who are older, of color, and/or disabled.

Adoptions in the second category – foster care – exhibit the opposite pattern: high supply of children and low demand by adoptive parents, generating what Elisabeth Landes and Richard Posner called a "glut" of children in foster care, a situation that they compared to "unsold inventory stored in a warehouse" (Landes and Posner 1978). Consequently, children in foster care may wait 2 years for adoption and get shuttled between 10 and 20 homes during those years. Only around ten percent – 50,000 – get adopted into permanent homes each year, and African-American children can wait twice as long as white kids to get adopted out of foster care (Beam 2013).

In the third category – independent adoption – comprehensive data is even more scarce than in agency adoptions. Many independent adoptions are stepparent adoptions – as when a divorced woman's new spouse adopts a child from the prior marriage – which do not require an intermediary to match the children to their new families nor a study of suitability of adoptive families. The fees are largely for lawyers, rather than agencies, and do not depend on the

characteristics of either parent or child. Some monetary payments are not legally binding, but nevertheless occur with regularity. A noncustodial birth father often demands a "price" for consenting to the adoption in the form of the mother agreeing not to pursue him for unpaid child support (Hollinger 2004). Another common payment – this one entirely nonmonetary – involves the birth parent retaining visitation rights in an agreement known as a Post Adoption Contact Agreement. These agreements are legally binding in about half the states (Ertman 2015).

Proposals to Narrow Gaps Between Supply and Demand

The market for adoption evidences inefficiencies. At least since the 1970s, supply and demand for two major types of adoption have been mismatched, with high demand coupled with low supply of healthy white infants and low demand coupled with high supply of children in foster care. Consequently, adoptive parents who are only willing to adopt a healthy, white infant may well remain childless, while as many as 125,000 adoptable children languish in foster care. At the fiscal level, adoption is less expensive than foster care. Between 1983 and 1986, adoption decreased government expenditures on foster care by $1.6 billion. Non-fiscal but nevertheless valuable benefits of adoption over foster care include the better health, behavioral, educational, and employment outcomes of children who are adopted over those who remain in foster care (Hansen 2008).

Foundational Work by Becker, Landes, and Posner

Economic proposals have sought to increase the supply for in-demand babies and increase the demand for hard-to-place children. Nobel Laureate Gary Becker's 1981 book *A Treatise on the Family* established the relevance of economic tools to understand family forms. On the supply side of adoption, he suggested that birth parents are more likely to put what he calls "inferior" children "up for sale or adoption." On the demand side, he postulated a "taste for own children, which is no less (and no more) profound than postulating a taste for good foods or any other commodity entering utility function" and that women are reluctant to commit "effort, emotion and risk" to children "without considerable control over rearing" as well as information ahead of time about the children's "intrinsic characteristics" (Becker 1981). Legal economists Elisabeth Landes and Richard Posner further developed the theoretical model of adoption as a market in their highly influential 1978 article "The Economics of the Baby Shortage." They identified the "potential gains in trade from transferring the custody of the child to a new set of parents" and cataloged the pros and cons of a "free baby market" in comparison to a likely black market that flourishes in part because of legal constraints on payments in adoption. Their proposal was quite modest. As "tentative and reversible steps toward a free baby market," they proposed that agencies use fees charged to higher-income adoptive parents to make "side payments" to pregnant women to induce them to relinquish the child for adoption instead of terminating the pregnancy (Landes and Posner 1978). Many readers mistook this proposal for an open market in children – akin to slavery – and criticized its tendency to treat children as commodities (Radin 1996; Williams 1995), despite Posner's clarification about the proposal's narrowness (Posner 1987).

Recent Applications

Recent market proposals further develop the idea of openly marketizing adoption.

Increase Supply of In-Demand Infants

Legal reforms aimed to increase the supply of the most sought-after babies who are available for adoption would focus on birth parents – especially birth mothers – because they are the ones who generally make the decision whether to keep a child or place him or her for adoption. Legal economists have argued that legal rules and adoption professionals could make birth parents' substitutes for adoption more expensive, use subsidies to encourage relinquishment, and lower transactions costs.

Make Birth Parents' Substitutes More Expensive Despite the apparent causal relationship between legal decisions granting women rights to birth control and abortion and plummeting number of infants available for adoption, empirical research has yielded mixed results regarding the effect of reproductive freedoms on adoption. One study found that states that legalized abortion before *Roe v. Wade* also had a 34–37% decrease in adoption of unrelated white children (Bitler and Zavodny 2002). However, other studies either found no statistically significant relationship between adoption and the price and availability of abortion (Medoff 1993) or a causal relationship between adoption and abortion availability but also found an unexpected effect of restrictive abortion laws decreasing the number of unwanted births (Gentenian 1999).

One proposal would impose barriers to abortion to increase the supply of healthy, white infants by making adoption a "two-sided market clearing institution" that remedies information asymmetries. It would increase the supply of sought-after babies through two legal changes: (1) allowing payments to birth mothers and (2) requiring women considering an abortion to hear a pitch from an adoption provider about the "economic incentives" should the birth mother "produce a child for the market" (Balding 2010).

Subsidize by Decreasing Birth Parents' Disincentives to Relinquish Birth parents, like other parents, generally prefer to raise the children they bear rather than surrender them for adoption. Legal economic scholars since the 1970s have proposed lifting the ban on payments to birth parents to help overcome the disincentive to relinquish. One scholar proposes that "baby market suppliers" – birth parents – "share in the full profits generated by their reproductive labor" (Krawiec 2009). Other scholars propose indirect payments, such as adoptive parents paying a birth mother's legal expenses to ensure she fully understands her rights or paying the birth mother's college tuition or job training expenses after the placement. These payments would have both monetary and nonmonetary benefits. Tuition payments in particular would reduce the chance of the birth mother having to relinquish another child since her education and training should improve her ability to support any child she gives birth to in the future (Hasday 2005).

Birth parents may deem one type of nonmonetary exchanges as more valuable than any monetary payments since it reduces somewhat the emotional cost borne by birth parents who lose all contact with their children. As adoption rates plummeted in the 1970s, that decreased supply and increased demand enabled birth mothers to exercise more bargaining power. Agencies began to let the birth mothers select the adoptive parents and also negotiate with prospective adoptive parents to get them to make promises to raise the child in a particular way (i.e., Catholic or with music education) and to provide the birth mother with periodic letters, pictures, email contact, and even in-person visits as the child grows up. These latter agreements, known as Post Adoption Contact Agreements, are increasingly common and increasingly enforced by courts (Ertman 2015).

Lower Transactions Costs Other scholars seek to increase adoptions by decreasing transactions costs. One scholar critiques the uncertainty created by statutes that allow a long period for birth parents to revoke their consent to an adoption on the grounds that courts could determine voluntariness of that consent at an early point, creating certainty for a child's place in his or her new family (Brinig 1994).

Increase Demand for Hard-to-Place Children

Legal reforms aimed to increase the demand for children in foster care focus on adoptive parents because they are the ones who decide whether and whom to adopt. Legal economists have argued for making adoptive parents' substitutes for adopting a foster care child more expensive, subsidizing foster care adoptions to reduce the price of adopting a special needs child, and lowering transactions costs.

Make Adoptive Parents' Substitutes More Expensive

Many, if not most, adoptive parents turn to adoption only after infertility treatments fail. Since the 1990s, reproductive technology techniques – especially in vitro fertilization (IVF) – have improved so that they more effectively help women become pregnant. In that same period, rates of domestic adoption have continued to decline. Accordingly, economists have asked whether IVF is a substitute for adoption. One study found that between 1999 and 2006, a 10% increase in adoption correlated with a 1.3–1.5% decrease in the number of IVF cycles performed. The correlation was higher when the researchers focused on infant adoptions, older adoptive mothers, and international adoption. Unsurprisingly, since adoption by relatives occurs in specialized circumstances such as stepparent adoption, prevalence of IVF does not correlate with those adoptions (Gumus and Lee 2012).

Legal reforms could make IVF more expensive to pull more adoptive parents toward children in foster care. A state legislature could remove subsidies to IVF such as state-mandated insurance coverage for the expensive procedure or require a formal legal process akin to adoption in IVF procedures using donor eggs (Appleton 2004).

Some scholars contend that race matching reduces demand for foster care children, since many adoptive parents are white and most children in foster care are either African-American or Latino. If the race of one child substituted for the race of another, the reasoning goes, then federal laws and adoption policies that prohibit delaying or denying an adoption because of the child or adoptive parents' race help the market clear faster. Yet empirical research indicates that race matching does not reduce the number of adoptions (Hansen and Hansen 2006).

Subsidize Adoption out of Foster Care

Subsidies have been among the most effective tools to spur demand for children in foster care. Because many prospective adoptive parents cannot afford the $30,000 or more required to adopt an infant through a private agency, adopting a child through a public agency may be the only available path to parenthood. Since the 1970s the federal government has enacted subsidies to encourage these adoptions, and empirical research has demonstrated the positive and statistically significant effect of subsidies on the rate of those adoptions (Hansen 2007).

Inefficiencies persist, however. Subsidies that continue after the adoption offset the costs borne by adoptive families who care for children with special needs (often due to early adverse experiences that landed them in foster care). But these adoption assistance programs are administered by states rather than the federal government, which has caused variation among states. Some states issue payment based on the type of adoptive family rather than the needs of the child, redirecting the subsidy away from its intended purpose (Hansen 2008).

One proposal to increase demand for children in foster care suggests an "all-pay simultaneous ascending auction with a bid cap" with prospective parents submitting sealed bids. Profits generated by placing healthy white infants would be used to place the children they call "less-desirable." (Blackstone 2004; Blackstone et al. 2008).

Lower Transactions Costs

Some legal economists have critiqued the agency gatekeeping function as extracting surplus (Blackstone et al. 2008) and increasing transactions costs (Brinig 1994). One proposal would reduce search costs by replacing that agency function with a national database, detailing the characteristics and requirements of prospective adoptive and birth parents and children on a platform akin to the Multiple Listing Service used by real estate agents (Balding 2010). A second proposal would impose a 10% tax on adoption expenses and channel the funds generated in high-price adoptions to subsidize adoption of children out of foster care (Goodwin 2006). A third proposal, already in place in many states, reduces search costs borne by both agencies and the children shuttled through foster care placements while they wait to be adopted. Instead of limiting the search for appropriate parents to married

heterosexuals, this approach expands the definition of suitable parents to include single people and same-sex couples. This expansion shortens the search because single and gay adoptive parents are more willing to create interracial families than their married, heterosexual counterparts (though whites in all three groups generally express a preference for white children) (Baccara and Collard-Wexler 2012).

Conclusion

Despite sharp criticism of early legal economic literature proposing economic frameworks for viewing adoption and remedying the two problems of queues of adoptive parents waiting to adopt healthy infants and queues of foster care children waiting to be adopted, economic and legal researchers continue to propose market solutions to remedy the situation.

Cross-References

▶ Becker, Gary S.
▶ Contract, Freedom of
▶ Gender Diversity
▶ Posner, Richard

References

Appleton S (2004) Adoption in the age of reproductive technology. University of Chicago Legal Forum 393–450
Baccara M, Collard-Wexler, A (2012) Child-adoption matching: preferences for gender and race, NBER Working paper 16444
Balding C (2010) A modest proposal for a two-sided market clearing institution under asymmetric supply constraints with skewed pricing: the market for adoption and abortion in the United States. J Public Econ Theory 12:1059–1080
Beam C (2013) To the end of June: the intimate life of American foster care. Houghton Mifflin Harcourt, Boston
Becker GS (1981) A Treatise on the family. Harvard University Press, Cambridge
Bernal et al (2007) Child adoption in the US, manuscript
Bitler M, Zavodny M (2002) Did abortion legalization reduce the number of unwanted children? Persp. Sex & Repro Health 34:25–33
Blackstone EA (2004) Privatizing adoption and foster care: applying auction and market solutions. Child Youth Serv Rev 26:1033–1049
Blackstone EA et al (2008) Market segmentation in child adoption. Int Rev Law Econ 28:220–225
Brinig MF (1993–1994) The effect of transactions costs on the market for babies. Seton Hall Legis J, 18:553
Ertman M (2015) Love's promises: how formal and informal contracts shape all kinds of families. Beacon Press, Boston
Ertman M (2003) What's wrong with a parenthood market. N C Law Rev 82:1
Gentenian LA (1999) The Supply of Infants Relinquished for Adoption. Econ Inq 37:412–431
Goodwin M (2006) The free market approach to adoption. Boston Coll Third World Law J 26:61–79
Gumus G, Lee J (2012) Alternative paths to parenthood: IVF or child adoption. Econ Inq 50:802–820
Hansen ME (2007) Using Subsidies to Promote the Adoption of Children from Foster Care. J Fam Econ Iss 28:377–393
Hansen ME (2008) The distribution of a federal entitlement. J Socio-Econ 37:2427–2442
Hansen ME, Hansen B (2006) The Economics of Adoption of Children from Foster Care. Child Welfare 85:559–583
Hasday J (2005) Intimacy and economic exchange. Harv Law Rev 119:491–530
Herman E (2008) Kinship by design: a history of adoption in the modern United States. Chicago University Press, Chicago, p 37
Hollinger (2004) State and federal adoption law. In: Cahn NR, Hollinger JH (eds) Families by law: an adoption reader. New York University Press, New York
Joyce K (2013) The child catchers: rescue, trafficking, and the new gospel of adoption. Public Affairs, New York
Krawiec KD (2009) Altruism and intermediation in the market for babies. Washington & Lee Law Rev 66:203–257
Landes EM, Posner RA (1978) The economics of the baby shortage. J Legal Stud 7:323–348
Medoff MH (1993) An Empirical Analysis of Adoption. Econ Inq 31:59–70
Moriguchi C (2012) The evolution of child adoption in the United States, 1950–2010: an economical analysis of historical trends. Discussion Paper Series A No. 572
Posner RA (1987) The regulation of the market in adoptions. Boston U L Rev 67:59–72
Solinger R, Susie WUL (1992) Single pregnancy and race before Roe v Wade. Routledge, New York
Williams PJ (1995) The rooster's egg: on the persistence of prejudice. Harvard University Press, Cambridge, MA, pp 218–225
Zelizer V (1985) Pricing the priceless child: the changing value of children. Basic Books, New York

Further Reading

Statutes regulating payment of birth mother expenses: Ind. Code § 35–46–1-9 (2013); Md. Code Ann., Fam. Law § 5–3A-45 (2013); In re Adoption No. 9979, 591 A.2d 468 (Md. 1991)

Alcohol Prohibition

Nicholas A. Snow
Department of Economics, Indiana University, Bloomington, IN, USA

Definition

Alcohol prohibition was a period from 1920 to 1933 in the United States of America where the manufacture, transportation, and sale of intoxicating beverages was made illegal.

Alcohol Prohibition

The national prohibition of alcoholic beverages in the United States of America lasted from January 16, 1920 to December 5, 1933. This period of time has become known as the "Noble Experiment," a phrase coined by President Herbert Hoover. Alcohol prohibition was instituted by the 18th Amendment to the United States Constitution, which stated, "the manufacture, sale, or transportation of intoxicating liquors within, the importation thereof into, or the exportation thereof from the United States and all territory subject to the jurisdiction thereof for beverage purposes is hereby prohibited." The Volstead Act was put into place for the purpose of enforcing this constitutional amendment. The experiment was meant to alleviate the various social ills that many perceived alcoholic beverages created such as violence, crime, corruption, and poor health.

In theory, prohibition would reduce consumption, which, in turn, would be followed by a reduction of these social problems. To judge this policy from an economic perspective, we look at whether or not this end goal was achieved. From this perspective, the "Noble Experiment" is often seen as failure.

It is interesting to note, however, that economists right before alcohol prohibition in the United States were primarily *for* prohibition. Their arguments rested on a belief that a sober society would outcompete heavy-drinking societies. Economist Irving Fisher's three books on the subject, 1927s *Prohibition at its Worst,* 1928s *Prohibition Still at its Worst*, and 1930s *The Noble Experiment*, all provided statistical evidence for the increased efficiency of prohibition and argued that the failures are a result of suboptimal enforcement. Fisher even claimed he could not find a single economist opposed to the prohibition of alcohol. Perhaps it was hindsight that changed this perception, as few would look back on alcohol prohibition as a success. The overwhelming empirical evidence and a look at economic theory clearly illustrate the difficulties alcohol prohibition had in successfully achieving its ends.

While alcohol consumption did decrease in the 1920s, prohibition did not help achieve the stated goals of prohibition. In fact, in a 1991 essay entitled "Alcohol Consumption During Prohibition," Miron and Zwiebel estimate that while consumption fell to 30% of the pre-prohibition level in the 1920s, it steadily rose throughout the rest of the prohibition era, reaching about 70% of the pre-prohibition level. And this occurred despite increased enforcement efforts.

From a theoretical standpoint, this is not surprising. Prohibition is a supply-reduction policy, which shifts the supply curve up and to the left. This leads to a reduction in quantity demanded; however, it also leads to an increase in price. The higher price offers a compensation for the increased risk prohibition creates in continuing to operate in the market. In other words, the increases in enforcement efforts do lead more risk adverse individuals to exit the market; however, the higher prices provide incentives for more daring entrepreneurs, and those more adept at using violence and secrecy, to remain or enter the market.

The prohibition also had a profound effect on the quality and nature of the goods within alcohol markets. Quality tended to fall dramatically as consumers turned to the black market or to produce their own in order to meet their demand. The largest effect is what Richard Cowan referred to as the "Iron Law of Prohibition," which states that the greater the level of enforcement, the greater to potency of the goods in question will become. Within the alcohol market, this meant a shift away from lower alcohol by volume drinks, and

thus bulkier, such as beer and wine, toward higher alcohol by volume drinks, such as whiskey.

Tellingly, economist Mark Thornton in a 1991 study entitled "Alcohol Prohibition was a Failure" analyzed the effectiveness of alcohol prohibition at achieving its stated goals of increase public health and decreasing crime. He found that prohibition had no discernable effect on public health, and perhaps more shockingly, he found that there was an increase in the crime rate, particularly among homicides. Thus, we must judge alcohol prohibition as having failed to accomplish its goals. These findings echo the work done by Clark Warburton at the tail end of prohibition with his 1932 book *The Economic Results of Prohibition.*

Fortunately, America's experiment with alcohol prohibition ended with the passage of the 21st amendment. After which, there was a dramatic reduction in crime, including organized crime, and an increase in health, jobs, and prosperity for all Americans.

Cross-References

▶ Prohibition

References

Cowan RC (1986) A war against ourselves: how the narcs created crack. Natl Rev 5:26–31
Fisher I (1927) Prohibition at its worst. Alcohol Information Committee, New York
Fisher I (1928) Prohibition still at it worst. Alcohol Information Committee, New York
Fisher I (1930) The noble experiment. Alcohol Information Committee, New York
Thornton M (1991) Alcohol prohibition was a failure. Policy Anal 157
U.S. Treasury Department (1930) Statistics concerning intoxicating liquors. Bureau of Industrial Alcohol December, 1930
Warburton C (1932) The economic results of prohibition. Columbia University Press, New York

Further Reading

Okrent D (2010) Last call: the rise and fall of prohibition. Scribner, New York
Thornton M (1991) The economics of prohibition. University of Utah Press, Salt Lake City

Alimony

Cécile Bourreau-Dubois and Myriam Doriat-Duban
Department of Economics, BETA UMR 7522, Université de Lorraine, Nancy, France

Abstract
Alimony or spousal support is a transfer of income between spouses intended mainly to reduce inequality in living standards following a divorce. Economic analysis provides two justifications for maintaining it in societies where women are now less financially dependent on their husbands and fault is no longer a decisive factor in divorce: the efficiency of marriage and the prevention of opportunism. It also provides some theoretical justifications for the methods used to calculate it.

Synonyms

Spousal support

Introduction

In many Western countries, the law provides that in the event of divorce, one of the spouses (generally the husband) will pay a sum of money to the ex-spouse (generally the wife) in the form of a lump-sum or regular payments, in particular when the separation leads to a significant difference in living standards. This transfer has different names in different countries: *pension alimentaire pour époux* in Quebec, *spousal support* in Canada, *alimony* in the United States, *spousal maintenance* in the United Kingdom, and *prestation compensatoire* in France. For a long time, the justification for it was based on the wife's financial dependence on the husband and/or adultery. In the past decades, two phenomena have raised questions about whether it is still justified to maintain this obligation between former spouses. The first one is the mass entry of women into the job market. The second one is the

no-fault divorce revolution (Leeson and Pierson 2017; Bracke and Mulier 2017), which leaves less room to the fault in divorce law. The economic analysis of alimony has provided new justifications, based on notions of efficiency and the prevention of opportunism; it also provides methods of calculation compatible with those justifications.

Alimony and the Economic Efficiency of Marriage

The first model of alimony was developed by Landes (1978), following on from the work done by Becker (1973, 1974). The approach is a normative one: alimony is supposed to encourage spouses to maximize their joint domestic production (income, upbringing of children, preparation of meals and activities contributing to the family's well-being).

The model rests on three hypotheses: the household's production is specific in the sense that it loses all value in the event of divorce, the domestic assets and wife's income depend on the amount of time she devotes to the home, and the husband's earnings depend on the time he devotes to his work but also on the time his wife devotes to domestic tasks. These hypotheses therefore fix a framework in which only the wife takes part in housework. The optimum degree of the wife's specialization is determined in a two-period model, where the variable to be maximized is the present value of the couple's total income. In this context, alimony is justified by the existence of transaction costs (costs of negotiating and implementing the transfers of wealth between the members of the couple). More precisely, if the household's revenue is totally divisible, if the transfers take place at no cost, and if the estimated likelihood of divorce is identical for both spouses, the optimum level of specialization is obtained without a judge having to decide alimony when divorce occurs. On the other hand, positive transaction costs (due to information asymmetries, indivisible public goods like children) can prevent the optimum level of domestic production being reached by "voluntary" transfers of income. Because it guarantees the wife that she will recoup the fruits of her investment in her household, alimony encourages her to specialize in domestic tasks, thereby maximizing the couple's expected income. Alimony is therefore nothing other than the implicit price, paid to the wife upon divorce, that encourages the spouses to behave in optimum fashion during the marriage.

This concept has aroused a considerable amount of criticism, in particular from feminist legal experts who object that it encourages women to specialize in domestic tasks and helps to perpetuate gender inequality (Singer 1994). Landes's hypothesis according to which only the wife can divide her time between market activities and nonmarket (domestic) activities is decisive in this result. Although Landes does not justify it and it can appear somewhat dated, Ellman (1989) points out that it can still be seen to correspond to a certain reality. Singer (1994) adds that this hypothesis also lacks any internal and external foundation. Thus, it ignores specialization between women rather than between spouses: nowadays the most productive household is one in which both spouses work full time and entrust care of the home and children to low-paid other people (most often women) (Carbone 1990; Brinig 1993). In addition, for many couples, a marriage is efficient when both members of the couple are committed to their career and devote a significant amount of time to their children (Brinig 1993).

Alimony and the Prevention of Opportunism

Marriage, in some ways, is quite similar to a long-term contractual relationship (Cohen 1987, p. 267). Alimony can then be justified by the concern to protect the weaker spouse who has made specific non-redeployable investments (i.e., domestic activities) from the opportunistic behavior of the other spouse.

Marriage is considered here as a long-term partnership, during which the spouses enter into reciprocal commitments, both explicit (imposed by law such as respect, faithfulness, moral assistance, duty of support) and implicit (promises). Nevertheless, marriage is different to a long-term commercial contract. Even if marriage has an instrumental value in the sense that the spouses can be considered as inputs enabling the production of an output (Becker's approach), above all it has an intrinsic value of its own (religious,

spiritual, proof of love) that a commercial contract does not have (Cohen 1987). By this specific dimension, marriage is a way of exchanging promises, whose value depends on the long-term behavior of each spouse. This is where a risk arises that is common to all long-term contracts: the risk of opportunism. The problem comes from the fact that the gains from the specific investments made generally by the wife are unequally spread over time. At the beginning of the union, the wife makes specific investments (bringing up children, cleaning, etc.) in return for the financial support of her husband, including once the children have left the parental home. The risk is then that the husband will adopt an opportunistic behavior consisting of breaking off the contract by a divorce just at the time when the spouse should be receiving her due, once the children have grown up. The husband thus appropriates what contract theory economists call the "quasi-rent" belonging to his wife. Geddes and Zak (2002) confirm that the husband's optimum strategy effectively consists of leaving his wife once she has made her contribution to the marriage. Dnes (1999) assimilates this behavior with a "greener grass effect" which encourages the husband to terminate the union before having to "pay" his wife for her investment in the family and to go to live with another – younger – woman. The consequences for the wife are all the greater when the investments made in the household cannot be redeployed on the labor and/or remarriage market. The change from fault-based divorce to no-fault divorce has increased this risk of opportunism because one of the spouses can now terminate the marriage at any time, virtually without compensation for the other spouse (Brinig and Crafton 1994).

The solution consists of making more costly for the potentially opportunistic party to break the contract costly for the potentially opportunistic party. Alimony can play this role since, on the one hand, the wife knows that in the event of a divorce, she will be compensated for her investment in domestic life (here we find Landes's incentive-based approach once again), while the husband is encouraged not to behave in an opportunistic way as divorce will not enable him to escape his commitments to his wife.

Alimony as a Means of Compensation

In both of these analyses, alimony constitutes a form of compensation justified by the recognition of a domestic investment made by the creditor during the marriage. This raises the question of how to compensate the party adversely affected by the ending of the marriage. By analogy with the notion of contract damages as compensation proposed in American contract law (Fuller and Perdue 1936), three forms of compensation are envisaged in the literature on alimony (Starnes 2011).

According to logic based on *restitution*, when one of the parties has made an investment that has led to the enrichment of the other party, but from which she cannot benefit, then it is necessary to pay her compensation. In this perspective, relying on the idea that marriage increases men's productivity, and thereby their income, Carbone and Brinig (1991), like Landes (1978), consider that divorce would deprive the wife of her due in terms of what she has contributed to her husband's success, especially when she has enabled him to continue his studies (Rea 1995) or when she has sacrificed her own career to create conditions more favorable to her husband's professional success (Ellman 1989). The logic of restitution then leads alimony to be calculated by placing the party that has benefited professionally from the marriage (generally the husband) in the same situation as if the marriage had never existed.

According to the *reliance interest* logic, the compensation aims to compensate the injured party for the loss incurred (e.g., expenses incurred) following the termination of the contract. According to Brinig and Carbone (1988), the application of this logic is relevant to divorce because the domestic investment can lead to opportunity costs, in this case a loss of human capital due to the slowing of the professional career (Ellman 1989). This involves placing the injured party (generally the wife) in the same situation as if she had never been married.

Finally, according to the logic based on *expectation*, the injured party is compensated for the benefits that could have been expected if the marriage had not been terminated. As Cohen (1987) emphasizes, the gains of a marriage are often asymmetrically divided between the spouses.

The spouse who works rapidly enjoys the benefits of the marriage (comfort of home and family life, facilitated professional career), while the spouse who invests in the domestic sphere at the beginning of the marriage reaps the gains in the long term, once the children have grown up. The loss caused by the divorce then resides in the loss of those gains, which is, according to Cohen (1987), more a loss of conjugal services (affection, sex, complicity, etc.) than a loss of career opportunities or a deterioration in standard of living. This therefore involves placing the "victim" in a situation identical to that which would have prevailed if the contract has been met. More precisely, the assessment of the loss incurred then involves the determination of the minimum sum the spouse will have to pay the other spouse to accept the divorce, if only divorce by mutual consent were possible. In other words, this involves determining the minimum sum that the spouse must pay to the other spouse so that it makes no difference whether they divorce or remain married. The expected loss criterion therefore guarantees that only efficient marriage terminations take place.

Bolin (1994) establishes a link between Landes's model (Landes 1978) and the three logics of compensation identified by American contract law. According to Bolin, this model, in which alimony guarantees an efficient specialization in the two spouses' roles, is compatible with an infinity of amounts of alimony, which depend on the parties' powers of negotiation. Nevertheless, these amounts are bounded by two extreme values: the highest amount that the husband would agree to pay and the lowest amount that the wife would agree to receive. More precisely, Bolin shows that the highest amount would correspond to the gain made by the husband as a result of the socially optimal investment of his wife in the household, which in fact would correspond to compensation according to the *restitution* principle. For its part, the minimal amount of alimony would correspond to the minimal amount accepted by the wife. This would be such that the wife is just compensated for the losses she has incurred due to her increased investment in domestic tasks. Such an amount would be very close to compensation based on the *reliance interest* logic. On the other hand, according to Bolin, compensation in terms of *expectation* could not be integrated into the efficiency model proposed by Landes insofar as it would lead to a level of investment on the part of the wife that would be higher than the optimal level.

Summary

Economic analysis provides justifications for the existence of alimony, including in modern societies where it no longer corresponds to a sanction for misconduct or a prolongation of the duty of support toward a financially dependent wife. The justifications are based first of all on an efficient division of tasks between the spouses and secondly on the prevention of a risk of opportunism to the detriment of the spouse that has made the specific investments. In addition, economic analysis proposes criteria for calculating alimony.

References

Becker GS (1973) A theory of marriage: part I. J Polit Econ 81(4):813–846

Becker GS (1974) A theory of marriage: part II. J Polit Econ 82(2):S11–S26

Bolin K (1994) The marriage contract and efficient rules for spousal support. Int Rev Law Econ 14:493–502

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

Brinig M (1993) The law and economics of no-fault divorce, family. Law Quart 26(453):456–58.

Brinig M, Crafton S (1994) Marriage and opportunism, J Leg Stud 23(2):869–894.

Brinig M, Carbone J (1988) The reliance interest in marriage and divorce. Tulane Law Rev 62:855–884

Carbone J (1990) Economics, feminism, and the reinvention of alimony: A reply to Ira Ellman. Vanderbilt Law Rev 1463:1463–1501.

Carbone J, Brinig M (1991) Rethinking marriage: feminist ideology, economic change, and divorce reform. Tulane Law Rev 65:953–1010

Cohen L (1987) Marriage, divorce and quasi-rents: or I gave him the best years of my life. J Leg Stud 16:267–272

Dnes A (1999) Marriage contracts. In: B. Bouckaert and De G. Geest (Eds), Ency Law Econ, Edward Elgar 2000:864–885.

Ellman I (1989) The theory of alimony. Calif Law Rev 77(1):1–81

Fuller LL, Perdue WR (1936) The reliance interest in contract damages. Yale Law J 46:52–96
Geddes R, Et Zak P (2002) The rule of one-third. J Leg Stud 31(1):119–137.
Landes E (1978) Economics of alimony. J Leg Stud 7(1):35–63
Leeson PT, Pierson J (2017) Economic origins of the no-fault divorce revolution. Eur J Law Econ 43: 419–439
Rea S (1995) Breaking up is hard to do: the economics of spousal support, working paper number UT-ECIPA-REAS-95-01. http://www.epas.utoronto.ca:5680/wpa/wpa.html
Singer J (1994) Alimony and efficiency: the gendered costs and benefits of the economic justification for alimony. Georgetown Law J 82:2423–2460
Starnes CL (2011) Alimony theory. Fam Law Q 45(2): 271–291

Allowance Trading

► Emissions Trading

Alternative Dispute Resolution

Nathalie Chappe
University of Franche Comté CRESE, Besançon, France

Definition

Alternative dispute resolution (ADR) refers to any mode of dispute resolution that does not utilize the court system, such as arbitration, neutral assessment, conciliation, and mediation. Methods of ADR are different from one another, but share common points, notably the feature that a third party is involved and a less formal and complex framework than courts. The third party offers an opinion about the dispute to the disputants or chose a binding decision. In recent decades, many countries have adopted rules requiring parties to go through some form of ADR before resorting to trial. ADR programs currently operate in a wide variety of contexts: union-management negotiations, commercial contract disputes, divorce negotiations, etc. The utilization of ADR mechanisms is championed by parties and lawyers, as well as by politicians or judges. Parties and lawyers hope to withdraw from ADR a benefit in terms of time and costs. Politicians and judges seek to relieve congestion in the courts by providing new methods of dispute resolution to meet the increasing demand for justice.

The economic analysis of the ADR mechanisms returns to the following question: is ADR actually an efficient and cost-reducing system? To answer the first point about efficiency, we need to know the incentives created by ADR: parties' incentives and third party's incentives. The second argument concerns the choice of an ADR mechanism by the parties. Once these answers are given, we will be able to determine if the ADR should be subsidized, provided, or mandated by the state. The economic approach is based on the assumption that individuals are rational and act by comparing costs and benefits.

ADR can be binding or nonbinding (or hybrid), voluntary or mandatory, and ex ante or ex post. Ex ante ADR concerns arrangements to use ADR made before a dispute arises, while ex post ADR refers to the use of ADR once the dispute has arisen. Secondly in binding ADR, the parties will be bound and abide by the decision of the third party. In nonbinding ADR, each party is free to reject the decision of the neutral and either takes the matter to court or does nothing. A hybrid procedure could include a nonbinding procedure followed by a binding procedure if needed. Thirdly, ADR can be voluntary or mandatory. These distinctions are important because the costs and benefits of different kinds of ADR may be different, hence a different impact on the parties' decision to turn to ADR, on the third party's behavior and on the incentives to settle.

Why Parties Turn to ADR to Resolve Disputes?

The incentives to turn to ADR are different between ex ante and ex post agreements.

Ex Ante Agreements

Ex ante ADR may be adopted because it would lead to mutual advantages. ADR may lower the cost of resolving disputes. Mediation and arbitration are indeed less expensive methods than court. ADR may also lower the risks attending disputes. Arbitration may be chosen because the arbitrator will be an expert whose award may be anticipated. Note, though, that the two previous benefits are equally advantages of ex post agreements.

Ex ante ADR may act as a quality signal. Chappe (2002) shows that in a contractual relationship between one buyer and two sellers (one of which offers a high-quality product, the other a low-quality product), the high-quality seller offers an arbitration clause, which then serves as a quality signal. Holler and Lindner (2004) analyze mediation as a signal; it possibly gives away information about the party who call for it. More precisely, to reject mediation can be interpreted as a "negative signal," while the interpretation of accepting or proposing mediation is ambiguous and provides no clear information.

Moreover, since ADR provides greater social benefits and information than litigation, the dynamics of the process should tend to induce the parties to include a clause submitting future disputes to ADR (e.g., an arbitration clause). Such a clause has a deterrent effect on behavior that may prevent the dispute. Defendants are induced to take a more efficient level of care since their anticipation in case of dispute is better.

Such benefits cannot be obtained by ex post agreement since it is too late to disclose information or to take care.

Ex Post Agreements

Shavell (1995) follows the standard model of litigation which assumes that disputing parties choose ADR or not depending on the potential outcomes. "Parties are risk neutral, bear their own legal costs, and know the judgment amount that would be awarded, but may hold different beliefs about the likelihood of a plaintiff victory." The expected outcome is defined as the probability of winning multiplied by the judgment amount won minus the costs. Parties will choose an agreement if and only if it is Pareto superior to their alternatives. Parties determine the net expected values (expected value for the plaintiff minus the expected value for the defendant) for ADR, settlement, and trial. Depending on this value, parties will opt for different method of dispute resolution. Shavell (1995) also holds that parties have probabilistic beliefs about how ADR influences the outcome of the trial: ADR perfectly predicts trial outcomes, ADR does not predict trial outcomes, and ADR imperfectly predicts outcomes. If ADR perfectly predicts trial outcomes, parties will opt for ADR which have a positive net expected value. If ADR does not predict trial outcome, nonbinding ADR will never occur.

Third Party

The role of the third party differs between binding and nonbinding ADR. In nonbinding ADR, as mediation, the third party never explicitly recommends how a dispute should be resolved. The third party assists parties in the resolution of their dispute and acts as a settlement facilitator. In binding arbitration, the neutral evaluates the merits of the case and issues a decision which is binding.

Binding

The most famous binding ADR is arbitration. Compared to court adjudication, the advantages and drawback of arbitration come from three differences. Firstly, unlike a judge (who is assigned), the arbitrator is a private person chosen by the parties. He is chosen for his expertise in the subject matter of the dispute. Consequently, the arbitrator's award may be more predictable (there is less risk) and faster (Ashenfelter 1987; Bloom and Cavanagh 1986).

In conventional arbitration, arbitrators make awards that are weighted averages of some notion of what is appropriate based on the facts and the offers of the parties where the weighs depend on the quality of offers as measured by the difference between the offers. In final offer arbitration, the arbitrators choose the offer that is closest to some notion of an appropriate award based on the facts (Bazerman and Farber 1985). Farber and Bazerman (1986) find strong results which support this framework.

Nonbinding

The mediator is not charged to solve the litigation, but simply to make proposals and to facilitate agreement. The mediator helps parties clarify issues, explore settlement options, and evaluate how best to advance their respective interests. The role of the mediator is thus above all to avoid the failure of agreement. Settlement may fail because of informational and psychological barriers (Arrow et al. 1995).

Models of settlement negotiations explain the occurrence of costly trials by assuming that parties have private information about the outcome of a trial. The role of the mediator will be to collect and distribute information in order to make sure that parties become aware of the mutual benefits that the mediation allows. Two points distinguish the mediator from the judge: distribution of information and confidentiality of the exchanges. Indeed, contrary to the judge who is held by the principle of contradictory, the mediator is free to disseminate or not informations. Ayres and Brown (1994) study how mediator resolve dispute by "caucusing" privately with the individual disputants. Mediators can create value by controlling the flow of information to mitigate moral hazard and adverse selection. Moreover, the information disclosed to the mediator cannot be used at trial. If such were not the case, the parts would not be encouraged to reveal information knowing that in the event of failure, the court could be put in possession of information which could be used for its decision, whereas in the absence of mediation, it would not have mentioned these facts within the framework of the lawsuit.

The main cognitive barriers are the attitude toward risk, the aversion for the losses and the framing effects. The mediator will limit the cognitive and the "reactive devaluation" barriers to conflict resolution (Arrow et al. 1995).

What Incentives to Settle Are Created by ADR?

Adding ADR to the traditional litigation process has complex and important effects on several ways. Notably, ADR will have an effect on the incentives to settle. What is the incentives for parties to settle their disputes if ADR makes them less costly? We consider firstly binding ADR and secondly nonbinding ADR. In the two cases, incentives depend on the third party's role which we have described above.

Binding ADR: The Case of Arbitration

It is often claimed that the final offer arbitration (FOA) system is more likely than the conventional arbitration (CA) system to lead the parties to settle. Under conventional arbitration (CA), the parties present their cases to an arbitrator who has the flexibility to impose any award he or she deems appropriate. The major criticism of CA is that arbitrators are perceived to compromise between the parties' final offers, which encourages them to take extreme positions into arbitration resulting in a "chilling effect" on negotiated settlement (Farber 1981). FOA is claimed to induce disputants to stake out more reasonable positions (Farber and Katz 1979; Farber 1980). Under FOA, the arbitrator must choose either one party's or the other party's final offer. If the offers of the two disputants do not converge or crisscross, then the arbitrator chooses the one that is closest to his/her notion of a fair settlement as the settlement. However, while FOA improves convergence and outperforms CA, several papers show that the design of the mechanism is not sufficient to harmonize the parties' positions (Chatterjee 1981). To restore convergence between offers, several procedures have been presented. Brams and Merill (1986) propose a "combined" arbitration procedure (CombA): if the arbitrator's notion of a fair settlement lies between the disputants' final offers, then the rules of FOA are used. Otherwise, the rules of CA are used. Under some conditions (concerning the density function of the estimate about the arbitrator's notion of a fair settlement), this procedure succeeds in overcoming the defects of FOA and increases voluntary settlement rates. Zeng et al. (1996) propose to improve FOA by letting each disputant make double offers (double-offer arbitration, DOA). They show that the two secondary offers converge. DOA succeeds in inducing the convergence of the offers made by the

disputants without assumptions about the density function as in CombA. Zeng (2003) proposes an amendment to FOA. The final arbitration settlement is determined by the loser's offer instead of the winner's offer, if the offers diverge. The author shows that this design induces two disputants to submit convergent offers. Finally, Armstrong and Hurley (2002) propose to incorporate what they call the closest offer principle. Arbitrator may put a higher weight on the disputant's offer which is closest to his/her own assessment. Their model allows them to generalize previous models of both CA and FOA. They derive the equilibrium offers that risk neutral disputants would propose and show how these offers would vary under different arbitration procedures: FOA and CA. They show that the offers made under CA are always more extreme than those made under FOA.

Chappe and Gabuthy (2013) highlight the strategic implications of legal representation in arbitration by analyzing how the presence of lawyers may shape the disputants' bidding behavior. They show that the contingent payment mechanism has interesting properties in that it enhances the lawyer's incentives to provide effort in defending her client (in comparison with the case without legal representation), without altering the gap between the parties' claims in arbitration.

Nonbinding ADR

Doornik (2014) considers a mediator who acts as a channel to verify and communicate the informed party's beliefs about the expected value of the judgment, without sharing any supporting evidence that would advantage her opponent in court. This increases the incentive for the informed party to share their private information about the likely outcome in court and hence increases settlement rates.

Dickinson and Hunnicutt (2010) examine the effectiveness of implementing a nonbinding suggestion prior to binding dispute settlement. On the one hand, a nonbinding suggestion may serve as a focal point, thereby improving the probability of settlement. On the other hand, a nonbinding suggestion may reduce uncertainty surrounding the outcome from litigation, thereby increasing dispute rates. Their results suggest that nonbinding recommendations improve bargaining outcomes by reducing optimism. Dispute rates are significantly lower when a nonbinding recommendation is made prior to arbitration.

Conclusion

ADR is now an integral part of the litigation system. Ex ante ADR agreements may result in mutual advantages for the parties to a contract, in superior incentives to disclose information or to avoid disputes, and in improved incentives to settle. Consequently, ex ante ADR agreement should be enforced. Nevertheless, some ADR (specifically arbitration) may be more expensive in comparison to the courts. Incorporating ex post ADR agreements to the standard litigation system has effects on incentives to settle immediately and to go to trial. Since parties settle in the shadow of ADR, their behavior is affected by the introduction of ADR. The conclusions about the desirability of ADR depend significantly on several factors: (i) the cost of ADR (mediation may be cheaper than trial, while arbitration may be more expensive), (ii) the predictive power of the ADR about the trial outcome, and (iii) the degree of coercion from the decision (binding or nonbinding). Consequently it is difficult to give a general recommendation requiring compulsory arbitration.

References

Armstrong MJ, Hurley W (2002) Arbitration using the closest offer principle of arbitrator behavior. Math Soc Sci 43(1):19–26

Arrow K. et al. (1995) Barriers to conflict resolution. W. W. Norton & Company, 1st ed, p. 368

Ashenfelter O (1987) Arbitrator behavior. Am Econ Rev 77(2):342–346

Ayres I, Brown JG (1994) Economic rationales for mediation. Va Law Rev 80:323–402

Bazerman M, Farber H (1985) Arbitrator decision making: when are final offers important? Ind Lab Relat Rev 39(1):76–89

Bloom D, Cavanagh C (1986) An analysis of the selection of arbitrators. Am Econ Rev 76:408–421

Brams SJ, Merill S (1986) Binding versus final-offer arbitration: a combination is best. Manag Sci 32(10): 1346–1355

Chappe N (2002) The informational role of the arbitration clause. Eur J Law Econ 13(1):27–34
Chappe N, Gabuthy Y (2013) The impact of lawyers and fee arrangements on arbitration. J Inst Theor Econ 139:720–738
Chatterjee K (1981) Comparison of arbitration procedures: models with complete and incomplete information. IEEE Trans Syst Man Cybern 11(2):101–109
Dickinson D, Hunnicutt L (2010) Nonbinding recommendations: the relative effects of focal points versus uncertainty reduction on bargaining outcomes. Theory Decis 69:615–634
Doornik K (2014) A rationale for mediation and its optimal use. Int Rev Law Econ 38:1–10
Farber H (1980) An analysis of final-offer arbitration. J Confl Resolut 24(4):683–705
Farber H (1981) Splitting-the-difference in interest arbitration. Ind Lab Relat Rev 35:70–77
Farber H, Bazerman, (1986) The general basis of arbitrator behavior: an empirical analysis of conventional and final-offer arbitration. Econometrica 54(6):1503–1528
Farber H, Katz H (1979) Interest arbitration, outcomes, and the incentive to bargain. Ind Labor Relat Rev 33(1):55–63
Holler MJ, Lindner I (2004) Mediation as signal. Eur J Law Econ 17:165–173
Shavell S (1995) Alternative dispute resolution: an economic analysis. J Legal Stud 24:1–28
Zeng DZ (2003) An amendment to final-offer arbitration. Math Soc Sci 46(1):9–19
Zeng D-Z, Nakamura S, Ibaraki T (1996) Double-offer arbitration. Math Soc Sci 31(3):147–170

Altruism

Sophie Harnay
EconomiX CNRS UMR 7235, University Paris Ouest Nanterre La Defense, Nanterre, France

Abstract

Until recently, most works in the economic analysis of law (EAL) have assumed that material self-interest exclusively motivates economic agents, including individuals involved in the functioning of the legal system. Altruistic choices, defined as a sacrifice that benefits others, thus remain outside the scope of the EAL. They become an issue for the EAL when the question whether to substitute external (legal) incentives for internal (moral) motives when altruism appears insufficient to have people help each other gradually emerges as a by-product of the liability controversy and as the outcome of the development of the economics of altruism in the 1970s.

Altruism

Since Adam Smith's *Theory of Moral Sentiments* and with landmarks in works of John Stuart Mill, Francis Ysidro Edgeworth, Vilfredo Pareto, or Léon Walras, the problem of altruism and giving, and more generally of the adoption of moral norms, has been a long-standing issue and a challenge within economics, due to the obvious tension between altruism and the self-interest assumption. Thus, most works in the economic analysis of law (EAL) assume that material self-interest exclusively motivates all people. Accordingly, individuals involved in the functioning of the legal system – criminals, litigants, judges, and lawyers – are supposed to behave as rational and self-interested utility maximizers. Furthermore, rules are seen as incentives designed to induce agents to behave in a way that is consistent not only with their private interest but also with social efficiency and public interest. Within this framework, thus, altruistic behaviors have not been an issue for the EAL for a long time, as they are apparently "irrational," and the study of helping behaviors and of a wide set of situations involving individuals who depart from the pursuit of their own interest to help others has been overlooked.

In the field of the economic analysis of law, disinterested behaviors start to become an issue in the early 1970s, as a by-product of the controversy on the economic analysis of liability and negligence that rages at the time. The questions that are raised then are the following: how is it possible to explain behaviors that are not obviously the consequence of selfishness? Can unselfish behaviors possibly be regulated? Should disinterested behaviors be encouraged and regulated? Is that possible to substitute external (legal) incentives for internal (moral) motives when altruism appears insufficient to have people cooperate and help each other?

A first step in the acknowledgment of disinterested behaviors in the law and economics literature is taken with Richard A. Posner's pioneering works on liability rules (Posner 1972). Arguing that liability in a case should be decided following a cost-benefit analysis, Posner explicitly refers to rescue situations, helping behaviors, and good and bad Samaritanism as illustrations for his economic theory of liability and support for its negligence rule within the debate on strict liability versus negligence (Posner 1973). However, no reference is made to the economic analysis of altruism at that time: within Posner's analytical framework, rescue situations do not differ from less extreme situations substantially and, therefore, they do not call for specific liability rules.

A second step focusing on the regulation of rescue behaviors more explicitly builds on the advances in the economic analyses of altruism that are made during the 1970s. During the decade, indeed, economists start to analyze altruistic behaviors with economic tools and assumptions. They approach altruism in three main ways. First, following Becker's works (1973, 1974a, b, 1976) and Dawkins (1976), a first family of works models altruism using interdependent utility functions in which the welfare of others enters an individual's utility function: thus, the higher the utility level of others, the higher the utility of the benefactor. A second family of analyses studies reciprocal altruism, as it is determined by the expectation of future gains within repeated interactions between players. A third set of models assumes dual utilities of agents: following up Harsanyi's works (1976), they consider a dual nature of man combining both an egoistic and an altruist side within a single individual. Such duality then accounts for pro-social behaviors, including pure altruism, and explains disinterested behaviors as the consequence of a moral imperative ruling out rational calculation and a natural propensity to moral behaviors. The economic debate on altruism is also fueled by the discussion of economists with biologists and the emergence of sociobiology (Wilson 1975).

In the field of law and economics, analyses concentrating on rescue behaviors mostly build on models of interdependent utility functions, in which the altruistic rescuer's utility function becomes a function of the utility of the endangered person (Landes and Posner 1978a, b). Analyses therefore endorse Becker's claim according to which economic analysis is sufficient to explain altruistic behaviors, without the need to resort to sociobiology and models of group selection to account for altruistic behavior. Furthermore, reciprocal altruism is also considered as irrelevant to account for rescue behaviors, as rescue most often involves strangers and relies on some form of pure altruism, devoid of all expectation of compensation. Rather, altruistic behavior is justified by a "recognition factor" that rescuers receive from helping others. Consequences in terms of regulation are then drawn from the existence of altruism. If people are mainly motivated by recognition in their altruistic transactions, including their rescue activities, it may not be efficient to impose liability for non-rescue: because it reduces the public recognition allotted to the altruistic individual, liability rules may then have a pernicious effect and decrease the number of altruistically motivated actions. In that sense, altruism thus provides a substitute for costly legal rules intended to internalize external benefits of rescues in emergency situations, when transaction costs are too important to allow an efficient use of legal rules.

References

Becker GS (1973) A theory of marriage: part I. J Polit Econ 81(4):813–846

Becker GS (1974a) A theory of social interactions. J Polit Econ 82(6):1063–1093

Becker GS (1974b) A theory of marriage: part II. J Polit Econ, Part 2. Marriage, family human capital and fertility 82(2):S11–S26

Becker GS (1976) Altruism, egoism and genetic fitness. J Econ Lit 14(3):817–826

Dawkins R (1976) The selfish gene. Oxford University Press, Oxford

Harsanyi JC (1976) Essays on ethics, social behaviour, and scientific explanation. Reidel, Dordrecht/Boston

Landes WM, Posner RA (1978a) Salvors, finders, good Samaritans, and other rescuers: an economic study of law and altruism. J Leg Stud 7(1):83–128

Landes WM, Posner RA (1978b) Altruism in law and economics. Am Econ Rev 68(2):417–421

Posner RA (1972) A theory of negligence. J Leg Stud 1(1):29–96
Posner RA (1973) Strict liability: a comment. J Leg Stud 2(1):205–221
Wilson EO (1975) Sociobiology, the new synthesis. Belknap/Harvard University Press, Cambridge, MA

Amnesty Programs

► Leniency Programs

Analytical

► Rationality

Anarchy

Peter T. Leeson
Department of Economics, George Mason University, Fairfax, VA, USA

Abstract

Anarchic environments are ubiquitous. Unable to rely on government, individuals in such environments develop private institutions of governance to promote social cooperation. Private governance consists of privately created social rules and mechanisms of their enforcement. Anarchy may secure better socioeconomic development than government in least-developed countries.

Synonyms

Statelessness

Definition

Anarchy is the absence of government. Government is more difficult to define. Typically government refers to a governance institution with a territorial monopoly on force. However, when government is defined this way, the presence or absence of government, and thus the presence or absence of anarchy, depends on how one construes the territory in question. Other attempts to define government in terms of characteristics with which government is often associated, such as compulsory recognition of a governance institution's authority, or a "high" cost of exiting a governance institution, are similarly problematic. Many governance institutions ordinarily characterized as nongovernmental share these characteristics, confounding efforts to define government, and thus anarchy, in a way that consistently reflects common intuitions.

Despite this conceptual difficulty, there is widespread agreement about whether or not government, and thus anarchy, prevails in concrete cases. Although a criminal gang, for example, may enjoy a monopoly on force in a neighborhood, few would consider the gang a "government." On the contrary, most would characterize the gang's presence as an outcome of government's effective absence and thus the environment in which the gang operates as anarchic. The distinction between government and anarchy is therefore useful in practice and poses little problem for applied research that seeks to study "government" or "anarchy" (Leeson 2014a).

Anarchic Environments

A large number and wide variety of social environments worldwide were or are anarchic. Most of the world for most of its history lacked government. And, if one goes back far enough, all historical environments are anarchic.

In the contemporary world, dysfunctional governments in least-developed countries have resulted in large populations' inability to rely on government to protect their lives and property. Indeed, in such countries, government is often the chief depredator of citizens' lives and property. In one least-developed country, Somalia, there has been no government at all for several decades.

Despite having high-functioning governments, the contemporary developed world also hosts numerous anarchic environments. Even where government is highly functioning, the state's eye cannot be everywhere all the time. Moreover, appealing to governmental institutions to enforce contracts and property rights is often expensive and slow (Stringham 2015). Many social and commercial activities in the developed world thus occur outside government's de facto reach. Black markets, which are criminal and, even in developed countries, large, also present anarchic environments. Such markets' participants cannot rely on government to secure their property rights or promote cooperation between them (Skarbek 2014).

Internationally, the world has always exited in, and continues to exist in, anarchy. Various supranational organizations, such as the United Nations, seek to facilitate international cooperation. However, such organizations' member states remain sovereign. There is therefore no supranational agency with the authority to create or enforce binding laws on nation states and no agency with the authority to act as the final arbiter of disputes between them.

Interactions between private international traders also occur in an anarchic international environment. Before 1958, private international commercial contracts could not be enforced in government courts. Even after 1958, when the first international treaty was created for this purpose, the vast majority of international commercial contracts continue to be enforced without government. Moreover, like all international treaties, those which promise government enforcement of international commercial agreements are not themselves enforceable by government, since no supranational government exists (Leeson 2008).

Anarchy and Private Governance

In the seventeenth century, Thomas Hobbes famously characterized life in anarchy as "solitary, poor, nasty brutish, and short." Contemporary conventional wisdom sees anarchy similarly: anarchy is widely believed to be chaotic, violent, and impoverishing. A growing body of research, however, finds that Hobbes' characterization of anarchy, and the contemporary conventional wisdom that echoes it, is wrong (Powell and Stringham 2009).

Although anarchic environments lack government, they do not lack governance: institutions that create and enforce social rules. In anarchy, such governance is provided privately – i.e., through the activities of private individuals seeking to make themselves better off. Private governance institutions in anarchy can be highly effective at producing social order, regulating violence, and more generally facilitating social cooperation.

Private governance institutions in anarchy take myriad forms. The particular forms they take reflect the particular problems of social cooperation that the people who find themselves in particular anarchic environments face. Private social rules may emerge in anarchy "organically" via the interactions of individuals, which, over time, generate customs or social norms that govern behavior. Private international commercial law, for example, emerged in this fashion via the interactions of medieval international merchants who, in the absence of supranational government, required rules to facilitate exchange between them (Benson 1989). Alternatively, private social rules may emerge in anarchy via individuals' deliberate construction. Nineteenth-century American settlers in the so-called "wild West," for instance, established private land clubs and cattle associations with the express purpose of creating laws to protect their property rights prior to the United States' government's appearance in frontier areas (Anderson and Hill 2004).

Private rule-enforcement mechanisms may also emerge in anarchic environments both "organically" and through private individuals' deliberate efforts. The most basic of such mechanisms is the "discipline of continuous dealings," according to which a social rule breaker is shunned or otherwise cut off from the benefits of future cooperative dealings by the members of his community. The members of Gypsy communities, for instance, whose gray market economic

activities typically prevent them from relying on government, use the threat of being ousted from the Gypsy world to enforce economic agreements between one another (Leeson 2013).

Other private punishments for the enforcement of social rules in anarchy include, for example, monetary fines, physical or capital punishment, and supernatural sanctions (Friedman 1979; Leeson 2009, 2014b). Private punishments for rule breaking in anarchy are often part of more elaborate and sophisticated private enforcement institutions. Eighteenth-century Caribbean pirates, for instance, who, as criminals, could not rely on government, used fines and capital punishment to punish violations of their private law, but developed, interpreted, and administered these punishments in the context of a broader system of constitutional democracy they developed to privately govern their ships (Leeson 2007a).

Anarchy and Development

Anarchy's relationship to socioeconomic development is nuanced. On the one hand, no anarchic society has come close to the level of socioeconomic development achieved by the most successful societies governed by states, namely, the countries of contemporary North America and Western Europe. This strongly suggests that government is necessary for maximal development.

On the other hand, governments in North America and Western Europe are unusual in the world in that they are highly functional. In most of the world, governments are not nearly so high functioning; and in much of the world, governments are highly dysfunctional. For a variety of reasons, governments in such places have failed to realize the potential exhibited by governments in the wealthiest nations.

This failure raises the question of whether private governance in anarchy may be able to secure a higher level of development than dysfunctional government. The developmental superiority of high-functioning government to private governance does not imply the developmental superiority of dysfunctional government to private governance. Somalia, whose dysfunctional government collapsed in 1991, is the only contemporary nation that has experienced both a significant period under dysfunctional government and one in anarchy and thus the only contemporary nation that permits a direct comparison of development under these alternative governance arrangements. Nevertheless, the results of such a comparison are instructive.

On nearly every available development indicator, Somalia exhibited a marked improvement in anarchy compared to under its former government (Leeson 2007b; Powell et al. 2008; Leeson and Williamson 2009). This does not mean, of course, that Somalia achieved high levels of development in anarchy (it did not), nor does it mean that Somalia could not achieve dramatically higher levels of development if it could achieve a high-functioning government. It does, however, suggest the possibility that where dysfunctional government and private governance are the only governance alternatives – as would seem to be the case in much of the least-developed world – anarchy may be superior to government for development.

References

Anderson TL, Hill PJ (2004) The not so wild, wild west: property rights on the frontier. Stanford University Press, Stanford

Benson BL (1989) The spontaneous evolution of commercial law. Southern Econom J 55:644–661

Friedman D (1979) Private creation and enforcement of law: a historical case. J Legal Stud 8:399–415

Leeson PT (2007a) An-*arrgh*-chy: the law and economics of pirate organization. J Polit Econom 115:1049–1094

Leeson PT (2007b) Better off stateless: Somalia before and after government collapse. J Comp Econ 35:689–710

Leeson PT (2008) How important is state enforcement for trade? Am Law Econ Rev 10:61–89

Leeson PT (2009) The laws of lawlessness. J Legal Stud 38:471–503

Leeson PT (2013) Gypsy law. Public Choice 155:273–292

Leeson PT (2014a) Anarchy unbound: why self-governance works better than you think. Cambridge University Press, Cambridge

Leeson PT (2014b) God damn: the law and economics of monastic malediction. J Law Econ Organ 30:193–216

Leeson PT, Williamson CR (2009) Anarchy and development: an application of the theory of second best. Law Develop Rev 2:77–96

Powell B, Stringham EP (2009) Public choice and the economic analysis of anarchy: a survey. Public Choice 140:503–538
Powell B, Ford R, Nowrasteh A (2008) Somalia after state collapse: chaos or improvement? J Econ Behav Organ 67:657–670
Skarbek D (2014) The social order of the underworld: how prison gangs govern the American penal system. Oxford University Press, Oxford
Stringham EP (2015) Private governance: creating order in economic and social life. Oxford University Press, Oxford

Anticommons, Tragedy of the

Qianwei Ying
Business School, Sichuan University, Chengdu, Sichuan, China

Definition

The tragedy of the anticommons is a type of coordination breakdown, in which a single resource has numerous owners, each of whom has the right to individually exclude others from its use but no effective privilege of using it independently. This coordination failure, i.e., the exercise of the veto power from any one of the owners will result in the underuse of the resource, frustrating what would be a socially desirable outcome.

The Concept of the Tragedy of the Anticommons

According to the definition by Heller (1998), an *anticommons property* is a scarce resource in which multiple owners have the right to individually exclude others from its use, and no one has an effective privilege of using it independently. The use of anticommons property requires these numerous right holders of a single resource to reach a consensus on the permission. The coordination breakdown of the owners, i.e., the exercise of the veto power from any one of the excluded right holders, will result in the underuse of the property, frustrate what would be a socially desirable outcome, and thus lead to the "tragedy of anticommons." The "tragedy of the anticommons" can be considered as a mirror image of the older concept of tragedy of the commons described notably by Hardin (1968), where multiple individuals are endowed with the privilege to individually use a given resource but without a right or an effective way to monitor and constrain each other's use, leading to the overuse of the resource. In terms of efficiency, the problem of the anticommons is based on a positive externality, while the problem of the commons is based on a negative externality.

The concept of the "tragedy of anticommons" is especially important for the transitional economies during the privatization process. One solution to the traditional tragedy of the commons is to fragment the common property into pieces and distribute them to different private owners. Private owners tend to avoid overuse because they can benefit directly from conserving the resources they control. Unfortunately, in the privatization process, too many separate owners of a single resource may be created, so that each one can block the others' use of the resource as a whole. Meanwhile, the separate use of the part of the resource that each separate owner controls may be impossible or with little value. As a result, nobody can use the resource efficiently if cooperation fails, leading to the tragedy of anticommons. For example, suppose an acre of land is distributed to 100,000 individuals who each own one square foot. The land will never be used for anything as long as thousands of people have to agree on what to do.

The anticommons is a paradox. While private ownership usually increases the efficiency of the use of scare resources and the social wealth, too many private owners from the fragmentation of a single property right has the opposite effect: it stops efficient use of the resource as a whole, hinders future innovation based on existing numerous intellectual property rights hold by different owners, and may cost our lives (Heller 2008). Anticommons property appears when new property rights are being defined and allocated without mechanisms for resolving

ownership disputes. While Heller devotes most of his attention to the underuse of commercial property in Moscow and other cities in the former Soviet Union, the concept of the anticommons has much wider implications in many other cases. Anticommons property can also emerge in developed markets wherever new property rights are being defined. This may occur when new technologies make possible uses of, for example, intellectual property and environmental rights, unanticipated by the previously existing legal mechanism.

Once anticommons property appears, it is difficult to remedy the situation either through markets or subsequent regulation. Care must be taken to avoid the accidental creation of anticommons property when new property rights are being defined by conveying core bundles of rights rather than multiple rights of exclusion.

Typical Examples of the Tragedy of Anticommons

Tragedy of Anticommons in the Transition from Marx to Markets in Moscow

A classical example of the tragedy of anticommons is proposed by Michael Heller in his famous article "The tragedy of the anticommons: Property in the transition from Marx to markets" published in Harvard Business Review, in which he examines a paradox in Moscow after the dissolution of the Soviet Union (Heller 1998). Storefronts remained empty even though the economy was growing and there was demand for consumer goods. In contrast, street kiosks in front of them filled with goods and customers. Why did the new merchants not come in the storefronts? One reason was that transition governments often failed to endow any individual with a bundle of rights that represents full ownership. Instead, fragmented rights were distributed to various stakeholders, including private or quasi-private enterprises, workers' collectives, privatization agencies, and local, regional, and federal governments. The fragmentation of the property right on the storefronts leads to a tragedy of the anticommons, an underuse of the storefronts due to the allocation of multiple new owners with the rights to exclude others from its use. An owner benefits from keeping a storefront empty, by excluding others because exclusion preserves the value of the right, perhaps for later trade to property bundlers or perhaps for use in rent seeking. The right of exclusion is valuable precisely because others want to use the resource and will pay something to collect the right. Because multiple parties may hold the same right, almost any use of the storefront requires the agreement of multiple parties. Even if only one party opposes the use, that party may be able to block others from exercising their rights.

Moving a storefront from anticommons to private property ownership requires unifying fragmented property rights into a usable bundle. In other words, creating private property requires moving from too many owners, each exercising a right of exclusion, to a sole decision maker, controlling a bundle of rights. Unfortunately, this process of collecting the fragmented property rights back together could be extremely difficult. In markets, entrepreneurial property bundlers may assemble control over stores by negotiating with all the holders of rights of exclusion. However, the market route to bundling rights might fail altogether if the transaction costs of bundling exceed the gains from conversion or if owners engage in strategic behavior such as holding out for the conversion premium.

The tragedy of the storefront anticommons is that owners waste the resource when they fail to agree on a use. Empty stores result in forgone economic opportunity and lost jobs. As of 1995, about 95% of commercial real estate in Russia remained in some form of divided local government ownership. Of this commercial real estate, a significant portion was unused.

Tragedy of Anticommons in Patents and Intellectual Property Rights

Patents and other forms of intellectual property protection for upstream discoveries may stimulate incentives to undertake risky research projects and could result in a more equitable distribution of profits across all stages of R&D. However, too many owners holding patents or intellectual

property rights in previous discoveries may constitute obstacles to future research. A researcher who may have felt entitled to coauthorship or a citation in an earlier era may now feel entitled to be a coinventor on a patent or to receive a royalty under a material transfer agreement. The result has been a spiral of overlapping patent claims in the hands of different owners. Researchers and their institutions may resent restrictions on access to the patented discoveries of others, yet nobody wants to be the last one left dedicating findings to the public domain. The tragedy of the anticommons in patents and intellectual property rights thus appears when a user needs access to multiple patented inputs to create a single useful product. Each upstream patent allows its owner to set up another tollbooth on the road to product development, adding to the cost and slowing the pace of downstream innovation.

The severity of the "tragedy of the anticommons" problem in patents and intellectual property rights depends on the difficulty in negotiating with the upstream patent holders. The larger is the number of separate patent holders, the more severe is the tragedy of the anticommons. On the other hand, if the upstream patent holders are united to one entity to make decisions on the aggregate royalty, it would be much easier for the downstream producers to obtain the right of using all these patents and thus avoid the tragedy of the anticommons. For example, building a DVD player requires using hundreds of patented inventions. No company could ever build a DVD player if it had to negotiate with all patent holders and obtain their unanimous consent. These patents would be worthless due to gridlock. Fortunately, the owners of the patents used in building DVD players have formed a single entity authorized to negotiate on their behalf. But if you're creating something new that does not have an organized group of patent holders, there will be "tragedy of the anticommons."

As suggested by Heller and Eisenberg (1998), an anticommons in biomedical research may be more likely to endure than in other areas of intellectual property because of the absence of organized group of patent holders, the high transaction costs of bargaining, heterogeneous interests among owners, and cognitive biases of researchers. In this case, policy-makers should seek to ensure coherent boundaries of upstream patents and to minimize restrictive licensing practices that interfere with downstream product development. Otherwise, more upstream rights may lead paradoxically to fewer useful products for improving human health.

The same logic can be applied to other high-tech industries besides of biomedical research. It is simply impossible to create a high-tech product these days without infringing on patents. For example, a new software or electronic device may use thousands of patents. It may not be practical even to discover all the possible patents involved, and it is certainly difficult to negotiate with thousands of patent holders individually. Therefore, overprotection on the patents or intellectual property rights may easily lead to a tragedy of anticommons for future product or technology development.

Tragedy of the Anticommons in Water Markets in the United States

Bretsen and Hill (2009) proposed another typical example of tragedy of the anticommons in water markets of the United States. Water shortages are becoming an increasingly important concern in much of the American West. Urban and environmental demands for water are increasing rapidly, and both physical and institutional constraints prevent new water supplies from being developed. With increasing demands for water for municipal, industrial, and environmental uses, transfers of water from agricultural irrigation to other uses will produce economic gains. A study of water transfers between 1987 and 2005 revealed dramatic differences in the value of water in urban uses versus agricultural uses (Brewer et al. 2008). Other estimates place the marginal value of water in municipal and industrial use at three to four times its marginal value in agricultural uses (Carey and Sunding 2001). However, in the face of increasing demand for water for urban and environmental uses in the western United States, water transfers out of agriculture have been fewer than one would expect

based on price differentials, and most of those that have occurred have required extensive negotiations. The reason for the difficulty in water transfer lies in the tragedy of the anticommons, since the existence of multiple rights of exclusion unbundled from the rights of use under the prior appropriation doctrine in the American West creates an anticommons that has impeded water transactions.

The multiple exclusion rights are the result of the evolutionary path of water institutions and the expansion of certain legal doctrines, such as the public interest and public trust doctrines which were originally designed to protect the property rights of people who used navigable waters and who depended on return flows from other irrigators.

The existence of multiple veto rights over water that are not bundled with use rights makes marketing of water difficult in the American West. In some cases no water trades will occur because of the multiplicity of veto rights; in other cases the tragedy of the anticommons leads to protracted and expensive negotiations with the need for side payments to secure the approval of all concerned. Furthermore, even after water transfers have been placed under contract, there is still the strong possibility that lawsuits can be brought to invalidate such transfers. The end result is an anticommons in which exclusion rights prevent the exercise of use rights and lock water into lower-value agricultural uses rather than allowing voluntary transfers into higher-value municipal, industrial, and environmental uses.

Tragedy of the Anticommons in Enterprise Licensing in China

As a developing country in the process of transition, China has inherited a complicated bureaucratic system of regulation from its past as a centralized planned economy. For several decades (which continues on today), almost all areas of business have been more or less under tight regulation. Routine enterprise licensing procedures involved multiple sign-offs and approvals, each with an independent right to reject any application. To get an investment project approved, an entrepreneur needs the permissions from all the related government agencies in the chain of licensing procedures, such as the local governments' administration for industry and commerce, local environmental protection bureau, local health bureau, local construction bureau (for safety work concern), fire prevention guards, and so on.

Without bribery, a government licensing agency may benefit little from approving the project but may bear great responsibilities if the project they passed caused environmental, safety, or other accidents later. The government licensing agency faced with such an asymmetry between return and responsibility may simply choose to exercise its veto right to avoid the potential responsibilities in the future. So long as any one of the licensing agencies chooses to exercise its veto right, the entrepreneur's investment project will be rejected. Therefore, with the assumption of no bribery, an entrepreneur's project may be easily rejected when there are multiple licensing agencies, leading to a tragedy of the anticommons. As shown in Ying and Zhang (2008), the larger the number of licensing agencies in the procedure is, i.e., the higher the fragmentation of the licensing right is, the higher the possibility for the occurrence of the tragedy of the anticommons is.

In reality, the entrepreneur may bribe each licensing agency to get its project approved. However, a licensing agency may deliberately hold back its excluding right or suspend the project to ask for a higher bribery later. Meanwhile, the government officials in charge of the licensing process are running at the risk of getting caught as criminals by accepting the bribes and thus may ask for more compensations of their risk. Consequently, it is still difficult for an entrepreneur to get its project approved by each licensing agency even with bribery, due to the high transaction costs of bargaining, heterogeneous interests and asymmetric information among entrepreneurs and licensing agencies, and cognitive biases of the licensing agencies. It is not an exaggeration that entrepreneurs may have to spend months, or even years, to obtain the necessary approvals to start their investment projects (if not completely rejected), leading to another case of tragedy of the anticommons.

Since the 18th National Congress of the Communist Party of China, the Chinese government has been reforming the current bureaucratic licensing system to significantly reduce the number of administrative approval and licensing items, which should be good news to ease the tragedy of the anticommons in enterprise licensing.

Other Examples of Tragedy of the Anticommons

Besides of the above typical examples, there are many other cases of anticommons in reality. As argued by Hunter (2003), the tragedy of the anticommons does not just happen in the physical world but may also occur in cyberspace. Since the opening of the Internet to commercial exploitation in 1992, commercial operators have grown exceedingly fat, relying on the public character of the Internet and their easy access to the vast public commons on the Internet that was created before they ever arrived. However, by carving out remarkable new property rights online, the commercial operators are enclosing cyberspace. They have set up barriers to the use of their online resources, which may erode cyberspace's public commons. To obtain access to some online resources that are all necessary to create a new online innovation or product, a user may need the permissions from a bunch of existing commercial operators, each of whom is endowed with an excluding right. The difficult negotiations with each excluder may lead to a tragedy of digital anticommons.

Tragedy of the anticommons may also occur in the construction of municipal infrastructures and new building or during the urbanization process in transitional economies such as China. During the process of urbanization, or for the sake of improving municipal infrastructures, it might be necessary to construct new buildings in cities, roads, railroads, and similar transportation arteries, as well as other public facilities. Although the benefit to society from these constructions may be substantial, every single property owner along the way of construction must agree on the plan. This provides conditions for the tragedy of the anticommons, as even if hundreds of owners agree, a single landowner can stop the new construction, e.g., the road or railroad. The ability for one person to veto the construction drastically increases the transaction costs and thus may cause tragedy of the anticommons. In some cases when negotiations with each excluder are extremely difficult, even some illegal violence instead is used to drive the "nail household," i.e., the ones who insist on staying without accepting the proposed compensations, out of the way, leading to brutal conflicts and other social tragedies.

Similarly, Heller (2008) indicates that the rise of the "robber barons" in medieval Germany was the result of the tragedy of the anticommons. Nobles commonly attempted to collect tolls on stretches of the Rhine passing by or through their fiefs, building towers alongside the river and stretching iron chains to prevent boats from carrying cargo up and down the river without paying a fee. The multi blockers along the river thus provided conditions for the tragedy of the anticommons.

Types of the Tragedy of Anticommons

Heller (1998) classified the tragedy of the anticommons into two types: legal anticommons and special anticommons. In a legal anticommons, substandard bundles of rights are allocated to competing owners in a normal amount of space, such as a storefront. By contrast, in a spatial anticommons, an owner may have a relatively standard bundle of rights but too little space for ordinary use.

In another way, Parisi et al. (2004) classified the tragedy of the anticommons into simultaneous and sequential anticommons. In the simultaneous case, various right holders exercise exclusion rights at the same time, independently. This may involve two agents linked in a coincident relationship, such as multiple co-owners with cross-veto powers on other members' use of a common resource. As an illustration, we can think of the choice of multiple land owners on whether to contribute their property for a joint venture development. Each parcel of land enters at the same

level as an input of production for the joint development. If all parcels of land are necessary for the realization of the development, the multiple parcel owners would effectively hold a cross-veto power for the realization of the joint project. In the sequential case, the exclusion rights are exercised in consecutive stages, at different levels of the value chain. The various right holders exercise exclusion rights in succession. This may involve multiple parties in a hierarchy, each of whom can exercise exclusion or veto power over a given proposition. As an example, a land owner grants building rights to a third party, who, in turn, grants a partial use right to another. If reunification of the land is desired, a vertical chain of contracting would be necessary. The grantor of the building rights would have to repurchase those rights from his/her grantee; in turn, the grantee of building rights would have to repurchase the partial use rights that he/she granted, and so on. The difficulty in reaching all these sequential repurchase agreements will cause a sequential tragedy of the anticommons. Both simultaneous and sequential anticommons problems are the result of nonconformity between use and exclusion rights.

A Formal Model of the Tragedy of Anticommons

Buchanan and Yoon (2000) proposed a formal economic model to explain the tragedy of the anticommons as a mirror image of the tragedy of commons. In their model, Q measures the quantity of usage, and P measures the average value of a unit product, where there is a linear relationship between P and Q as follows:

$$P = a - bQ \quad (1)$$

where a and b are positive constants. Consider, first, the two-person case, where A and B are to be assigned either (i) usage rights or (ii) exclusion rights. In either case, explicit collusion will allow for attainment of the efficient solution. In collusion, the optimal choice of Q is to maximize the total rent:

$$\underset{Q}{\text{Max}}\, PQ = (a - bQ)Q \quad (2)$$

The solution to this optimization problem is $Q^* = a/2b$, while the corresponding maximum rent (i.e., the total value of the products) is $a^2/4b$.

Now suppose that the required mutual trust is absent and joint action is impossible. In the first case, if each person is assigned a right to use the facility but cannot exclude the other from usage, the interaction will converge on an equilibrium that is analogous to Cournot-Nash duopoly. Person A chooses the level of usage, Q_1, to maximize his/her rent, given person B's choice of usage, Q_2. The rent to person A will be

$$\underset{Q_1}{\text{Max}}\, PQ_1 = (a - bQ_1 - bQ2)Q_l \quad (3)$$

Similarly, person B chooses the level of usage, Q_2, to maximize his/her rent, given person A's choice of usage, Q_1. The rent to person B will be

$$\underset{Q_2}{\text{Max}}\, PQ_2 = (a - bQ_1 - bQ2)Q_2 \quad (4)$$

The Nash equilibrium of the maximization problems (3) and (4) yields $Q_1 = Q_2 = a/3b$, and the total usage $Q = Q_1 + Q_2 = 2a/3b$, which is larger than the efficient usage ($a/2b$) in the collusion case, leading to the overuse of the resource, i.e., the "tragedy of commons."

Alternatively, if each person is assigned a right to exclude but no independent usage right, then he/she can exercise this excluding right by setting the price of his/her ticket independently from the practice of the other owner. Let P_1 denote the price of person A's ticket and P_2 denote person B's ticket. Users are required to secure both tickets by paying the total price $P = P_1 + P_2$ for each unit of usage of the facility. Using Eq. 1, the quantity demanded by the users, Q, will be determined by

$$P_l + P_2 = a - bQ. \quad (5)$$

A Nash equilibrium can be obtained by formulating a game in which each owner tries to maximize

his/her rent by setting his/her ticket price. Person A chooses P_1 so as to

$$\max_{P_1} P_1 Q = P_1\,(a - P_1 - P_2)/b \quad (6)$$

From the first-order condition of this maximization problem, we can express P_1 as a function of P_2:

$$P_l(P_2) = a/2 - P_2/2 \quad (6a)$$

Similarly, we can express P_2 as a function of P_1:

$$P_2(P_1) = a/2 - P_1/2 \quad (6b)$$

Solving the system of simultaneous equations gives a stable solution, $P_1{}^* = P_2{}^* = a/3$. The price a user pays is $P^* = P_1{}^* + P_2{}^* = 2a/3$. The corresponding total usage is $Q = (a - P^*)/b = a/3b$, which is smaller than the efficient usage in collusion case, leading to an underuse of the resource, i.e., the “tragedy of the anticommons.”

Although the above model only includes two owners for simplicity, it can be generalized to any number of owners, while the main conclusion remains unchanged: in the case of commons when each owner is endowed with usage right but without excluding right, there will be overuse of the resource; in the case of anticommons when each owner is endowed with excluding right but without independent usage right, there will be underuse of the resource. It can be further proved that the larger the number of owners is, the less efficient the usage of resources will be, either in the case of commons or in the case of anticommons.

The above model also demonstrates that the tragedy of the anticommons is just a mirror image of the tragedy of commons. The equilibrium in either the multiple-user model or the multiple-excluder model is structurally analogous to that familiar in Cournot-Nash duopoly-oligopoly settings of interfirm competition. Competition among users, on the one hand, or among excluders, on the other, tends to reduce the total rents to the owners, which represents the efficiency of usage in the commons/anticommons model setting.

References

Bretsen SN, Hill PJ (2009) Water markets as a tragedy of the anticommons. William Mary Environ Law Rev 33(3):723–783

Brewer J, Glennon R, Ker A, Libecap G (2008) 2006 presidential address water markets in the west: prices, trading, and contractual forms. Econ Inq 46(2):91–112

Buchanan JM, Yoon YJ (2000) Symmetric tragedies: commons and anticommons. J Law Econ 43(1):1–13

Carey JM, Sunding DL (2001) Emerging markets in water: a comparative institutional analysis of the Central Valley and Colorado-Big Thompson Projects. Nat Resour J 41(2):283–328

Hardin G (1968) The tragedy of the commons. Science 162(3859):1243–1248

Heller MA (1998) The tragedy of the anticommons: Property in the transition from Marx to markets. Harv Law Rev 111(3):621–688

Heller M (2008) The gridlock economy: how too much ownership wrecks markets, stops innovation, and costs lives. Basic Books, New York

Heller MA, Eisenberg RS (1998) Can patents deter innovation? The anticommons in biomedical research. Science 280(5364):698–701

Hunter D (2003) Cyberspace as place and the tragedy of the digital anticommons. Calif Law Rev 91(2): 439–519

Parisi F, Schulz N, Depoorter B (2004) Simultaneous and sequential anticommons. Eur J Law Econ 17(2):175–190

Ying Q, Zhang G (2008) Fragmentation of licensing right, bargaining and the tragedy of the anti-commons. Eur J Law Econ 26(1):61–73

Anti-dumping

Arianna Ciabattoni and Roberto Ippoliti
Department of Management, University of Turin, Turin, Italy

Definition

Anti-dumping measures are represented by differential duties aimed at countering dumping, which is a form of unfair competition by foreign companies achieved through international price discrimination.

The procedure for defining anti-dumping duties is long and complex and takes place under the supervision of the WTO.

Anti-Dumping Under the WTO

Anti-dumping measures – taken by a government or an international organization – are aimed at counteract the dumping practices that occur when a good or service is sold on foreign markets at a lower price, in some cases, even at its production cost. Therefore, dumping is a practice of unfair competition and international price discrimination.

However, anti-dumping measures, which are defensive tools of the national economy, are also susceptible of an offensive use by the country that imposes them and, for this reason, are regulated within the World Trade Organization (WTO) by art. VI of the General Agreement on Tariffs and Trade 1994 (GATT 1994). The WTO Member States may therefore apply these measures only under the circumstances provided for in Article VI and pursuant to investigations initiated and conducted in accordance with the provisions of the Agreement on Implementation of Article VI of the GATT 1994.

In particular, for the purpose of this Agreement, a product is to be considered as being dumped if it is introduced into the commerce of another country at less than its normal value, i.e., the export price of the product exported from one country to another is less than the comparable price, in the ordinary course of trade, for the like product when destined for consumption in the exporting country. The Agreement on Implementation sets out further criteria for the determination of dumping in cases where it is more difficult to quantify the normal value of a product.

This practice is to be condemned, in force of the Article VI, if it causes or threatens material injury to an established industry in the territory of a contracting party or materially retards the establishment of a domestic industry, referring to the domestic producers as a whole of the like products. The determination of prejudice is based on positive evidence and involves an examination of the volume and the effect on prices of the dumped imports in the domestic market for like products, and the impact of these imports on domestic producers of such products.

The initiation of an investigation to determine the existence, degree, and effect of any alleged dumping takes place normally upon a written application filed by the national industry or on its behalf, which, for the purposes of its validity, is necessarily supported by evidence of:

- Dumping
- Injury within the meaning of article VI of GATT 1994 as interpreted by its Implementation Agreement
- A casual link between the dumped imports and the alleged injury

The same evidence is required from the authorities concerned when, in special circumstances, they decide to initiate an investigation without having received the aforementioned written application.

It is up to the authorities to examine the adequacy and accuracy of the supporting material attached to the application in order to justify the initiation of an investigation which must be completed within 1 year, and in any case not later than 18 months from its opening.

All parties involved in an anti-dumping investigation are notified of the information required by the authorities and have ample opportunity to present in writing all the evidence they consider relevant, as well as the power to acquire non-confidential information. Exporters or foreign producers receiving an anti-dumping investigation questionnaire have a time limit for the reply of at least 30 days, which may be extended. In order to verify the material received or to obtain further details, the authorities may carry out the necessary investigations in the territory of other Members, if they obtain the consent of the Government and of the companies concerned.

For the duration of the investigation, interested parties are allowed to defend their interests. To this end, the authorities organize, upon request, meetings with the opposing parties to start a possible contradictory.

In order to neutralize or prevent dumping, each Contracting Party may receive, on any dumped product, an anti-dumping duty not exceeding the dumping margin for that product.

As a general rule, the authorities should determine the dumping margin on a case-by-case basis for each exporter or producer of the product under investigation.

Provisional or definitive measures may be authorized if the investigation is successful. The former takes the form of a provisional duty or, preferably, a guarantee – deposit in cash or security – equal to the amount of the provisional anti-dumping duty, which must not exceed the provisionally estimated dumping margin.

The decision to introduce an anti-dumping duty in cases where all the conditions are met, as well as to impose an anti-dumping duty equal to or less than the dumping margin, is taken by the authorities of the importer Member. Once applied to any product, the duty shall be levied, for the amount appropriate to the case and without discrimination, on all the relevant imports considered to be dumped and causing injury, whatever their origin.

An anti-dumping duty remains in force for the period of time and to the extent necessary to neutralize the dumping which is causing the injury. Any definitive anti-dumping duty should be revoked no later than 5 years after its imposition.

The European Anti-Dumping System

Council Regulation (EC) No 1225/2009 of 30 November 2009, and subsequent amendments, concerning the defense against dumped imports by countries not members of the European Union, transposes into the European law the anti-dumping rules contained in the Agreement on Implementation of Article VI of the GATT 1994. Specifically, the legislation contains provisions concerning the calculation of dumping, the procedure for the opening and subsequent conduct of investigations, for the imposition of provisional and definitive measures, and for the duration and review of anti-dumping measures.

In relation to anti-dumping duties, the European Union can choose to adopt one or more of the three basic forms:

- Ad valorem duty, the most frequent duty form, the amount of which is established in proportion to the net price of the goods at the EU border.
- Specific duty, which refers to the physical structure of the goods (weight, length, capacity, volume) and is related to it as a fixed value, e.g., 100 euros per ton of product.
- Variable duty, calculated on the basis of representative prices (the levy is in this case a variable element that can adapt to fluctuations in the international price).

The investigation to determine the existence, degree, and effect of the alleged dumping practices is opened following a written complaint lodged by any natural or legal person, or any association not having legal personality, acting on behalf of the EU industry. The complaint is considered to have been presented by the EU industry or on its behalf if it is supported by EU producers, which together account for over 50% of European production. The complaint can be brought to the Commission or to a Member State which sends it to the Commission, and it must contain evidence of the existence of dumping, injury, and causal link between the allegedly dumped imports and the alleged injury. It is examined in the Advisory Committee, composed of representatives of each Member State, and chaired by a representative of the Commission. If, after consulting the Committee, the evidence is sufficient to justify the initiation of an investigation, the Commission shall commence it within 45 days of the date on which the complaint was lodged. The investigation ends normally within 15 months of its initiation, with the closure or imposition of a definitive anti-dumping measure.

If the existence of dumping and of material injury results from the definitive findings of the facts and the interests of the EU are relevant, the EU Council, acting on a proposal submitted by the Commission after hearing the Advisory Committee, shall impose a definitive anti-dumping duty. As in the case of provisional measures, the amount of the anti-dumping duty must not exceed the dumping margin and should be less than this margin, if a lower amount is sufficient to eliminate

the injury caused to the EU industry. The duty shall be established for the amount appropriate to each case and without discrimination on imports of dumped products. The regulation imposing the duty specifies the amount imposed on each supplier or, if this is not possible, to the supplier country concerned.

The amount of the duty depends on the established value of the dumping margin. This is given by the difference between the normal value and the export price of a product.

For the purpose of determining the normal value of a product and the dumping margin, it is necessary to distinguish between States whose economic system is assessed as being based on a market economy and countries that are not considered as such. The normal value of a product originating in a non-market economy country is identified by referring to a third country governed by a market economy, for example, comparing China to Brazil's production costs. Thanks to this procedure the EU has active anti-dumping duties for about 50 Chinese products.

Anyway, the Section 15 of the Chinese WTO Accession Protocol of China, which attributes it the non-market economy status, expired in December 2016 but EU refused to grant China market economy status. Thus, during its plenary session in Strasbourg on November 15, 2017, the EU Parliament passed a new anti-dumping rule. The ex-ante definition between market or non-market economy disappears: all countries are equal. And China is like the other members of the WTO. But a punctual mechanism is introduced to defend against competitors who distort markets (anyone, not just China). If EU companies or trade unions or other stakeholders have suspicions of possible distortions, they can report to the Commission, which initiates an investigation and draws up a report. If it is established that the distortions are really there, then the old method of the analogue country can be applied.

With two substantial differences compared to the previous system. The first is that it will be possible to be selective, distinguishing also between sectors and not just between countries. It will thus be possible to treat Chinese exports not subject to excessive distortions such as those of any market economy. The second is that the principle of distortion has been considerably enlarged, including social and environmental dumping. It will be argued that there is a distortion if workers are not treated according to ILO criteria or if environmental conventions are not respected. These rules reduce the most unfair sources of unfair competition but open the door to greater randomness in the identification of distortion.

Cross-References

- ▶ European Integration
- ▶ Preferential Tariffs
- ▶ State Aids and Subsidies
- ▶ TRIPS Agreement
- ▶ WTO: Procedural Rules

References

European Council (2009). Council Regulation (EC) No 1225/2009

European Parliament (2017). Legislative resolution of 15 November 2017 on the proposal for a regulation amending Regulation (EU) 2016/1036 and Regulation (EU) 2016/1037

World Trade Organization (1994). Agreement on Implementation of Article VI of the General Agreement on Tariffs and Trade 1994

World Trade Organization (1947). General Agreement on Tariffs and Trade (GATT 1947), Article VI

World Trade Organization (2001). The WTO Accession Protocol of the People's Republic of China (2001), Section 15

Asylum Law

Marc Deschamps[1] and Jenny Helstroffer[2]
[1]CRESE, Univ. Bourgogne Franche-Comté, Besançon, France
[2]BETA, Université de Lorraine, Nancy, France

Abstract

The purpose of this essay is to indicate that asylum law essentially derives from international commitments made in the aftermath of

the Second World War. We then present, by way of illustration, the structure and main mechanisms of the common European asylum system. Finally, we will present the analyses of some of the economic studies dealing specifically with the question of asylum law.

Synonyms

Refugee law

The Economic Analysis of Asylum Law

Immigration is the source of passionate controversies throughout the world among populations and the governments. The issue is tense on all continents, and it has substantial, though varying, effects on electoral results and policies. For example, the results of the Standard Eurobarometer 85 (2016) of May 2016 show that 48% of European Union citizens consider immigration to be the main European problem, thus ranking higher than terrorism (39%), the economic situation (19%), public finances (16%), and unemployment (15%). It is the most frequently cited concern in 20 Member States, with peaks in Estonia (73%) and Denmark (71%). However, at the national level, the issue of immigration ranks second to unemployment, and it is of minor importance when Europeans are asked about the problems with which they are confronted individually.

Three clarifications are necessary here. First, a distinction must be made between immigrants and asylum seekers. Indeed, the latter are by definition person who apply for refugee status on the grounds that he (or she) is at a significant risk to his safety or to his life in his country of origin. Asylum seekers are therefore to be distinguished from persons who migrate primarily for economic or family reasons. Secondly, according to United Nations High Commissioner for Refugees (UNHCR), more than 65 million people worldwide have been forced to flee their homes, including more than 21 million persons who have left their country of origin. Sixty-eight percent are hosted in Africa and the Middle East. Thirdly, the reduction of legal channels of immigration since the 1970s leads to an increasing pressure on asylum mechanisms.

In order to better understand asylum law, we will present the international instruments on which it is based (1), describe the structure of the common European asylum system (2), and present the economic literature on asylum (3).

International Instruments Concerning Asylum Law

The population movements connected with the Russian Revolution, the collapse of the Austro-Hungarian and Ottoman empires, and the rise of totalitarianism at the beginning of the twentieth century led the League of Nations and the International Labor Organization to examine the question of millions of displaced or endangered people. One of the strongest acts was undoubtedly the introduction of a passport allowing stateless refugees to travel, on the initiative of Fridtjof Nansen, the first High Commissioner for Refugees of the League of Nations. However, the four main international instruments that currently define asylum law appeared only after the Second World War.

The Universal Declaration of Human Rights was adopted on December 10, 1948, by the 58 member states which then constituted the United Nations General Assembly. Article 13 provides that “every person has the right to leave any country.” Most importantly, article 14 states in its first paragraph that: “Everyone has the right to seek and to enjoy in other countries asylum from persecution.”

It is nevertheless with the Convention of 28 July 1951 relating to the status of refugees (the so-called Geneva Convention) that a true right to asylum first arises. Indeed, the Geneva Convention still presents the centerpiece of the legal definition of the right to asylum for signatory states, since article 1 defines as a refugee any person who “[...] owing to well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country

of his nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country; or who, not having a nationality and being outside the country of his former habitual residence as a result of such events, is unable or, owing to such fear, is unwilling to return to it." The remainder of the text defines the rights and duties of refugees as States, in particular article 31, under which States undertake not to apply penal sanctions against refugees entering or remaining on their territories without authorization, provided that they report to the authorities without delay and explain their irregular entry or presence.

The Protocol of 31 January 1967 on the Status of Refugees (the so-called New York Protocol) generalizes the scope of the Geneva Convention by lifting all time and space limitations, since initially it was applicable only to events occurring before 1 January 1951 either "in Europe" or "in Europe and elsewhere."

Finally, in order complete the multifaceted picture of asylum law, we must add the United Nations Convention on the Rights of the Child of 20 November 1989. Its article 22 deals specifically with refugees below 18 years of age, whether accompanied or not, and with other provisions relating in particular to family reunification, deprivation of liberty, or criminal protection.

The Common European Asylum System

The question of asylum is an issue that concerns all countries. Here we present the current architecture set up by the European Union to deal with this issue.

According to data provided by Eurostat in April 2016, in 2015, the European Union with 28 members experienced almost twice as many asylum seekers compared with the figure recorded in 1992 with 15 members. According to UNHCR, there were nearly 15,000 deaths solely in the Mediterranean between 2011 and 2016. Faced with the scale of this tragedy and to avoid the destabilization of certain countries of entry (such as Greece or Italy), the European Union decided to implement a common European asylum system (CEAS) at the Tampere summit in 1999. It is outlined in article 78 of the Treaty on the Functioning of the European Union.

After a preparatory period of introduction of minimum standards in European asylum law, the CEAS aims to achieve uniform speed, impartiality, quality, and protection for all those seeking protection in all Member Countries. The objective is to apply the following chronology: (1) the asylum seeker arrives on European territory and faces a harmonized procedure for the entire territory of the EU, (2) he (or she) is granted accommodation and means of subsistence during the processing of the application, (3) if he is at least 14 years old, his fingerprints are recorded and sent to the Eurodac database in order to determine which Member State is responsible for the application; (4) the asylum seeker is granted an interview to determine whether he is eligible for refugee status, subsidiary protection, or temporary protection; and (5) the procedure ends with either a refusal of his application (in which case the claimant is granted an appeal) and an order to leave the territory or the request is accepted and the person is granted a residence permit or citizenship, access to the labor market, and health care.

Legally, the CEAS is a set of legislative texts laying down common standards and procedures. Today it is essentially composed of two Regulations and three Directives: (a) Regulation 604/2013 (known as "Dublin III") laying down the criteria and mechanisms for determining the Member State responsible for the examination of an application for international protection lodged in one of the Member States by a third-country national or stateless person, (b) Regulation 603/2013 creating Eurodac for the comparison of fingerprints of asylum seekers, (C) Directive 2011/95/EU (the so-called Qualification Directive) laying down the conditions for benefiting from protection; (d) Directive 2013/32/EU on the granting and withdrawal of international protection, and (e) Directive 2013/33/EU (the so-called Reception Directive) governing the reception of asylum seekers in EU countries .

These texts are complemented by Regulation 439/2010 establishing the European Asylum

Support Office, whose task is to strengthen cooperation between Member States and to help them in times of crises, as well as Regulation 516/2014 establishing the Asylum, Migration, and Integration Fund.

Asylum in the Prism of Economic Analysis

The treatment of asylum by economists is the subject of both theoretical and empirical and positive and normative analyses. Without providing a complete account, we will now present some of its salient aspects.

Explaining asylum migration flows Unlike labor migration flows, for which expected differentials between countries of origin and destination are major determinants, for explaining asylum migration flows, absolute conditions in the country of origin carry a much greater weight. Empirical studies (see Hatton (2011) for a review) such as Schmeidl (1997) show that push factors, and especially political violence, are particularly important predictors of the generation of refugee flows, unlike economic variables. While this is a feature that distinguishes refugee migration from economic migration, a certain element of choice of the destination is common to both.

Determinants of destination can be grouped into factors influencing the migration path, such as geography and migration costs, and destination country characteristics, such as GDP, refugee stocks, and relevant policies. The latter include visa policies, asylum application recognition rates (see, e.g., Toshkov 2014), the speed of decision on asylum applications, labor market access of asylum applicants and refugees, asylum procedures, acceptance rates of asylum applications, expulsion rates, and living conditions of asylum applicants. Different combinations of these migration determinants have been analyzed in theoretical models of destination country choices by Czaika (2009), Schaeffer (2010), and Djajić (2014). However, Keogh (2013) shows that destination country-specific characteristics explain over 70% of the overall variation in asylum applications.

Asylum lawmaking Economics scholars have also studied normative policy-making aspect of asylum. Thus, Helstroffer and Obidzinski (2010) develop a theoretical model to predict which actors benefit from moving asylum law from the national to the EU level and from the common European asylum system. They find that in a context of regulatory competition, host states do not benefit from the Europeanization of asylum law because of the loss of discretionary power. Refugees however can benefit from common minimum standards, though not from the latest step of the development, which aims at total harmonization.

Helstroffer and Obidzinski (2014) show that the asylum lawmaking procedure in the European Union, which is based on codecision of the Council and the European Union, favors the institution least favorable to change and gives the Commission crucial influence over the outcome of the lawmaking procedure. It is Kaldor-Hicks inefficient.

Bubb et al. (2011) propose a model in which the 1951 Geneva Convention is a means of resolving the problem of the provision of the global public good of refugee protection between states. They consider two possible policies to resolve the issue of filtering between refugees and economic migrants: reforms that make states less attractive for potential immigrants and monetary transfers. They show that countries with high incomes should subsidize those with low incomes in order to avoid the problem of spatial concentration of refugees in southern states.

Fernandez-Huertas and Rapoport (2015) propose to create a market for exchangeable asylum quotas (i.e., countries participate in the public good of international protection either by visas or monetary contributions) coupled with a matching mechanism linking the preferences of asylum seekers and those of host countries. The study compares its recommendations with the EUREMA (European Relocation from Malta) program initiated in 2009 by the European Commission. The authors conclude that their proposal would be a “perfect instrument” to reallocate asylum seekers in the EU.

Conclusion

The question of asylum has long been the subject of specific analyses by philosophers, legal scholars, political scholars, psychologists, geographers, sociologists, and anthropologists. Since the turn of the century, economists have started to contribute to the understanding of asylum migration flows through empirical study and models of migration options. They have further studied the legislators' responses to asylum flows, with the objective of identifying a system of hosting asylum seekers that is in the interest of both host countries and refugees.

In our view, there are two main avenues to be pursued: firstly, a better understanding of the observable and non-observable characteristics of asylum seekers and their choices (as Alvin Roth pointed out in 2015 Article *Politico*: "refugees are not widgets") and, secondly, to find permanent and socially accepted mechanisms to guarantee the effective existence of the right to asylum. We are thus sure that in the years to come, economic studies will complement those of the other human sciences in order help find a solution to one of the most pressing humanitarian problems of our times.

Cross-References

▶ Asylum Law
▶ Court of Justice of the European Union
▶ Immigration Law

References

Bubb R, Kremer M, Levine D (2011) The economics of international refugee law. J Leg Stud 40(2):367–404
Czaika M (2009) The political economy of refugee migration and foreign aid. Palgrave Macmillan
Djajić S (2014) Asylum seeking and irregular migration. Int Rev Law Econ 39:83–95
Fernandez-Huertas M, Rapoport H (2015) Tradable refugee-admission quotas and EU asylum policy. CESifo Econ Stud 61(3–4):638–672
Hatton T (2011) Seeking asylum: trends and policies in the EU. Centre for Economic Policy Research
Helstroffer J, Obidzinski M (2010) Optimal discretion in asylum lawmaking. Int Rev Law Econ 30(1):86–97
Helstroffer J, Obidzinski M (2014) Codecision procedure biais: the European legislation game. Eur J Law Econ 38(1):29–46
Keogh G (2013) Modelling asylum migration pull-force factors in the EU-15. Econ Soc Rev 44(3):371–399
Schaeffer P (2010) Refugees: on the economics of political migration. Int Migr 48(1):1–22
Schmeidl S (1997) Exploring the causes of forced migration: a pooled time-series analysis, 1971–1990. Soc Sci Q 78(2):284–308
Standard Eurobarometer 85 (2016) Public opinion in the European Union, May, Brussels, 221 p
Toshkov DD (2014) The dynamic relationship between asylum applications and recognition rates in Europe (1987–2010). Eur Union Polit 15(2):192–214

Asymmetric Information in Litigation

Shay Lavie
The Buchmann Faculty of Law, Tel Aviv University, Tel Aviv, Israel

Definition

Information asymmetries in litigation refer to situations in which one party is privately informed concerning an issue that is relevant to the outcome of the legal process. In particular, information asymmetries are widely conceived to be an obstacle to settlements. They are prevalent in the legal discourse and have been extensively analyzed in the last decades through game-theoretical models. This entry briefly discusses the theoretical models of asymmetric information in litigation, their underlying assumptions, real-world applicability, and possible avenues to bridge informational gaps.

Introduction

Litigation is costly. We would therefore expect the parties to seek a contractual solution that saves their joint litigation costs – a settlement. Most of the filed cases indeed settle, but some are not (Spier 2007, p. 268). The literature has offered a variety of reasons to explain settlement failures, among

which asymmetric information is considered a major obstacle to settlements.

Asymmetric information in litigation refers to situations in which one party is privately informed concerning an issue that is relevant to the outcome of the legal process. It can stem from many reasons (Spier, p. 272). Plaintiffs may have private information regarding their level of damages; defendants may have private information concerning their fault. Asymmetric information can have various other sources. Parties could be privately informed as to the quality of their lawyers and their capacity to withstand a long trial. Intuitively, under information asymmetries, the uninformed party could not evaluate its case, and trials are more likely.

Theoretical Perspectives

There are several common perspectives to theoretically analyze the effects of asymmetric information on settlements. In a now-classic paper, Bebchuk (1984) modeled asymmetric information as a game in which the uninformed litigant makes a single take-it-or-leave-it offer to the informed litigants. Informed parties decide whether to accept or reject that offer by comparing it to their expected utility from trial, i.e., the expected judgment and the trial costs they likely to incur should they reject the offer. In this so-called screening model, the informed parties whose case is relatively strong – beyond what the single offer represents – reject the offer and go to trial. Similarly, informed litigants whose case is relatively weak would accept the offer. In this sense, the offer creates a cutoff that "screens" different types of informed litigants. The uninformed litigant sets that cutoff according to its gain from the settlement offer and the risk that the offer would be rejected. The screening model can thus predict both the proportion of cases that go to trial and their average type –note that the cases that made it to trial under this model are those in which informed litigants are more confident regarding their case. This basic screening model was developed and extended by other papers (Spier 2007, pp. 273–275, provides a brief survey).

Reinganum and Wilde (1986) present another influential direction to analyze asymmetric information situations. In their model, it is the informed party who makes a take-it-or-leave-it offer to the uninformed, and that offer can "signal" to the uninformed the strength of the informed litigant's case. Reinganum and Wilde's signaling model results in an elegant, fully separating equilibrium. In this equilibrium, informed litigants propose a truthful offer, i.e., an offer corresponding to their expected judgment, and the uninformed mixes between accepting the offer and rejecting. In order to prevent weak informed types from masking as stronger ones, the rate of acceptance that the uninformed type employs in this model should depend on the offer. To demonstrate, where the defendant holds private information, the plaintiff should more frequently reject low offers. Knowing that the plaintiff tends to reject low offers, weak defendants would hesitate to mimic as strong ones, i.e., offer a settlement lower than their type. As before, the basic signaling model predicts both the proportion of cases that fail to settle and their merits, where the cases that made it to trial are not representative of the entire population.

There is now a considerable body of game-theoretical literature that discusses asymmetric information in litigation, and it is important to highlight some of the common assumptions that underlie these models (e.g., Daughety and Reinganum 2014a, pp. 84–86; Bone 1997, pp. 567–571). First, this literature assumes that aside from the private information, all relevant parameters are common knowledge. Moreover, while the uninformed litigant does not know the strength of its rival litigant, it does know the background distribution of the different potential types of its rival, i.e., the probabilities that the rival's case is of a certain strength. This is a nontrivial description, which facilitates the theoretical analysis of asymmetric information. Second, the legal proceedings are assumed to resolve at least some of the informational gaps, for instance, through testimonies or pretrial discovery. However, this revelation process is costly. The foregoing literature could therefore be described as modeling the bargaining that

precedes costly revelation. Hence, the asymmetric information story does not fit situations in which the private information could be conveyed cheaply before trial, e.g., through credible disclosure by the informed or by the uninformed litigant's pre-filing investigation (Bone 1997). There can be a variety of other theoretical complications. For example, most of the papers depict simple situations in which one litigant is informed and the other is not; in reality, both parties can have private information regarding different issues (for instance, Daughety and Reinganum 1994).

Asymmetric Information in Real-World Litigation

Given this discussion, one wonders whether asymmetric information situations are prevalent in real-world settings. As in actuality most of the cases settle, rigorous empirical investigation is quite complicated. One strategy is to focus on the characteristics of cases that made it to trial. As noted above, asymmetric information models predict that the cases that fail to settle are a non-random sample of the entire population. Particularly, the predictions of asymmetric information models could be compared to competing theories of settlements and in particular to the influential models that assume that both parties diverge on their expectations regarding the outcome at trial (e.g., Priest and Klein 1984). Hylton and Lin (2012) survey empirical papers according to this logic and conclude that asymmetric information appears to be relevant in pretrial settlements, though competing theories are also confirmed by the literature. They also caution that "the empirical work so far has to be considered preliminary" (Hylton and Lin 2012, p. 505). There is also some experimental analysis that uses ultimatum games to test the screening and signaling models of asymmetric information in litigation. This literature has generated mixed results, particularly with regard to the signaling setting (where the informed party makes the offer) (Pecorino and van Boening 2016).

Beyond robust empirical findings, the notion of asymmetric information in litigation is quite prevalent in the current legal discourse, at least in the USA. Asymmetric information is sometimes referred to by law professors as "[t]he most important problem in dispute resolution" (Rhee 2009, p. 548). The academic interest in asymmetric information has risen following several decisions of the US Supreme Court that encourage judges to assume a greater gatekeeping role (Bell Atl. Corp. v. Twombly, 550 U.S. 570 (2007)). These precedents direct trial courts to screen out – at the outset of litigation and before proceeding to discovery – cases that do not present a sufficient factual threshold. These stricter rules allegedly deter meritorious filings, particularly in areas where the information about the merits of the case is thought to reside with the defendant – asymmetric information settings. Salient examples are medical malpractice and employment discrimination cases (Hubbard 2016; Bone 2009).

Bridging Informational Gaps

To the extent parties can somehow transmit information, the importance of asymmetric information diminishes. Given the notion that asymmetric information is common, and the recent legal trend to make filing cases harder, one may wonder how parties could bridge informational gaps. A straightforward way to convey information is voluntary disclosure; alternatively, formal, court-supervised discovery proceedings are aimed at forcing the informed to reveal its information (Shavell 1989; Farmer and Pecorino 2005). There are more subtle alternatives. Information can be conveyed indirectly, through the employment of various litigation tactics (Daughety and Reinganum 2014a, pp. 90–92). For example, recent papers have shown how litigants can use various litigation features such as filing for costly injunctions (Jeitschko and Kim 2012) or investing in observable pretrial preparation (Choné and Linnemer 2010) to credibly signal the value of their case. The use of intermediaries such as

attorneys (Leshem 2009) and litigation funders (Daughety and Reinganum 2014b; Avraham and Wickelgren 2014) can likewise play an indirect role in facilitating settlements under asymmetric information. More generally, recent papers attempt to show the extent to which parties can signal information through employing costly signals, e.g., filing fees (Hubbard 2017), or by committing to augment the judgment should the rival party win (Lavie and Tabbach 2017).

Conclusion

Asymmetric information – settings in which one of the parties enjoys better information – is seemingly a prevalent phenomenon in actual litigation. Game-theoretical models constitute an effective analytical tool to understand pretrial bargaining and settlements under informational asymmetries. In the last decades rich literature has developed the basic models, adding a considerable theoretical depth. However, some of the underlying assumptions of the theoretical models seem questionable under the current litigation landscape, and the empirical applicability of the theory to the field is a direction for future research.

Cross-References

- ▶ Access to Justice
- ▶ Information Disclosure
- ▶ Legal Disputes
- ▶ Litigation Decision
- ▶ Signalling
- ▶ Third-Party Litigation Funding

References

Avraham R, Wickelgren A (2014) Third-party litigation funding—a signaling model. DePaul Law Rev 63:233–264

Bebchuk LA (1984) Litigation and settlement under imperfect information. RAND J Econ 15:404–415

Bone RG (1997) Modeling frivolous suits. Univ Pennsylvania Law Rev 145:519–605

Bone RG (2009) Twombly, pleading rules, and the regulation of court access. Iowa Law Rev 94: 873–936

Choné P, Linnemer L (2010) Optimal litigation strategies with observable case preparation. Games Econ Behav 70:271–288

Daughety AF, Reinganum JF (1994) Settlement negotiations with two-sided asymmetric information: model duality, information distribution, and efficiency. Int Rev Law Econ 14:283–299

Daughety AF, Reinganum JF (2014a) Revelation and suppression of private information in settlement-bargaining models. Univ Chicago Law Rev 81:83–108

Daughety AF, Reinganum J (2014b) The effect of third-party funding of plaintiffs on settlement. Am Econ Rev 104:2552–2566

Farmer A, Pecorino P (2005) Civil litigation with mandatory discovery and voluntary transmission of private information. J Leg Stud 34:137–159

Hubbard WHJ (2016) A fresh look at plausibility pleading. Univ Chicago Law Rev 83:693–757

Hubbard William HJ (2017) Costly signaling, pleading, and settlement. University of Chicago Coase-Sandor Institute for Law & Economics Research Paper No. 805; U of Chicago, Public Law Working Paper No. 622. Available at SSRN: https://ssrn.com/abstract=2947302

Hylton KN, Lin H (2012) Trial selection theory and evidence. In: Sanchirico CW (ed) Encyclopedia of law and economics—procedural law and economics, vol 8, 2nd edn. Edward Elgar, Cheltenham/Northampton, pp 487–506

Jeitschko TD, Kim B-C (2012) Signaling, learning, and screening prior to trial: informational implications of preliminary injunctions. J Law Econ Org 29: 1085–1113

Lavie, Shay, Avraham Tabbach (2017) Judgment contingent settlements https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3000218

Leshem S (2009) Contingent fees, signaling and settlement authority. Rev Law Econ 5:435–460

Pecorino, Paul, van Boening, Mark (2016) An empirical analysis of the signaling and screening models of litigation. Available at SSRN: https://ssrn.com/abstract=1988244

Priest GL, Klein B (1984) The selection of disputes for litigation. J Leg Stud 13:1–55. 1984

Reinganum J, Wilde L (1986) Settlement, litigation, and the allocation of litigation costs. RAND J Econ 17:557–566

Rhee RJ (2009) Toward procedural optionality: private ordering of public adjudication. N Y Univ Law Rev 84:514–571

Shavell S (1989) Sharing of information prior to settlement or litigation. RAND J Econ 20:183–195

Spier KE (2007) Litigation. In: Mitchell Polinsky A, Shavell S (eds) The handbook of law and economics. Amsterdam-Oxford, North-Holland, pp 259–342

Audit Committees

Patrick Velte
Leuphana University, Lüneburg, Germany

Abstract

A profound international competition between corporate governance and corporations constitutions systems has been going on since the middle of the last century (La Porta et al. 2002, p. 1147). A basic categorization has been made with regard to the ratio between internal and external corporate governance as well as to the management and supervising structure of publicly owned firms (one-tier and two-tier system). Amongst others, a partial convergence of both constitutional models indicates a high acceptance of audit committees in both systems of corporation's constitutions. However, the committee's competences are different in the one- and two-tier system as well as the main motives of their implementation. Within the two-tier system, the audit committee has been implemented to support and relieve the supervisory board in preparing various tasks. In addition the committee is expected to strengthen corporate governance as a consequence of the high number of supervisory board members. Moreover, the appointment of financial experts as audit committee members is to counteract the lack of respective knowledge in the supervisory board. In contrast, the one-tier system is by trend forcing a stronger personal separation between executive and nonexecutive directors in the board. In addition, major importance is placed on the independence of the committee members in the one-tier system which is usually symptomatic for the separation of functions within the two-tier system. As with the example of audit committees, it becomes clear that both models try to use the advantages of the respective constitutional systems. However, a general superiority of one system cannot be concluded.

The aim of the present analysis is to provide an overview of empirical survey results with regard to the acceptance of audit committees on the capital market and the influence of audit committees on corporate governance. Major attention is paid to a statistically proven relation between certain corporate governance variables and the implementation of audit committees, especially with regard to the independence and financial expertise of its members. The German stock corporation law will be used as an example to demonstrate the importance of audit committees within the two-tier system. Similarly, the US-American capital market with its particular regulations of the stock exchange commission will be used for the one-tier system.

Definition

Audit committees are part of the board of directors (one-tier system) or the supervisory board (two-tier system). They play a major role in modern corporate governance to supervise the management, the risk management (system), and the external auditor. The following analysis gives an overview of empirical survey results with regard to the acceptance of audit committees on the capital market and the influence of audit committees on corporate governance. To compare the different status in the one-tier and the two-tier system, audit committees in German- and US-American-listed companies will be presented.

Normative Analysis

Germany (Two-Tier System)

Implementation

The discussion regarding the implementation of audit committees to enhance corporate governance grew more intense with the "control and transparency law" in 1998. Amongst others, the empirical survey of Coenenberg et al. (1997) supports this relatively young movement deriving from a scientific economic source. The majority of the management board members of the 100 top German corporations in question were not aware

of the necessity to implement audit committees in 1995. In contrast, the survey of Quick et al. (2008) was able to prove an implementation quota of 100% for the DAX30 and 86% for the MDAX. Hence, the present survey results suggest that the majority of the listed stock corporations in the German prime standard have implemented audit committees to strengthen corporate governance.

In contrast to the USA, the German stock corporation law and commercial law have not yet stipulated a general, legally binding obligation for the implementation of audit committees. In fact, the decision to implement audit committees is subject to the autonomy of the supervisory board in terms of § 107 III 1 AktG. This voting right has already been part of the stock corporation law of 1937 and was reinforced by further reform act. The national legislator on purpose did not include the obligation to implement audit committees in order to guarantee highest flexibility with regard to the corporation's management. However, the demand for due diligence of the supervisory board accounting for an appropriate organization of its activities will lead to the implementation of audit committees with rising number of board members. Without audit committees, the necessary intensity of the supervision is no longer ensured. Consequently, the corporations' voting right to implement audit committees becomes redundant with rising number of supervisory board members. Accordingly, this fact gives reasons for the recommendation in the German Corporate Governance Code (GCGC). According to this recommendation, the implementation of qualified audit committees should depend on the specific circumstances of the company as well as its numbers of members. Since its introduction, the GCGC explicitly advises the implementation of audit committees. In case the supervisory board is composed of only three to six members, the prevailing opinion allows for an abandonment of the implementation of audit committees. In this case, no explanation according to § 161 AktG is required, since such small supervisory board usually would not relate to the implementation of an audit committee.

The audit committee has been explicitly mentioned for the first time within the context of the commercial and stock corporation law since 2009. However, a general obligation to implement audit committees is still missing. In principle, only capital market-oriented stock corporations in terms of § 264d HGB that do not have a supervisory board with the respective job specifications are obligated to implement audit committees with at least one independent financial expert. Since all tasks of the audit committee may also be fulfilled by the plenum of the supervisory board, all stock corporations that are legally forced to implement supervisory boards still hold a voting right concerning the implementation of audit committees. Hence, the national legislator relies on the empirically proven high quota in complying with the GCGC.

Job Specification

The matter of independence is implemented in the German stock corporation law in § 105 I 1 AktG. Thus, a member of the audit committee is not allowed to be an active management board member or permanent deputy, authorized officer, or a general agent authorized for the entire corporations management at the same time. In addition and according to the prohibition of crosswise intersection, a member of the audit committee is not allowed to be a legal representative of a dependent company or of another corporation that engages a management board member of the corporation in question in the supervisory board. These regulations are common practice in the German two-tier system. Therefore, the audit committee needs to evolve from the supervisory board, and its members are not allowed to fulfil any managerial functions. In accordance with § 264d HGB, capital market-oriented stock corporations need to appoint at least one independent member in the supervisory board or audit committee. However, this is the only article with regard to the term "independence" so far. In fact, the recommendation of the EU commission of the 15th of February 2005 can be classified as a general guidance. A cooling off period of 2 years for former management board members to become supervisory board members of listed stock corporations is advised in 2009. An exception is granted, if shareholders holding more than

25% of the voting rights of the corporation are in favor of the nomination.

In addition to the stock corporation law standards, the GCGC recommends that supervisory board and hence audit committees should be composed of an adequate number of independent members. Thus, a member is independent if he has no commercial or personal relation to the corporation or its management that accounts for a conflict of interests. Furthermore, the GCGC advises a nomination of no more than two former management board members for the supervisory board. Moreover, the GCGC suggests that the present supervisory board chairman should not take the chair of the audit committee. However, the chairman of the audit committee should be independent. The cooling off period of two years for former management members to become audit committees chairman should be strictly adhered to. A missing compliance with the before-mentioned code suggestions will not account for a justification with regard to the conformity declaration, since the compliance statement only relates to recommendations.

In addition to the independency, the job specification of the audit committee emphasizes on the financial expertise of its members. In terms of § 100 I AktG, no specialist knowledge is mentioned explicitly. However, a minimum level of common, economic, organizational, and judicial knowledge, necessary for understanding and appropriately judging on all regular business transactions unassisted, is demanded (BGH 1982, p. 991). Nonetheless, financial expertise is not mentioned explicitly. At least one member of the audit committee is expected to have the necessary expert knowledge with regard to accounting or auditing (financial expert). Yet, no comment is made on whether and in how far the audit committee chair is to be involved in this part.

Similarly, the GCGC only recommends that the audit committee should be composed of members that are able to fulfil all tasks with the required knowledge, skills, and professional experience at all times. Though, the GCGC provides a detailed job description of the audit committee's chairman. According to this, the audit committees chairman is expected to have special knowledge and experience with regard to the application of accounting standards and internal control procedures. Consequently, the GCGC expects the chairman to be a financial expert, whereas the national legislator only demands for compliance with the legal minimum requirements.

USA (One-Tier System)

Implementation

The implementation of audit committees on the US-American capital market was first recommended in 1939/1949 by the Securities and Exchange Commission (SEC) and the New York Stock Exchange (NYSE). Since corporations did not put this recommendation into effect in the following years, the American Institute of Certified Public Accountants (AICPA) (1967) renewed and enhanced the recommendations of the SEC. Within this context, the composition of audit committee members and their tasks were discussed for the first time. A liability law case in 1968 led to a vote for an obligatory disclosure in the proxy statement with regard to the implementation of audit committee and its members by the SEC (1974). In addition to the name of the members, the disclosure of the number of meetings and their main tasks and responsibilities became obligatory on the 1st of July 1978. Since that time, it became mandatory for all listed corporations at the NYSE to implement an independent audit committee. This was stipulated by the SEC (1978). The American Stock Exchange (AMEX) followed in 1980 and the National Association of Securities Dealers Automated Quotations (NASDAQ) in 2001. In 1987, the results of the National Commission on Fraudulent Financial Reporting became public, also emphasizing the importance of audit committees regarding the corporation's supervision. Within this context, the national commission recommended the implementation of audit committees for all publicly owned firms. The “Blue Ribbon Report” went along with this after a couple of years in 1999. The Sarbanes Oxley Act stipulated an implicit obligation for the implementation of audit committees as permanent committees of the board of directors for all corporations listed at a US stock

exchange. In addition, the job specification of the audit committee's members was described in detail. Opposed to German stock corporation law, the corporations in question do not have an option with regard to the implementation of audit committees.

Job Specification

According to the Sarbanes Oxley Act, all members of the audit committee have to be financially and personally independent of the corporation's management (Section 301). The term independent is applicable only if no direct or indirect corporate or affiliate payment is collected by an audit committee member. The regulations of the Sarbanes Oxley Act are of exterritorial nature. Hence, the rules of financial independence would only hardly be applicable in countries with internal employee participation (e.g., German corporations that are secondary listed at a US-American stock exchange). The codetermination of the supervisory board would be dependent in terms of their salary. In order to preserve the exterritorial effect, the SEC is expecting only managing employees to comply with the rules of financial independence (see Altmeppen 2004, p. 401).

Depending on the stock exchange listing, supplementary regulations of the NYSE, respective the NASDAQ, may apply in addition to the ones of the SEC. According to Section 303 of the Sarbanes Oxley Act, a listing at the NYSE requires the independence of all audit committee members. Thus, an audit committee member is independent if he is not an employee of the (affiliated) corporation currently or has not been for the past three years. In addition his direct relatives are not part of the management and have not been for the past three years (NYSE 2004). With regard to the audit committee member's independence, the NASDAQ demands for an enhancement of the greater SEC criteria, thus demanding that an audit committee member has not participated in the preparation of the annual financial statements as a governing body within the last three years. The Sarbanes Oxley Act does not provide for such cooling off periods after termination of employment. However, as already described above, the German stock corporation law generally arranges for a cooling off period of two years for former management board members to become supervisory board members.

In addition to the requirements of independence, the Sarbanes Oxley Act is demanding for at least one financial expert within the audit committee. Initially, the SEC was interested in stipulating that this person ought to be an expert in terms of accounting. However, in the end they refrained from doing so. In addition to accounting, it is hence acceptable if the expert has knowledge of other finance areas. An exception to this rule may apply if it has been briefly described why no financial expert was appointed as an audit committee member. In general, this is not often the case in order to maintain a good reputation (see Altmeppen 2004, p. 397). The requirement to appoint at least one financial expert is consistent with the amendments of the German stock corporation law. Though in contrast to the German legislator, the SEC is specifying the financial expert qualification in detail. Thus the financial expert is expected to have good knowledge with regard to the preparation of annual financial statement and accounting standards. In addition, he must have the relevant skills to generally judge on the application of accounting policies with regard to estimation, amortization, and the setting up of accruals. Furthermore, he needs to be experienced in the preparation, assessment, analysis, and evaluation of financial statements which are comparable in scope and complexity to the registered corporation's financial statement. Moreover, he is expected to be experienced in actively supervising people that are assigned to the previously described tasks (Section 401 and 407 of the Sarbanes Oxley Act). Such requirements correspond to the job specifications of accountants, finance directors, accounting directors, or similar profession. The Sarbanes Oxley Act does not comment on the qualification of other audit committee members.

In case a corporation is listed at the NYSE, at least one member of the audit committee needs to be experienced in finance and accounting management (NYSE 2004, Section 307). This is consistent with the minimum requirement of one financial expert according to the Sarbanes Oxley

Act. Furthermore, all members need to prove basic knowledge in finance and accounting or are required to be financially literate after a reasonable time. Hence, the NYSE requirements are more demanding than the ones of the Sarbanes Oxley Act with regard to the professional qualification of the audit committee members.

In case a corporation is listed at the NASDAQ, all audit committee members are expected to understand and comprehend the respective corporation's financial statements at the time of their nomination. The regulation with regard to the financial expertise is comparable to the one of the NYSE. In accordance with the regulations of the NYSE, at least one audit committee member is to be experienced in finance and accounting (financial expert). Thus, a professional qualification with regard to accounting or any other comparable experience or basic background knowledge is expected (NASDAQ 2006, Section 4350).

Empirical Relevance of Audit Committees

Capital Costs and Market Reactions

Since only few multivariate empirical studies concerning the impact of audit committees on corporate governance are available for the German capital market (see Velte 2009, 2011; Velte and Stiglbauer 2011), US-American studies have been used primarily. The following explanation provides an overview of current research. According to Ashbaugh et al. (2004), the number of independent audit committee members is related to lower costs of capital. Anderson et al. (2003) empirically proved that audit committees with independent members imply lower interest on debt. In contrast, the results of Bhagat and Black (1999) suggest a lower corporate performance in case the majority of the audit committee members are independent. Similarly, this holds true for the analysis of Klein (1998). Likewise, no statistical significance exists regarding the number of nonexecutive directors and the enhancement of corporate performance.

In addition, the study of DeFond et al. (2005) was addressing the question whether the existence of an accounting expert or a member with any other financial expertise had an influence on the amount of accumulated abnormal return on investment. The results of this study provided a statistical significant positive evidence for an accounting expert. The studies of Wild (1994, 1996) found a significant positive increase of accumulated abnormal accruals, e.g., stock price fluctuation on the statement results.

The empirical results suggest that the implementation of independent and financially literate audit committees provides and increases confidence on the capital market. Hence, the demonstrated attempts of the standard setter regarding the job specifications of audit committee members (independence and financial expertise) are legitimated from an economic point of view for the US-American one-tier system.

Earnings Management and External Management Reporting

An offensive earnings management is sanctioned by the capital market with regard to balance sheet analysis. Hence, a conservative performance is approved. The earnings management performance is measured by means of abnormal accruals. By supervising managing directors, the audit committee is due to provide incentives for the reduction of earnings management.

According to Ebrahim (2007), a significant negative correlation exists between the number of independent audit committee members and the accounting policy, measured by means of abnormal accruals. Xie et al. (2001) analyzed the financial expertise of audit committee members. They were able to prove a significant evidence for a negative influence of investment banking members and nonexecutive directors on the amount of corporations accounting policy, measured by means of discretionary (disproportional) accruals.

Bédard et al. (2004) verified a significant negative influence on the accounting policy, in case at least one audit committee member had the

respective financial expertise. A corresponding relation applies to audit committees with solely nonexecutive directors without substantial corporate integration, provided that the corporate addressees have detailed knowledge of the audit committee's job specification. According to the research of Yang and Krishnan (2005), a significant positive relation exists between the share property of the audit committee members and the amount of nondiscretionary operative accruals. Further studies of Klein (2002) provide evidence for a significant negative correlation between the audit committee's independence and accounting policy in case the audit committee not solely but by majority consists of nonexecutive directors. The respective relation is measured by means of the absolute value of adjusted abnormal accruals.

Other areas of research seek to address the impact of audit committees on the occurrence of subsequent accounting adjustment. Reactive adjustment leads to negative market reactions. From a capital market point of view, they are caused by (intentional) misinterpretations of the corporate management. According to the empirical results of Abbott et al. (2004), the frequency of occurrence of subsequent adjustment of the annual financial statement may be reduced significantly by audit committees solely consisting of independent members and/or the existence of at least one financial expert.

Furthermore, accounting policy is directly influencing quality and quantity of the external management reporting. Hence, by pooling financial expertise, the audit committee fulfils an advisory function to the managing directors. The joint effort is to provide the capital market with the best available management reporting. The survey of Karamanou and Vafeas (2005) proves a significant positive correlation between financial expertise in the audit committee and the frequency, e.g., quality of the management's performance forecast. The results differentiate in how far the corporation responds to negative forecasts ("bad news") and how well they are documented. In addition, attention has been paid to the conformity of corporation information with the analyst's opinion.

Management Fraud

In addition to the impact on accounting policy, empirical corporate governance research is addressing possible consequences of audit committees on the existence and prevention of management fraud. Here, the occurrence of fraud is associated with an intentional erratic behavior of the management and results from information asymmetries between the corporation's management and the capital market. The continuous supervision of the management by the audit committee seeks to increase the likelihood of revealing fraud. In addition, it is likely that the implementation of audit committees may impede the occurrence of accounting fraud preemptively and avoid falsification of the balance sheet by means of due diligence.

In case the submitted financial statement documents are rejected by the SEC in the context of enforcement, negative publicity and damage to the corporation's reputation will be the consequences. According to Abbott et al. (2000), audit committees without continuous employees, holding a meeting for at least twice a year, might be able to alleviate the rejection of the SEC. A corresponding significant negative influence can be verified for audit committees without employees or managing directors having substantial relations to the corporation or its management. These findings are consistent with the research of Krishnan (2005). Hence, an independent and financially literate audit committee reduces the risk of internal control-system failure. However, the corporation is obliged to report on the weakness in case of a change of the auditor. The survey of McMullen (1996) reveals a significant negative correlation between the existence of audit committees and the sanctions of the SEC. According to Beasley et al. (2000), the likelihood of management fraud diminishes with the implementation of audit committees that solely consist of independent members. The sole existence of audit committees leads to a corresponding significant negative influence. The research of Uzun et al. (2004) corresponds with the mentioned empirical findings. Thus, the occurrence of fraud is negatively correlated with the existence of audit committees and, respectively, positively

correlated with audit committees consisting of dependent, nonexecutive directors. These results are supplemented by the research of McMullen and Raghunandan (1996). By trend, corporations with no financial statement fraud have audit committees solely consisting of non-managing directors, i.e., independent audit committees nominating at least one financial expert (e.g., auditor).

External Audit

Amongst others, US-American surveys emphasize on the relation between audit committees and external audit. In addition to the supervision of management and accounting, this activity aims at supervising the external auditor. By continuous monitoring of the auditor's qualification, the audit committee is able to enhance the quality of corporate governance. Amongst others, the relation between compensation of audit and non-audit activities provides a basis for judging on the independence of the external auditor. According to the prevailing opinion, an increase in compensation of audit (non-audit) activities leads to an increase (decrease) in the auditor's independence ceteris paribus. By trend, non-audit activities such as consulting promote the annual auditor's dependence on the management. In addition, they imply the risk of financial side transfers, leading to an inferior audit quality. Hence, the auditor might be willing to grant a concession with regard to the certification of the financial statement he/she might not be granting in case he/she had no consulting mandate.

Carcello et al. (2002) provided evidence for a significant positive relation between audit committees solely or by majority consisting of independent members and the amount of compensation for audit activities of the auditor. According to Abbott et al. (2003a), a completely independent audit committee with respective financial expertise has a positive influence on audit fee. Another survey of Abbott et al. (2003b) concludes that audit committees with solely independent members holding a meeting at least four times a year might reduce the ratio for the compensation of the non-audit activities, since they might endanger auditor independence. Consequently, this implies a significant positive relation between the independence of audit committee members and auditor independence. However, the results of Vafeas and Waegelein (2007) are opposed to the aforementioned findings. Their results suggest a significant positive relation between the requirement of appointing at least one managing director or person being a member of an audit committee of another Fortune 500 corporation into the audit committee and the amount of the audit fee.

Auditor independence serves as a substitute for the audit quality. Within an international framework, it is measured not only by means of the auditor's fee but of the size of the audit company. According to the basic description of the audit theory of DeAngelo (1981), auditor independence and hence audit quality increase with the appointment of international awarded and top-selling audit firms in comparison with other audit and trust companies. Empirical surveys have been addressing possible relations between the implementation of audit committees and the nomination of the annual auditor. If an independent audit committee is responsible for the nomination of the auditor and thus might generate an adequate audit quality in favor of the shareholders, counterproductive intervention of the management is less likely.

The empirical survey of Eichenseher and Shields (1985) already verifies that corporations tend to implement audit committees in case a new auditor needs to be appointed and one of the eight top-selling audit companies is involved. Additional empirically proven relations between audit committees and the external audit refer to the independence of the audit committee members and the likelihood of a cancellation of the auditor's contract. According to Lee et al. (2004), a significant negative relation exists between a solely independent audit committee and the cancellation, e.g., resignation of the audit mandate. The research of Knapp (1987) suggests a significant positive influence of the existence of audit committees on the appointment of one of the eight top-selling companies, the economic situation of the corporation in question, and the likelihood of the board supporting the annual

auditor in case a conflict between auditor and management arises.

The majority of the US-American empirical research could verify a positive influence of audit committees on the quality of external annual audit resulting from the normative approach of the legislator. Until the end of the 1990s of the twentieth century, empirical research was emphasizing only on the existence of audit committees. Later, with the introduction of the Sarbanes Oxley Act, the job specification of the audit committee became more important in terms of empirical research. Attention needs to be paid to the trend that only a cumulative existence of independence and financial expertise leads to significant positive impacts on the amount of the audit fee. The surveys often comply with the normative status quo of the Sarbanes Oxley Act, e.g., all members of the audit committee are independent and at least one member is a financial expert.

Conclusion

Audit committees are of great importance in order to strengthen corporate governance within the US-American one-tier system and the German two-tier system. The comparative normative analysis suggests that audit committees are representative for the convergence of the one- and two-tier system. With regard to the one-tier system, the independent audit committee serves as a monitoring instrument for the managing directors of the board of directors. With regard to the two-tier system, the audit committee is responsible for preparing the plenum's decision. And with the nomination of at least one financial expert, it is ought to counteract the increase of professionalism within the supervisory board. The ideas of the European Commission regarding the job specification of audit committees have been realized in Germany. As a result, independence and financial expertise are of equal importance. This is due to the fact that the EU member states use one- and two-tier systems, therefore demanding the equality of both requirements.

Overall, the requirements for the implementation and job specification of audit committees are more restrictive in the US-American one-tier system. They ought to impede a potential self-assessment of the board of directors. An objective supervision of financial accounting and executive directors is not feasible with dependent audit committee members. Hence, the subject of the member's independence is of major importance within the one-tier system. In contrast, the two-tier system is characterized by a vast separation between managing and supervising tasks. As a result, the requirements for audit committee members are described in detail and more restrictive in the USA. However, the independence of audit committee members might be impaired as well in the two-tier system. The requirements of the German law (at least one independent member in the audit committee) might not be sufficient if a member accepts an additional position in the supervisory board of another corporation of the same industry. This would lead to an increase in risk of conflicts of interests of audit committee members. Though, with the implementation of audit committees, the German two-tier system aims at a professional execution of the supervisory board's tasks by a purposive preparation of the plenum's decision.

The normative concretion has been analyzed along with empirical findings of the international corporate governance research concerning audit committees. Yet, the present empirical results of capital market surveys are primarily based on the US-American one-tier system. With regard to the rising importance of audit committees in the two-tier system, further studies are needed. Emphasis should be placed on the question whether and in how far the implementation of audit committees, including respective job specification, has an actual influence on the improvement of corporate governance. With regard to the one-tier system, empirical results suggest a correlation between the implementation and job specification of audit committees and several corporate governance indicators. Many surveys conclude a significant positive correlation between the nomination of financial experts and independent members in the audit committee and the aforementioned corporate governance variables.

Hence, further studies should address the question whether and in how far the improvement of corporate performance within the one-tier system by the appointment of independent and financial literate audit committee members can be adopted to the German two-tier system. Yet, it needs to be considered that the competencies of the German audit committee cannot be compared to the US-American as a result of the separation between the corporation's management and supervision. By trend, the majority of the respective studies suggest that the US-American capital market has more confidence in corporations with independent and financially literate audit committee members. Thus, the certification of an increase in corporate governance quality might become more realistic. Again, this fact should lead to an increase in research on audit committees within the German two-tier system.

Cross-References

- ▶ Auditing
- ▶ Hedge Fund Activism in Corporate Governance
- ▶ Institutional Review Board
- ▶ Risk Management, Optimal

References

Abbott, Park, Parker (2000) The effects of audit committee activity and independence on corporate fraud. Manag Finance 26:55–67

Abbott et al (2003a) The association between audit committee characteristics and audit fees. Auditing 22:17–32

Abbott et al (2003b) An empirical investigation of audit fees, nonaudit fees, and audit committees. CAR 20:215–234

Abbott, Parker, Peters (2004) Audit committee characteristics and restatements. Auditing 23:69–87

AICPA (1967) Executive committee's statement. JoA 163:10–11

Altmeppen (2004) Der Prüfungsausschuss. ZGR 33:390–415

Anderson RC, Mansi SA, Reeb DM (2003) Board characteristics, accounting report integrity, and the cost of debts. Working paper, Washington, DC

Ashbaugh, Collins, LaFond (2004) Corporate governance and the cost of equity capital. The University of Wisconsin, Madison

Beasley et al (2000) Fraudulent financial reporting. Consideration of industry traits and corporate governance mechanisms. Account Horiz 14:441–454

Bédard, Chtourou, Courteau (2004) The effect of audit committee expertise, independence and activity on aggressive earnings management. Auditing 23:13–35

BGH (1982) Bundesgerichtshof II ZR 27/82 vom 15.11.1982. NJW 36:991–992

Bhagat, Black (1999) The uncertain relationship between board composition and firm performance. Bus Lawyer 54:921–963

Carcello, Hermanson, Neal (2002) Disclosures in audit committee charters and reports. Account Horiz 16:291–304

Coenenberg, Reinhart, Schmitz (1997) Audit committees. DB 50:989–997

DeAngelo (1981) Size and audit quality. JoAE 3:183–199

DeFond, Hann, Hu (2005) Does the market value financial expertise on audit committees of boards of directors? JoAR 43:153–193

Ebrahim (2007) Earnings management and board activity. An additional evidence. Rev Account Finance 6:42–58

Eichenseher, Shields (1985) Corporate director liability and monitoring preferences. J Account Public Policy 4:13–31

Karamanou, Vafeas (2005) The association between corporate boards, audit committees, and management earnings forecasts. An empirical analysis. JoAR 43:453–486

Klein (1998) Firm performance and board committee structure. JoLE 41:275–303

Klein (2002) Audit committee, board of director characteristics and earnings management. JoAE 33:375–400

Knapp (1987) An empirical study of audit committee support for auditors involved in technical disputes with client management. Account Rev 62:578–588

Krishnan (2005) Audit committee quality and internal control. An empirical analysis. Account Rev 80:649–675

La Porta et al (2002) Investor protection and corporate valuation. JoF 57:1147–1170

Lee, Mande, Ortman (2004) The effect of audit committee and board of director independence on auditor resignation. Auditing 23:131–146

McMullen (1996) Audit committee performance. An investigation of the consequences associated with audit committees. Auditing 15:87–103

McMullen, Raghunandan (1996) Enhancing audit committee effectiveness. J Account 179:79–81

NASDAQ (2006) Section 4350 qualitative listing requirements for Nasdaq issuers except for limited partnerships. NASDAQ, New York

National Commission on Fraudulent Financial Reporting (1987) Report of the National Commission on Fraudulent financial reporting, Washington, DC

NYSE (2004) Listed company manual. Section 303A.00 Corporate Governance Standards (Last Modified 3 Nov 2004), New York

Quick, Hoeller, Koprivica (2008) Prüfungsausschüsse in deutschen Aktiengesellschaften. ZCG 3:25–35

SEC (1974) SEC accounting release no 165, 20 Dec 1974

SEC (1978) SEC accounting release no 13-346, 9 Mar 1978
SEC (2002) Proposed rule disclosure required by 404, 406 and 407 of the Sarbanes Oxley Act of 2002, Washington, DC
SEC (2003) SEC accounting release no 33-8820, Washington, DC
Uzun, Szewczyk, Varma (2004) Board composition and corporate fraud. Financ Anal J 60:33–43
Vafeas, Waegelein (2007) The association between audit committees, compensation incentives and corporate audit fees. Rev Quant Finance Account 28:241–255
Velte (2009) Die Implementierung von Prüfungsausschüssen/Audit Committees des Aufsichtsrats/Board of Directors mit unabhängigen und finanzkompetenten Mitgliedern. JfB 59:123–174
Velte (2011) The link between audit committees and corporate governance quality. Corp Ownersh Control 8:5–13
Velte, Stiglbauer (2011) Impact of audit committees with independent financial experts on accounting quality. Probl Perspect Manag 9:17–33
Wild (1994) Managerial accountability to shareholders: audit committees and the explanatory power of earnings for returns. Br Account Rev 26:353–374
Wild (1996) The audit committee and earnings quality. J Account Audit Finance 11:247–276
Xie, Davidson, DaDalt (2001) Earnings management and corporate governance. The roles of the board and the audit committee. J Corp Finance 9:295–316
Yang, Krishnan (2005) Audit committees and quarterly earnings management. Int J Audit 9:201–219

Auditing

Ari Hyytinen[1] and Juha-Pekka Kallunki[2]
[1]Jyväskylä School of Business and Economics, University of Jyväskylä, Jyväskylä, Finland
[2]Oulu Business School, University of Oulu, Oulu, Finland

Definition

In its commonly cited definition, the American Accounting Association defines (financial) auditing as "A systematic process of objectively obtaining and evaluating evidence regarding assertions about economic actions and events to ascertain the degree of correspondence between those assertions and established criteria and communicating the results to interested users."

More broadly, auditing is a mechanism for quality assurance: An auditor is a third-party agent that acts as a certification intermediary and that examines, corrects, and verifies the financial accounts of a business in order to make them more credible to various stakeholders, such as investors.

Audit Regulation and Profession

Auditing is regulated by laws, by global and local guidelines published by professional bodies, and also by professional ethics and practice. Audit legislation varies across countries, but typically, auditing of firms listed in stock markets (public firms) is strictly regulated, whereas that of non-listed (private) firms is either not required or is clearly less regulated. The International Standards on Auditing (ISA) issued by International Federation of Accountants (IFAC) gives guidelines for auditing profession globally.

The audit industry has consolidated over the years, and there are currently four global audit firms. These so-called Big 4 audit firms are the PricewaterhouseCoopers (PWC), Ernst & Young (EY), KPMG, and Deloitte Touche Tohmatsu (Deloitte). Other audit firms are smaller and act more locally.

Accounting and audit failures, including those of Enron, WorldCom, and Tyco International, shaped accounting industry and raised serious questions about the auditors' independence of their clients. One consequence was the adoption of Sarbanes-Oxley Act of 2002 (SOX) in the US, which includes rules aimed to improve audit quality and auditor independence. The issue has evolved during the financial crisis, and consequently, many countries have either suggested or already taken regulatory initiatives to (further) improve auditor independence and audit quality. These initiatives include the separation of mandatory auditing services from consulting services, the restriction of the market share of a given auditor in a country, and mandatory audit rotation, which requires that an audit firm can audit the same client firm only over a limited period of time.

Audit Research

Identifying the determinants of audit fees, which auditors charge from their customers, has been probably the most intensively explored area in audit research. Since the seminal paper by Simunic (1980), many researchers have shown that audit fees are related to client characteristics, such as size, complexity, and risk. For instance, firms with greater financial leverage, lower liquidity, or greater reported losses pay larger audit fees because the perceived audit risk and the required amount of audit work are greater in these firms.

Another frequently cited finding is that clients are willing to pay a premium for the Big 4 audit firms. Some studies also report a fee premium for certain audit offices and audit firms with industry-specific expertise (e.g., Francis et al. 2005).

Much research efforts have also been devoted to examining auditor independence. High audit fees and non-audit services provided by auditors may particularly impair auditor independence. However, many academics and practitioners remind that an incumbent auditor's deep knowledge about its client may actually improve audit quality. Research evidence on the area has not been entirely conclusive. Some recent papers find no evidence that auditors compromise their independence (e.g., Hope and Langli 2010).

Recently, audit research has focused, e.g., on how auditors' personal traits, such as their attitudes toward risk-taking, affect audit outcomes (Amir et al. 2014) and on the effects and possible design of reforms that aim at improving the workings of the auditing industry (e.g., Ronen 2010).

Third-party auditing of the compliance of firms is also common in areas other than financial accounting, such as in the enforcement of product and safety standards and environmental regulation and standards (see, e.g., Dranove and Jin 2010; Duflo et al. 2013).

Cross-References

▶ Information Disclosure

References

Amir E, Kallunki JP, Nilsson H (2014) The association between individual audit Partners' risk preferences and the composition of their client portfolios. Rev Account Stud 19:1

Dranove D, Jin GZ (2010) Quality disclosure and certification: theory and practice. J Econ Lit 48(4):935–963

Duflo E, Greenstone M, Pande R, Ryan N (2013) Truth-telling by third-party auditors and the response of polluting firms: experimental evidence from India. Q J Econ 128(4):1499–1545

Francis JR, Reichelt K, Wang D (2005) The pricing of national and city-specific reputations for industry expertise in the US audit market. Account Rev 80:113–136

Hope O-K, Langli JC (2010) Auditor independence in a private firm and low litigation risk setting. Account Rev 85(2):573–605

Ronen J (2010) Corporate audits and how to fix them. J Econ Perspect 24(2):189–210

Simunic DA (1980) The pricing of audit services: theory and evidence. J Account Res 18:161–190

Austrian Economics

▶ Austrian Perspectives in Law and Economics

▶ Austrian School of Economics

Austrian Perspectives in Law and Economics

Shruti Rajagopalan[1] and Mario J. Rizzo[2]

[1]Department of Economics, State University of New York, Purchase College, Anderson, NY, USA

[2]Department of Economics, New York University, New York, NY, USA

Abstract

This encyclopedia discusses some of the important representative ideas in the Austrian tradition such as spontaneous orders, individuals coping with decentralized knowledge and uncertainty, coordination, and entrepreneurship. The encyclopedia highlights the essential and distinctive feature of the Austrian school of law and economics is its emphasis on both

economic and legal *processes.* The Austrian emphasis on *processes* can be applied to both these branches of law and economics: individual behavior within institutions, as well as individual behavior, leading to the emergence of institutions.

Synonyms

Austrian Economics

Introduction

The Austrian tradition is distinct in emphasizing the role of uncertainty and ignorance of the individual in decision-making. Austrian scholars emphasize that the knowledge in society is fragmented and dispersed across individuals. Therefore, the main problem faced in society is one of coordination and social cooperation. The essential and distinctive feature of the Austrian school of law and economics is its emphasis on both economic and legal *processes.* The Austrian emphasis on *processes* can be applied to both these branches of law and economics: individual behavior within institutions, as well as individual behavior, leading to the emergence of institutions. This encyclopedia discusses some of the important representative ideas in the Austrian tradition such as spontaneous orders, individuals coping with decentralized knowledge and uncertainty, and coordination.

Origin of Institutions

The Austrian law and economics is most closely associated with Nobel Laureate Friedrich Hayek and his work on the origins of legal institutions. Hayek described evolved law as "conceived at first as something existing independently of human will" and distinguished it from legislation, which was "the deliberate making of law" by few individuals (Hayek 2011 [1960], pp. 118–119; and Hayek 1973, pp. 72–73). Hayek's analysis is the direct outgrowth of the earliest Austrian insights as well as those of the Scottish Enlightenment of the eighteenth century. Carl Menger, the founder of the Austrian approach, stated that social scientist must explain "how can it be that institutions which serve the common welfare and are extremely significant for its development come into being without a common will directed toward establishing them" (Menger 1963 [1883], p. 146). Menger's question links the Austrian approach to Scottish enlightenment scholars, who explained the spontaneous orders as "the result of human action, but not the execution of any human design" (Ferguson 1782[1767], p. III, S.2). This line of enquiry has continued in the Austrian tradition where modern scholars like Mario Rizzo and Gerald O'Driscoll ask, "How can individuals acting in the world of everyday life unintentionally produce existing institutions?" (O'Driscoll and Rizzo 1996, p. 20).

Menger argued that the emergence of money is one such example of spontaneous development of institutions. To solve the problem of the double co-incidence of wants, individuals find more highly valued commodities to exchange and therefore, add an exchange value to the use value of these goods, increasing the demand. As more individuals participate in such exchange, they converge to one or two generally accepted media of exchange, which we call money (Menger 1892).

In the same spirit of Menger's explanation for the emergence of money, Ludwig von Mises, one of the most prominent scholars in the Austrian tradition, attempted to explain the emergence of legal rules. Mises argued that property law originally arose from recognition of simple possession and contract law from primitive acts of exchange within localized areas. While the former may have had as its primary motive the avoidance of violence and the creation of peaceful conditions, the latter was almost bound to arise under conditions of de facto property in order to pursue the gains from exchange. But ultimately the world created by these early efforts produced institutions that could be viewed as "a settlement, an end to strife, an avoidance of strife" and thus "their result, their function" is to produce peace within a community (Mises 1981 [1922], p. 34).

Hayek applied spontaneous order analysis, not just to specific legal institutions, but to the entire legal tradition of common law. Hayek described it as "deeply entrenched tradition of a common law that was not conceived as the product of anyone's will but rather as a barrier to all power" (Hayek 1973, p. 84). Scholars have written about the role of litigants, judges, lawyers, etc., in the emergence of common law rules. Some have described the efficiency of common law as a result of private interests of litigants to resolve the dispute. On the supply side, Zywicki (2003) describes the common law system in the Middle Ages as polycentric law-making and attributes the emergence of efficient rules to courts and judges competing for litigants and fees in overlapping jurisdictions. Since the evolution of legal rules and institutions is a continual process, institutional entrepreneurs have an important role in finding opportunities to resolve conflicts and form more efficient, context-specific rules. This may inadvertently give rise, not only to the development of the specific rule, but the entire legal tradition.

Time and Ignorance

Perhaps the most important contribution of Austrian economics, as exemplified especially in the work of F.A. Hayek, is the understanding that individual behavior and social cooperation takes place in the face of decentralized knowledge. However, individuals have limited knowledge, and social knowledge is dispersed or decentralized. Furthermore, this knowledge may not be costless to acquire, or even exist in the form required for decision-making.

In his essay *The Use of Knowledge in Society*, Hayek notes "economic problem of society is thus not merely a problem of how to allocate 'given' resource . . . It is rather a problem of how to secure the best use of resources known to any of the members of society, for ends whose relative importance only those individuals know" (1945, pp. 519–520). So, for Hayek, the function of the law is to provide dispersed economic decision-makers the "additional knowledge" or, more exactly, surrogates for the knowledge necessary to rationally plan (Ibid, p. 521).

The problem of decentralized knowledge and uncertainty is at the very core of the Austrian approach to economics, especially the Austrian approach to law and economics. Karen Vaughn, while describing the overall Austrian approach, wrote, it is "impossible to think of Austrian economics as anything but the economics of time and ignorance" (Vaughn 1994, p. 134). Rizzo clarifies that it is the economics of individuals *coping* with real time and radical ignorance (O'Driscoll and Rizzo 1996, p. xiii). If individuals were not continuously faced with uncertainty and ignorance, a system of legal rules would be quite different. If one takes seriously the fact that all knowledge is decentralized, institutions must have certain characteristics to solve the problem of dispersed knowledge in society.

The knowledge problem may take a static or dynamic form. The static knowledge problem is the utilization of the current stock of dispersed factual knowledge so that individuals can coordinate their actions and plans at a particular time. The dynamic knowledge problem is the growth of new knowledge, not currently in the system, so that improved forms of coordination can occur through time (Kirzner 1973, 1992; Leoni 1991 [1961]). Given static and dynamic knowledge problems, legal institutions are important in the coordination of different individuals at a point in time, and over a period of time.

Ludwig Lachmann argued that legal institutions act as "signposts," because these institutions as they help individuals overcome problems of exchange by enabling individuals to form expectations and coordinate plans. And the most important characteristic of these legal institutions is their stability. He argues that incremental changes of rules, especially in their application to particular new conflicts, must not change the predictable nature of legal rules. "If institutions are to serve us as firm points of orientation their position in the social firmament must be fixed. Signposts must not be shifted" (Lachmann 1971, p. 50). However, this does not mean that a specific legal rule cannot or should not change. It is that the system of rules must be predictable. Lachmann's unchanging

signposts are about the stability of the system, rather than the stability of each particular rule. This echoes Hayek's argument that laws "are intended to be merely instrumental in the pursuit of people's various individual ends. . . . They could almost be described as a kind of instrument of production, helping people to predict the behavior of those with whom they must collaborate, rather than as efforts toward the satisfaction of particular needs" (Hayek 1944, pp. 72–73).

In a world filled with ever-changing and complex interactions, what is required is a simple system of rules to guide individual behavior. Epstein (1995) argues that *simple* legal rules can act as signposts in a complex world that is uncertain and dynamic. As systems go from hierarchical orders to spontaneous orders, the degree of coordination required, and the inherent complexity in mutual compatibility of plans, magnifies. Once it is appreciated that in referring to legal rules we are referring to inputs into *individual* decision-making, it becomes evident that the more decentralized and complex a system is the more critical it is that rules are simple.

Coordination and Optimality

The emphasis on ignorance, decentralized knowledge, and uncertainty leads us to the Austrian emphasis on coordination in society. It is important to understand the difference between coordination and the neoclassical concept of optimality. In the Austrian approach, the focus is on coordination, and not on optimality.

The fundamental meaning of coordination is simply the mutual compatibility of plans. This requires two things. First, each individual must base his plans on the correct expectation of what other individuals intend to do. Second, all individuals base their expectations on the same set of external events (Rizzo 1990, p. 17). In this basic meaning, the existing dissemination of knowledge has led to a state of affairs where each party is able to implement his plans. All offers to buy are accepted by sellers. All offers to sell are accepted by buyers. This is to be distinguished from the *process* of coordination whereby through trial and error learning and entrepreneurial discovery agents are able to make their plans compatible or more nearly compatible with those of others.

Coordination is analytically different, though not incompatible, with the concept of optimality. Pareto optimality implies that individuals exhaust all the potential gains from trade. This is a special case of coordination. However, there can be coordination, or the execution of mutually compatible plans, which do not exhaust all potential gains from trade. "...these plans are mutually compatible and that there is consequently a conceivable set of external events, which will allow all people to carry out their plans and not cause any disappointments" (Hayek 1937, p. 39). A state of mutually compatible plans "represents in one sense a position of equilibrium, it is however clear that it is not an equilibrium in the special sense in which equilibrium is regarded as a sort of optimum position" (Hayek 1937, p. 51). Everyone within a system may have mutually compatible plans, and yet there may be better trading opportunities out there so that at least some parties can improve their positions by alternative trades. Thus, if there is a sense in which the mutual compatibility of plans is an optimum, it is only a local optimum, that is, between the direct parties to an exchange.

In a fully or perfectly coordinated state of affairs, each individual correctly takes into account: (1) the actions being taken by everyone else in the set and (2) the actions which the others might take, if one's own actions were to be different (Kirzner 2000, p. 136). The latter ensures that no buyer transacts at a price higher than that which a potential seller would offer. And that no seller transacts at a price lower than that which a potential buyer would offer. In this sense, the Austrian idea of coordination is compatible with the neoclassical concept of Pareto optimality. If each individual fully takes account of the actions (and potential actions) of every other individual, all courses of action, which might be preferred by any one participant without hurting anyone else, must already have been successfully pursued. In this sense, Pareto-optimality corresponds to perfect coordination (Kirzner 2000, p. 144).

The importance of law to basic and perfect coordination is indirect. Legal rules obviously

cannot affect a state of affairs in which plans are mutually compatible and all arbitrage opportunities are eliminated. Nevertheless, they can, by facilitating exchange and protecting or ensuring the right to entrepreneurial (arbitrage) profit, make the discovery processes that move the system *toward* mutual compatibility and full coordination more likely to be unleashed. Lachmann emphasizes the aspect of institutions that aid the formation of expectations and argues that institutions "enable each of us to rely on the actions of thousands of anonymous others about whose individual purposes and plans we can know nothing. They are nodal points of society, coordinating the actions of millions whom they relieve of the need to acquire and digest detailed knowledge about others and form detailed expectations about their future action. But even what knowledge of society they do provide in highly condensed form may not all be relevant to the achievement of our immediate purposes" (1971, p. 50).

More generally in the field of law and public policy, simply to assume that the lawmakers, paternalist, or central planner each has the relevant knowledge to bring out his stated goals is to assume the knowledge problem away (e.g., Rizzo and Whitman 2009, p. 905). The task is actually the opposite, to *solve* the problem of decentralized knowledge. When lawmakers act on the basis of a pretense of knowledge to which they have no access, they *increase* uncertainty relative to attainment of individuals' goals.

Legal Order

Economic activity takes place within the framework of a "given" legal order. However, some explanation is required to produce clarity about the meaning of such "givenness." Something can be given in the objective sense, in which the legal rules are given to the omniscient observing economist. This is a conceptual expedient for the creation of narrowly specified and limited models. More important is the subjective sense, in which they are given to the individuals whose actions we are trying to explain (Hayek 1937, p. 39). Givenness in this second sense means that the framework, at least insofar as it affects the plans of the individual, is predictable.

However, it is impossible for any system of legal rules to be completely defined, specified, unambiguous, and hence perfectly predictable either in theory or in application. If ignorance and genuine uncertainty is taken into account, then it is problematic to assume a completely specified set of legal rules. While legal institutions may help individuals cope with ignorance in the market, these institutions are themselves subject to the knowledge problem. Hayek emphasized the knowledge problem not only in the context of the market, but extended it to other orders. The language of rules and legal decisions is always characterized by some ineradicable degree of uncertainty or vagueness, and therefore even if it were conceptually possible to define all rules clearly, it would be prohibitively costly (Whitman 2002, p. 6).

Whitman argues that it is inaccurate to see law as a process of one-way causation where a given set of exogenous legal rules resolves conflicts (2002, p. 3). In this area, an important aspect of the Austrian approach to law and economics is to endogenize the system of legal rules. Especially in a legal system in which judges make law by establishing, modifying, overturning, and reaffirming precedents, the actions of participants play a pivotal role in determining the direction of the law. Even when the there is no new legal rule, or a novel application of an old legal rule, the law still changes as a result.

If the legal rules are and constantly evolving, then what do we mean by the "givenness" of rules to the agents in the system? If the rules are constantly challenged and modified, then how do they provide any kind of guidance for human action? And more importantly, how does a constantly evolving system of rules act as a constraint for individual behavior? Within any legal system, there is a tension between the need to produce certainty of the laws with the need for the law to evolve and be relevant to new situations. This is particularly the case in common law. The question is often posed as a tradeoff between certainty and flexibility of the law.

In the first place, at the moment of choice, a certain framework of rules is given. Today's market transactions must be executed within the framework of rights as given today, but that framework is itself the unintended result of the past actions of many individuals. These rules of the game are the "relics" of successful plans of earlier generations that have "gradually crystallized" into institutions (Lachmann 1971, pp. 68–69). Second, legal rules are not being changed in entirety, but the change is marginal. "A change in the law can be marginal in the sense that it is perceived as deviating only slightly from precedent" (Rizzo 1980a, p. 651). Third, Rizzo further argues that certainty of the law and its flexibility are not incompatible and that the "law endures by changing" (1999, p. 499). The law must have a certain plasticity to survive through economic changes. A rigid or static framework would break apart. These considerations imply that the "system" of rules is relatively stable, while marginal changes to specific rules adapt to the new or changing circumstances. This is the idea of the decomposability of the system of rules.

The most important factor that enhances the predictability of law, even as it adapts and changes to new circumstances, is the nature of the process involved. To see this we must distinguish between two forms of coherence in the law. Rizzo (1999) differentiates logical coherence of the law, from the praxeological coherence, or the coordination, which arises from the law. For Rizzo, and the Austrian approach more generally, it is praxeological coherence, or coordination in society, which is at the forefront of analysis. The logical consistency of laws is neither necessary nor sufficient for such coordination.

Another related reason for the emphasis on praxeological coherence of rules is the recognition that there is no one single correct set of legal rules. If the moral intuitions of individuals are not completely consistent or if they have gaps, then there may be more than one right answer in a particular case. All of these may be in the range of expectations of the agents. Presumably, this does not unduly disturb the order of actions as long as the acceptable range is within ordinary limits.

The process of rule-evolution or generation in a common law system is based on trial and error (Hayek 2011 [1960], pp. 122–125). Therefore, at any given point in time some rules or application of rules will simply be wrong. In other words, they will be ripe for revision as the process continues. For Hayek the primary focus is on the overall system. The system is or should be the primary object of normative evaluation. Its mistakes are in a sense simply part of the process.

Entrepreneurship

An important and recurring theme in Austrian economics is the role of entrepreneurial alertness in seizing profit opportunities and thereby enhancing the level of coordination in the market. Krecké argues that the Austrian concept of entrepreneurship is, in principle, applicable to legal decision-making. Decisions on which course to follow in a given case, and on which sources to rely, can be supposed to involve entrepreneurial judgments (2002, p. 8). Legal entrepreneurs, like their counterparts in the market, are alert to the "flaws, gaps and ambiguities in the law" (Krecké 2002, p. 10). Whitman (2002) also extends the idea of entrepreneurship to the role played by lawyers and litigants. He examines how legal entrepreneurs discover and exploit opportunities to change legal rules – either the creation of new rules or the re-interpretation of existing ones to benefit themselves and their clients. Harper (2013) believes that the entrepreneurial approach lays the groundwork for explaining the open-ended and evolving nature of the legal process – it shows how the structure of property rights can undergo continuous endogenous change as a result of entrepreneurial actions within the legal system itself.

The most important differentiating factor separating the entrepreneurship of the market process from legal entrepreneurship is the absence of the discipline of monetary profit and loss in the latter case. Although money may change hands in the process of legal entrepreneurship, its outputs may not be valued according to market prices, especially when there is a public-goods quality to the

rule at issue. Whether effective feedback mechanisms exist in the contexts is therefore an open question. Martin argues that, in such structures, the feedback mechanism in polities is not as tight as feedback in the market mechanism, and therefore, ideology plays a greater role in such decision-making (Martin 2010).

Legal entrepreneurship can be coordinating and yet also increase uncertainty and conflicts in society. It all depends on the kind of legal order in operation and the mechanism by which it is generated and maintained. Rubin (1977) and Priest (1977) originally analyzed how the openly competitive legal process tends to promote economic efficiency.

Conclusion

Austrian scholars extend the themes of ignorance, uncertainty, and fragmented knowledge, to the legal order; and this has important implications for their approach to law and economics. First, the legal rules cannot be assumed to be exogenously given; they must be evolved and discovered through a process. Second, legal systems can become important signposts enhancing expectational certainty, even if they are constantly evolving. And third, entrepreneurship is no longer restricted to within a given set of rules; entrepreneurs also operate in legal and political spheres attempting to create, change, and evolve rules. Finally, through these forces, legal institutions evolve spontaneously without a central mastermind.

Cross-References

- ▶ Austrian School of Economics
- ▶ Common Law System
- ▶ Customary Law
- ▶ Efficiency
- ▶ Hayek, Friedrich August von
- ▶ Liberty
- ▶ Mises, Ludwig von
- ▶ Rule of Law and Economic Performance
- ▶ Smith, Adam

References

Epstein RA (1995) Simple rules for a complex world. Harvard University Press, Cambridge

Ferguson A (1782 [1767]) An Essay on the History of Civil Society. 5th edition. London: T. Cadell

Harper DA (2013) Property rights, entrepreneurship and coordination. J Econ Behav Organ 88:62–77

Hayek FA (1937) Economics and knowledge. Economica 4(13):33–54

Hayek FA (1944) The road to sefdom. Chicago University Press, Chicago

Hayek FA (1945) The use of knowledge in society. Am Econ Rev 35:519–530

Hayek FA (1973) Law legislation and liberty: rules and order (Vol. I, Law Legislation and Liberty). Chicago University Press, Chicago

Hayek FA (2011 [1960]) Constitution of liberty: the definitive edition. Chicago University Press, Chicago

Kirzner IM (1973) Competition and entrepreneurship. Chicago University Press, Chicago

Kirzner IM (1992) The meaning of market process: essays in the development of modern Austrian economics. Routledge Press, London

Kirzner IM (2000) The driving force of the market. Routledge, New York

Krecké E (2002) The role of entrepreneurship in shaping legal evolution. J des Econ et des Etudes Humaines 12(2):1–16

Lachmann LM (1971) The legacy of max weber. The Glendessary Press, Berkeley

Leoni B (1991 [1961]) Freedom and the law. D. Van Nostrand Company Inc., Princeton

Martin A (2010) Emergent politics and the power of ideas. Stud Emergent Order 3:212–245

Menger C (1892) On the origin of money. Econ J 2(6):239–255

Menger C (1963 [1883]) Problems of economics and sociology. Univeristy of Illinois Press, Urbana

Mises L (1981[1922]) Socialism: an economic and sociological analysis. Liberty Fund, Indianapolis

O'Driscoll GP, Rizzo MJ (1996) The economics of time and ignorance. Routledge Press, New York

Priest GL (1977) The common law process and the selection of efficient rules. J Leg Stud 6(1):65–77

Rizzo MJ (1980) The mirage of efficiency. Hofstra Law Rev 8(3):641–658

Rizzo MJ (1990) Hayek's four tendencies toward equilibrium. Cult Dyn 3(1):12–31

Rizzo MJ (1999) Which kind of legal order? Logical coherence and praxeological coherence. J des Econ et des Etudes Humaines 9(4):497–510

Rizzo MJ, Whitman DG (2009) The knowledge problem of new paternalism. Bingham Young University Law Review Issue 4:905–968

Rubin PH (1977) Why is the common law efficient? J Leg Stud 6(1):51–63

Vaughn KI (1994) Austrian economics in America: the migration of a tradition. Cambridge University Press, New York

Whitman DG (2002) Legal entrepreneurship and institutional change. J des Econ et des Etudes Humaines 12(2):1–11
Zywicki TJ (2003) The rise and fall of efficiency in the common law: a supply side analysis. Northwest Univ Law Rev 97(4):1551–1633

Austrian School of Economics

Michael Litschka
Department of Media Economics, University of Applied Sciences St. Pölten, St. Pölten, Austria

Abstract
This entry describes the role of Austrian Economics as a branch of economics which has enriched economic theory with important concepts and theoretical alternatives. After focusing on main representatives and their ideas (specifically uncertainty, entrepreneurship, evolution, insufficient knowledge, and "rules") it stresses some differences to mainstream neoclassical economics and possible applications to the discipline of Law and Economics. It argues that some discussions within Austrian Economics (e.g. social design of rules vs. evolutions of rules, the role of markets and institutions, and a specific understanding of efficiency and rationality) have a direct bearing on this discipline and should therefore be analyzed in more detail.

Synonym

Austrian economics

Definition

A branch of economics focusing on evolution and institutions, which deals with states of disequilibrium, time, uncertainty, and incomplete knowledge and is represented by Austrian and international scholars.

Introduction

The Austrian school of economics is a branch of economics founded by an Austrian economist Carl Menger and his major work "Grundsätze der Volkswirthschaftslehre (sic)" (1871). While mainly known for his contribution to marginal analysis in economic methodology, Menger put forward many other ideas his pupils in Austria and followers all over the world (but mainly in the USA) took over and developed further. Today it seems not clear whether "Austrian" economics and its major theories have been incorporated into the neoclassical mainstream or make up a theoretical stream of its own, like other heterodox approaches (e.g., Post-Keynesianism or new institutional economics). However it can be said that Austrian ideas have changed economic thinking in a specific way and have contributed to a better understanding of issues of uncertainty, causal relationships in economics, disequilibrium, the role of entrepreneurs, evolution, and informational issues in economic life. Moreover, Law and Economics as a discipline has profited from some of the insights of Austrian economics over time.

Representatives and Main Ideas

According to Leube (1995), we can distinguish at least seven generations of scholars in this tradition, beginning with Carl Menger and his pupils Eugen Böhm-Bawerk and Friedrich Wieser and later on comprising popular names like Ludwig Mises, Joseph Schumpeter, Friedrich Hayek, Gottfried Haberler, Fritz Machlup, Oskar Morgenstern, Israel Kirzner, Murray Rothbard, and Ludwig Lachmann. They all shared a methodological understanding of economic science as being based on "individualism," i.e., all actions and intentions can be traced back to individual decisions, and "subjectivism," i.e., all knowledge in economics must come from the subjective interpretation of the environment by individuals. "Human action" (also the title of an important contribution by Ludwig Mises) should stand in the center of any explanation of the social world.

It is important for Austrian economics (AE) that human actions are always determined by insufficient knowledge and the course of time, which is contrary to most of the early neoclassical economics (NE), where complete knowledge plays an important part in theory building (see also section "Differences to and Similarities with Mainstream Economics" in this entry).

In the famous argument about the right methodology for economics (the so-called German debate on methodology), i.e., whether the historical-inductive (collecting single data and historical specifications) or the causal-deductive (constructing general laws by logically deducing from basic axioms) method suits the discipline better, the Austrians favored the last one. In a famous passage (Menger 1871, pp. 25ff.). Menger explained the evolution of money through a process of market forces, where in the end money arises as institutionalized and government approved means of exchange, but without the will of a social planner beforehand. This focus on institutions arising as spontaneous orders in unhampered market processes continued throughout the history of AE, culminating in the work of Friedrich Hayek (1945, 1952) who explained the role of dispersed knowledge within society and the importance of the relative price system as guiding line for individual rational decisions.

The "subjectivist" view embraced by AE, i.e., that correct explanations of human behavior can only be had by analyzing the subjective sense an individual attaches to his or her action, has had implications for the understanding of costs, marginal values, expectations, cause-effect relationships over time, and the meaning of entrepreneurs. As such, this kind of subjectivism is a more general concept than, e.g., utility maximization by rational individuals in NE. This also explains the important tasks of entrepreneurs in the market as explained by Schumpeter, Kirzner, Menger, and the like. The "watchful" entrepreneur seizes opportunities given by the discovery of states of disequilibrium and fills the gap within such states. Subjectivism in this understanding also has some bearing for an understanding of modern rational choice theories which themselves are important for Law and Economics as a discipline. As Law and Economics demands a fairly active role of judges in using economic reasoning to assess legal problems and the consequences of the arising rules, it seems clear that subjectivist analysis (like the evaluation of damages according to marginal cost analysis) is part of that task.

Lastly, the concept of spontaneous orders put forward by Hayek (1973) is a natural consequence of the above said. The system of rules a judge or lawyers have to accept and stick to was brought about by evolution: Rules prevail because they made a group successful; they were not adopted because of that group knowing their effects. Legal positivism (in the sense of deriving all law from the will of a law-making authority) is therefore a constructivist and unacceptable theory for AE; it is important that the common law judge (to name an example) infers general rules from precedents to apply them to new cases and thereby serves the legal order (the customs and rules evolution has developed). Law coordinates the actions of individuals according to general rules, and via this route and the spontaneously arising institutions, it also contributes to the efficiency of groups.

Differences to and Similarities with Mainstream Economics

There are mainstream neoclassical economists who believe that AE and NE share many premises, and therefore AE has been incorporated into NE (e.g., Stigler 1941). Both apply simplified logical models abstracting from complex actual events, both rely on methodological individualism, and both assume rational behavior by individuals. However, whereas NE stresses the predictive power of their theories, AE focuses on the explanatory power in their approaches. Also, when knowledge and rationality are concerned, NE concentrates on the conscious part of decisions, whereas AE deals with the meaning of tacit knowledge in such decisions.

Remarkable differences concern time and uncertainty. In AE, preferences and knowledge can change before a goal is achieved; in NE, knowledge is assumed to be perfect, at least in some defining approaches of the discipline (e.g.,

Gary Becker 1976). As regards the meaning of markets, the main difference is that Austrians concentrate on market processes and the role of the entrepreneur in the creative destruction of equilibria (see, e.g., Joseph Schumpeter's work), while neoclassicals come up with constrained optimization models to prove the states of equilibrium in markets. Applying these insights to Law and Economics, NE poses a stable framework of customary practices and legal rules, whereas AE sees continuous change and arising "spontaneous orders" (Friedrich Hayek) which develop over time and through the rational decisions by all individuals, in a so-called evolution of rules. As such an evolution is by definition an unintentional process, Austrian economists doubt the possibility of applying efficiency criteria (like Pareto or Kaldor-Hicks criteria) to legal rules, at least if they have quantitative implications. They would rather deem efficiency a criterion already implied in evolutionary processes. To give an example, Law and Economics scholar Posner and Austrian economist Hayek did not agree on the role of judges within the legal system, but share an understanding of the law having a built-in correctness (Posner 2005). While Hayek would see this correctness manifest in a spontaneous order, Posner would equal correctness with efficiency and use economic criteria for that task.

Contribution to Law and Economics

How has Law and Economics applied some of the "Austrian" insights mentioned above? If we take as example the methodological tenets of marginalism and subjectivism, one application is provided by the theory of efficient breach of contract. Roughly stated, a person A sells a good to person B but then finds another person C who is willing to pay more for this good. An efficient allocation of resources would require this good to be transferred to C (when there are equal wealth restraints), so A should breach the contract and pay B damages, as long as those damages are not too large (e.g., Posner 1998, p. 133). Other examples would be the protection of property rights as an incentive to produce goods and information, the task of judges to subjectively calculate the efficiency of rules (see above), issues of deregulation, and the fact that courts (and the law) have to mimic market processes whenever the market fails to optimally allocate resources.

While these and other points could be made in favor of a positive influence AE may have exercised on Law and Economics (see Litschka and Grechenig (2010), for a more detailed account), there are also "Austrian" criticisms directed toward a neoclassical understanding of Law and Economics: There are doubts, e.g., as to whether the legislator or court can have the necessary information to postulate an efficiency solution, a claim that Posner or Becker would readily make. Rather, law should provide for space for spontaneous orders which result from human action, but not human design.

Considering the problem of incomplete information and the problem of social design of rules and institutions, how could an "Austrian" version of Law and Economics look like? Generally speaking, it would focus on those institutions best suited to promote decentralized decision making, especially on customs, social norms, and legal rules devised by evolution. The government's task is, among other things, to specify clear and transparent property rights and keep markets free from state influence. Judges and legislators should then fill the gaps in existing rules which have evolved out of social economic activity. To enable people to make their own subjective evaluations of state of affairs is the normative criterion economic policy should adhere to (and not sticking to the usual Pareto criteria of NE) in order to internalize external effects the right way.

Conclusion

As Litschka and Grechenig (2010) argue, it was essential for an economic approach to be accepted in legal theory to show that the law as it should be is already the law properly understood. Hayek's theory of legal evolution constituted one important basis for such a transformation of law, finding its way to, e.g., Posner's theory on the efficiency of

legal rules. Inefficient rules endangering the social welfare of the litigating parties would be challenged by the parties and thereby less likely to survive. Transformations of such a kind allowed Law and Economics to develop as a discipline and enter the daily work of lawyers and judges. To support such normative claims, a positive theory of legal institutions and their evolution, states of disequilibrium, time issues, uncertainty, and informational asymmetries under incomplete knowledge was necessary. This is exactly what (among other important inputs from, e.g., new institutional economics) the Austrian school of economics provided.

Cross-References

- ▶ Becker, Gary S.
- ▶ Choice Under Risk and Uncertainty
- ▶ Coase and Property Rights
- ▶ Economic Analysis of Law
- ▶ Efficiency
- ▶ Hayek, Friedrich August von
- ▶ Institutional Economics
- ▶ Mises, Ludwig von
- ▶ Posner, Richard
- ▶ Rothbard, Murray
- ▶ Schmoller, Gustav von

References

Becker G (1976) The economic approach to human behaviour. The University of Chicago Press, Chicago
Hayek F (1945) The use of knowledge in society. Am Econ Rev 35(4):519–530
Hayek F (1952) Individualismus und wirtschaftliche Ordnung. Eugen Rentsch Verlag, Zürich
Hayek F (1973) Law, legislation, and liberty. A new statement of the liberal principles of justice and political economy, vol 1, Rules and order. Routledge, London
Leube K L (1995) Die österreichische Schule der Nationalökonomie. Texte. Band 1: von Menger bis Mises. Manz, Wien
Litschka M, Grechenig C (2010) Law by human intent or evolution? Some remarks on the Austrian school of economics' role in the development of law and economics. Eur J Law Econ 29:57–79
Menger C (1871) Grundsätze der Volkswirthschaftslehre. Braumüller, Wien
Posner R (1998) Economic analysis of law, 5th edn. Aspen, New York
Posner R (2005) Hayek, law, and cognition. NYU J Law Libert 1:147–165
Stigler G (1941) Production and distribution theories. Macmillan, New York

Avoidance

Jacob Nussim
Faculty of Law, Bar-Ilan University, Ramat-Gan, Israel

Abstract

The term "Avoidance" avoids clear definition in the academic literature. It generally describes illegitimate reactions to legal rules that wholly or partially thwart their legal effect. Self-interested, utility-maximizing individuals are obviously expected to try avoiding legal restrictions and limitations to their behavior. Indeed, examples of avoidance in reality are abundant. Research on avoidance was initiated by the economics of crimes literature in the 1970's and developed since in two additional spinoffs of that literature – environmental economics and economics of tax evasion. These developments are organized and reviewed here. The focus is on the relationship between avoidance and deterrence.

Introduction

Avoidance is not a term of art and no consistent definition exists. It refers to individual reactions to legal rules that circumvent the legal consequences of these rules to some extent. But it does not encompass any type of reaction. Avoidance typically refers only to noncompliant or illegitimate reactions rather than to behavioral changes that are in compliance with the law or considered legitimate reactions to legal rules.

Drawing the line between avoidance behavior and non-avoidance reactions is a challenging task. Individual behavior may face expected legal outcomes that consequently affect individual behavior. Earning income subjects taxpayers to income

taxes, which in turn may induce individuals to reduce their effort to earn income or to underreport their earned income. Driving activity is subject to various safety rules, which in turn make individuals drive more safely, drive less, or try to avoid the traffic police. Regulating the amount of air, water, and soil pollution may limit production and pollution, or it may induce production and pollution processes that are more difficult to observe and monitor. Individuals react to expected positive or negative legal consequences. Whether or not legal rules are regulatory, i.e., intended to change behavior, they trigger individual reactions. In particular, individuals change their behavior to reduce utility-decreasing legal consequences and to take advantage of utility-increasing ones. But not all these reactions are considered avoidance. If the behavioral reaction is intended by the social planner, it is clearly socially desirable. If the reaction is not intended or is not consistent with the goals of legal rules, it is commonly considered socially undesirable and represents a social cost, although it is not necessarily perceived illegitimate or noncompliant. For example, in response to income taxation, individuals may work less, change their tax status (e.g., from a corporation to a partnership), move investments to other countries, or register their business offshore. None of these responses are intended by income taxation; all of them are (locally) welfare reducing, but not all are commonly considered illegitimate. Another example is regulation of air pollution, which may induce polluters to move their operations elsewhere, pollute water rather than the air, misreport the extent of their pollution, change their product lines, litigate with the environmental protection agency, or capture regulators. Again, although none of these responses are intended by pollution regulation, not all of them are commonly considered avoidance.

Economists shy away from this type of discussion and assume certain individual reactions are avoidance. They typically neglect the definitional difficulty and treat avoidance as distinct, assuming that a clear legal dividing line exists between allowed and disallowed responses to legal rules, in other words, that a legal test can be applied. Legal scholars attempt to draw a line between legitimate and illegitimate responses to law, but are yet to offer a satisfactory theory. For example, economists and legal scholars who analyze tax systems adopt various definitions, but provide no justification for their choices.

Ignoring the definitional problem, the literature on avoidance has developed along three largely parallel paths: the economics of crime, environmental economics, and economics of taxation. The last one is concerned with revenue raising and therefore engages in both efficiency and distributive analyses, whereas the former two are regulatory in nature and focus on efficiency. Economists and legal scholars do not necessarily distinguish between enforcing offenses such as theft, murder, speeding, or insider trading and regulatory misconduct such as pollution. All are considered externalities by economists and crimes by legal scholars. But whereas most crimes are controlled through quantity regulation, various other control instruments are considered and used in controlling pollution (see also Shavell 2011). Therefore, the analyses of crime enforcement and of environmental enforcement have also developed in parallel. In the present chapter, I discuss these strands in the literature, focusing on the more recent development of costly avoidance in the economics of crime literature.

The economics literature provides conventional examples of avoidance activities. In the tax literature, misreporting the tax base (tax evasion) is the leading example of costless avoidance. Other tax planning examples, such as revising the tax entity, changing tax location, and non-arm's-length transactions (e.g., transfer pricing), are abundant in the tax avoidance literature (see, e.g., ___tax avoidance papers___). The environmental economics literature also uses the misreporting example in its models of costless avoidance, as well as other examples of costly avoidance, such as changing the mode of operation upon inspection or making surveillance difficult by purchasing private land around the polluting facility (see, e.g., Heyes 1994). In the economics of crime literature, we find examples such as covering incriminating evidence, lying, fleeing the scene of the crime, moving to other countries, committing perjury, destroying

evidence, falsifying financial books or other documentations, concealing assets, hoarding cash, moving money to offshore banks, litigating, and using sophisticated expert advice (e.g., legal, accounting, financial, and medical).

To these examples, we should add the largely independent political economics (of regulation) literature, which generally begins with Stigler (1971). This body of literature is grounded in observations of various activities undertaken by private entities who try to avoid the effects of regulatory rules by influencing politicians, administrators, or the judiciary. (The political economics literature is voluminous and not reviewed here. See, e.g., Persson and Tabellini (2000), Grossman and Helpman (2001), and Besley (2004).) Another general example that may be considered avoidance is substituting away from legally controlled to uncontrolled or lightly controlled behavior (see also Nussim and Tabbach 2008b below). For example, controlling air pollution may encourage replacing air-polluting production activities with water- or soil-polluting processes; punishing the shredding of document for the purpose of obstructing justice may induce new retention policies with more frequent shredding or limited documentation (Chase 2003); and punishing producers who do not reveal safety information to consumers may also dissuade them from collecting such information (Kronman 1978; Farrell 1986).

Economic analysis of avoidance can be organized along several recurring features, emphasized in the following discussion. First, positive versus normative analysis involves analyzing the response of individuals to legal rules against designing an optimal public mechanism to cope with avoidance behavior. Typically, positive models are more refined and may better conform to reality, whereas normative models, for reasons of tractability, are more general. Second, avoidance activities may exhibit various characteristics. They may be costless or costly; avoidance measures may be observable or not and hence punishable and controllable or not; the costs of controlling avoidance vary (e.g., ex ante vs. ex post control); avoidance measures may affect the probability of punishment, its magnitude, the effectiveness of public enforcement effort, etc.; and certain avoidance activities may be more suitable for price controls, while others to quantity or other control instruments. Some of these features are investigated in the literature.

This entry proceeds as follows. The first section "Crime and Avoidance" – sets the stage historically and traces the development of the economic research of avoidance in the tax and environmental literatures. The second section "Tax Avoidance and Tax Evasion" – briefly introduces the distinction between avoidance and evasion in the tax literature and fits it into the literature on avoidance. The following sections "Costless Avoidance" and "Costly Avoidance" – discuss the move from costless to costly avoidance, focusing on public and environmental economics. The next section "Economics of Crime and Costly Avoidance", goes into the recent developments in the economics of crime literature on costly avoidance. Section "Avoidance and Self-Reporting" takes the logical step of connecting avoidance with self-reporting, and the last section on "Avoidance in Private Law" extends the observation and analysis of avoidance into private law.

Crime and Avoidance

The economics of crime began largely with Becker's (1968) seminal work, in which he used economic tools to explain criminal behavior, and welfare economics to design crime enforcement schemes. Generally, incentives to engage in crime are affected by the relative market prices of legal and illegal activities. Crime generates externalities (harm) and hence represents a market failure that results in inefficiency. Therefore, controlling criminal behavior by central planning may be desirable. Crime control relies on several instruments, such as punishment and enforcement effort, changing the opportunity costs of crime, self-reporting subsidies, education, and more (Ehrlich 1973, 1975).

Avoidance was introduced into economic analysis in the 1970s, mainly as part of the economic analysis of crime, environmental

economics, and economics of tax evasion. (Avoidance was also discussed in other types of economic analysis, but not extensively. The emerging political economy literature of the 1970s also dealt, among others, with similar issues.) Although all three types of literature originated in Becker's (1968) deterrence-based framework of the economics of crime, they developed independently, probably because of their different focus of interests. The economics of crime literature is centered on deterrence of externality-producing behavior, whereas the focus of tax and environmental studies is different.

Tax and public economics scholars are interested mostly in optimal taxation and tax-related questions, in particular, inefficiency and distribution. Hence, the economic analysis of tax evasion developed around questions of labor supply (Pencavel 1979; Cowell 1981; Sandmo 1981; Weiss 1976), occupational choice (Pestieau and Possen 1991; Kolm and Larsen 2004 and empirical evidence by Parker 2003; Feldman and Slemrod 2007), tax rates and tax schedules (Srinivasan 1973). See also empirical evidence by Clotfelter (1983), Feinstein (1991), Alm et al. (1993), and Joulfaian and Rider (1996) and experimental findings by Friedland et al. (1978), Alm et al. (1992), and Baldry (1987), and redistribution (see, e.g., Christian 1994). Later, normative tax analysis, i.e., optimal taxation, naturally paid attention to both efficiency and distribution (and even complexity), rather than only to efficiency, as is the case under economics of crime studies (Chander and Wilde 1998; Boadway and Sato 2000).

The treatment of avoidance in environmental economics, although based on a deterrence rationale, was typically related to the choice of control instruments. Pollution can be controlled by several legal instruments, in particular, taxes (i.e., prices), standards (i.e., quantities), or tradable permits. The environmental economics literature examined the effects of avoidance activities, particularly on the choice of control instruments (Downing and Watson 1974; Harford 1978, 1987; Lee 1984; Linder and McBride 1984; Khambu 1990).

This entry focuses on the effect of avoidance on deterrence and hence on the economics of crime literature. I begin by briefly reviewing the tax and environmental literatures and then delve into the economics of crime, situating it in relation to the general development and analysis of avoidance.

Tax Avoidance and Tax Evasion

The tax literature has come up with two arguably distinct terms for illegitimate behavior: "tax avoidance" and "tax evasion." These have become terms of art among tax and public economics scholars. Both are differentiated from taxpayers' legitimate behavior or "real responses," but the dividing line between them is blurred, and therefore the definition of "tax avoidance" remains to be settled. Unlike real responses to changes in relative prices due to taxes, tax avoidance embodies "a variety of tax planning, renaming, and retiming activities whose goal is to directly reduce tax liability without consuming a different basket of goods" (Slemrod 2001, p. 119). Other economists use other tentative definitions (Slemrod and Yitzhaki 2002, p. 1428; Piketty and Saez 2013, p. 417; Mayshar 1991, p. 78; and Sandmo 2005, p. 645), whereas legal scholars and legal reality still differ (Gunn 1978; Isenbergh 1982; Bankman 2000; Weisbach 2002).

The dividing line between tax avoidance and tax evasion is also unclear. Whereas tax avoidance is considered legal, although illegitimate, tax evasion is illegal. Whether or not legality can function as a clear separating criterion (Weisbach 2002), economists fail to see the normative underpinning and hence meaningfulness of such a distinction (Cross and Shaw 1981; Slemrod and Yitzhaki 2002). Other economists define the difference between avoidance and evasion differently (Cowell 1990b).

Both tax evasion and tax avoidance are considered avoidance activities by the economics of crime literature, which can cause some confusion. Both conform to the definition of avoidance adopted in this entry: tax avoidance and tax evasion represent attempts to avoid the legal consequences of behavior. They seem to differ by whether they are punishable (tax evasion) or not

(tax avoidance) (see Nussim and Tabbach 2008b, in the crime enforcement context, and Neck et al. 2012 in the tax context). The punishable/non-punishable divide is related to the tax-originated legal/illegal divide, but the two are not congruent. In any case, the feasibility of sanctioning avoidance activities is important, and it is discussed below.

Another potential difference between tax evasion and tax avoidance is the cost of such behaviors. This difference does not necessarily manifest in reality, and economic modeling is also inconsistent in making such a distinction. Tax evasion is commonly characterized as underreporting of the tax base (e.g., income for revenue-based taxes or emissions for Pigouvian taxes). Underreporting involves either no direct costs or negligible costs. Tax avoidance, by contrast, is commonly described as requiring costly planning, advice, means, or actions. But this distinction is not a sharp one. In practice, cheating (evasion) may also require substantial investment of resources.

In sum, rather than making a decisive distinction between avoidance and evasion, legal reality and economic modeling have taught us that in thinking about avoidance and its impact on behavior and enforcement of legal rules, we should pay attention to the feasibility of sanctioning avoidance and to its private costs.

Costless Avoidance

Costless avoidance is probably best associated with reporting issues. The economics of crime is founded on the built-in assumption of non-reporting: criminals do not self-report their crimes, and therefore enforcement effort is required. The non-reporting of crimes appears to be a costless avoidance measure. Although notice that non-reporting can generate risk-bearing costs, more effective enforcement, or emotional burdens (Polinsky and Shavell 1979; Mayshar 1991). But the literature mostly ignored these costs of non-reporting. Non-reporting was adopted into enforcement-based economic studies of taxation and environmental control and was naturally extended into a continuous measure of avoidance in the form of partial reporting, which is more suitable for these issues (See also Bebchuk and Kaplow 1993 for heterogeneous costless avoidance).

Allingham and Sandmo (1972) and Srinivasan (1973) offered positive models of tax evasion – i.e., non-reporting or concealment of taxable income – which prompted a rich literature on tax evasion (See Surveys by Cowell 1990b; Andreoni et al. 1998; Slemrod and Yitzhaki 2002; Sandmo 2005). Notice that unlike non-reporting of crimes or environmental misbehavior, non-reporting in tax issues constitutes the criminal behavior.) The tax evasion literature is a spin-off of the positive analysis of crime. It is formulated as a portfolio model in which a taxpayer makes allocation choices, the returns to which are partially uncertain (Schmidt and Witte 1984). In these models, a taxpayer must allocate income between legal reporting and illegal evasion or allocate effort between observable-taxable and non-observable-nontaxable income-producing behavior, and at times must also allocate effort between taxable income-producing behavior and nontaxable leisure production.

At its inception, the tax evasion literature focused on the choice of legal and illegal activity in the form of reporting or misreporting taxable income. In light of the costless evasion of taxes – i.e., the costless misreporting of income – it examined the effects of various designs of ex post punishment (or tax rates) and enforcement effort on taxpayers' behavior (e.g., income production, occupational choice, misreporting), assuming different types of risk preferences (Pencavel 1979; Cowell 1981; Sandmo 1981; Weiss 1976; Yitzhaki 1974; Pestieau and Possen 1991), or various enforcement responses by the tax authority (Reinganum and Wilde 1985, 1986). Only later was the tax enforcement literature incorporated into normative models of optimal tax theory. First, the normative analyses assumed away the costs of avoidance, i.e., tax evasion (Sandmo 1981; Cremer and Gahvari 1996), but subsequently the costs of avoiding punishment were accounted for, i.e., tax avoidance (Usher 1986; Mayshar 1991).

The environmental regulation literature also used crime enforcement theory and incorporated avoidance activities. Similarly to the tax evasion literature, environmental studies first considered avoidance in its costless form, as misreporting of pollution levels. Unlike the tax literature, however, the environmental literature focused on comparing environmental control instruments, in particular Pigouvian taxes, quantity standards, and tradable permits, and examined the effect of costless avoidance activities on the choice of control instruments (Downing and Watson 1974; Harford 1978, 1987; Viscusi and Zeckhauser 1979; Jones 1989; Garvie and Keeler 1994).

Costly Avoidance

Assuming that avoidance consumes real resources conforms better to observations of behavior and therefore allows for a much richer set of circumstances. The potential use and effect of avoidance has already been mentioned by Isaac Ehrlich (1972, 1973), who further developed Becker's theory. Diligent readers have noted not only that Beccaria (1764) and Bentham (1789) had laid the groundwork for Becker's (1968) neoclassical economic treatment of crime, but that Beccaria (1764) also acknowledged the likelihood of reactive avoidance behavior (Sanchirico 2006, p. 1350 n. 64). The first serious investigation of Ehrlich's observation is in Malik's (1990) influential study, discussed in greater detail below.

Anticipating, in a way, Malik's (1990) contribution, both the tax avoidance and environmental economics literatures addressed costly avoidance in the mid-1980s. In environmental economics, Lee (1984) studied the positive effects and normative implications of costly avoidance on the design of affluent taxes. Linder and McBride (1984) examined the expected effects of costly avoidance on the choice of control instrument under an agency-hierarchy framework. These studies were followed by further positive analyses in the environmental literature. Khambu (1989) showed that increased regulatory standards may reduce compliance because of engagement in costly avoidance and that larger fines or more stringent enforcement may not improve compliance. He also compared taxes and quantity standards given regulated entities invest in costly avoidance (Khambu 1990). Shaffer (1990) analyzed nonlinear penalties in enforcement of standards in the face of costly avoidance by firms. Nowell and Shogren (1994) examined the effect of costly avoidance on the enforcement of illegal dumping of hazardous waste. Heyes (1994) investigated the difference between thoroughness of inspection and probability of inspection under costly avoidance activity by polluting firms. Huang (1996) modeled costly avoidance and examined its positive effect on the control of externalities, using quantity standards and emission charges.

The tax literature introduced costly avoidance into both positive and normative analyses. Cowell (1990a) studied tax reporting behavior with risky evasion and costly avoidance. Slemrod (2001) incorporated costly tax avoidance into a positive model of labor supply under a linear wage tax, and examined how the availability of tax avoidance activity as a substitute for income-producing effort affects the extent of both tax avoidance and labor. Usher (1986) applied costly avoidance and evasion to a normative model and focused on their effect on the marginal costs of funds, and hence on investment in public goods. Mayshar (1991) demonstrated the normative consequences of costly tax avoidance in a cost-benefit framework in terms of its effect on the marginal costs of public funds.

Kaplow (1990) and Cremer and Gahvari (1993) analyzed the effect of costly tax evasion and enforcement using optimal commodity tax setups. Both studies focused on optimal taxation rather than the economics of crime enforcement, and hence the effect of evasion on enforcement was not emphasized. They showed, inter alia how costly evasion affects the optimal mix of taxes and enforcement effort. Their analyses, however, implicitly reveal a Malik-like effect of costly evasion on enforcement. Kaplow (1990) modeled enforcement of evasion with no specific penalties, because in his model taxpayers face no enforcement uncertainty: they know in advance whether they are going to be audited and pay taxes or not

audited and evade them. (This feature in Kaplow's model limits its analogy with crime enforcement. To put it in the terms of crime enforcement and accordingly embed uncertain enforcement in the model, taxpayers either evade taxes successfully and do not pay taxes, or they pay the same tax rate whether they do not evade or evade unsuccessfully.) Larger investment in enforcement reduces evasion, whereas lower private marginal costs of evasion increase its rate. Kaplow showed that a higher tax rate increases wasteful investment in evasion and therefore should be lowered. (Note that Kaplow's model ignores the social value of tax revenue (i.e., production of public goods); it parallels the common assumption in the economic analysis of crime enforcement.) He additionally showed that enforcement effort may also be limited by potential inducement of additional evasion.

Cremer and Gahvari (1993) modeled uncertain enforcement in a crime-like fashion, but they collapsed the expected penalty into an expected tax rate and therefore did not differentiate between their effects. They also collapsed the costs of evasion into product prices and exogenously limited the penalty rate, assuming away Becker's (1968) main result to begin with. As a result, we cannot draw clear results about the effect of penalties on evasion from their model. But they generally showed that a higher tax rate, which may stand for a higher penalty, increases evasion and, hence, should be limited. Cremer and Gahvari (1994) are similar, but in an optimal income tax framework. Costly evasion is socially wasteful and as such affects optimal tax rates and enforcement. Again, the penalty for evasion was treated as constant.

The normative analysis of tax avoidance is either rather abstract or quickly becomes quite complicated. The reason is that unlike the economics of crime literature, which focuses on efficiency only, optimal taxation theory also necessarily involves social preferences for the distribution of utility (Slemrod 1994; Cremer and Gahvari 1994). Indeed, enforcement of tax evasion and avoidance may affect distributive outcomes (Reinganum and Wilde 1985; Slemrod and Yitzhaki 1987; Border and Sobel 1987; Cremer and Gahvari 1996; Chander and Wilde 1998).

Economics of Crime and Costly Avoidance

Isaac Ehrlich (1972, 1973), one of the founders of the economics of crime literature, identified the potential effect of crime control on individual incentive to avoid punishment, and hence on deterrence: "an increase in [punishment]... will generally increase an offender's incentive to spend resources on self-protection... This may decrease the probability of his being apprehended and punished which in turn may at least partly offset the deterrent effect of [punishment]" (Ehrlich 1972, p. 266). (Using the terminology of Ehrlich and Becker (1972), avoidance measures that reduce the probability of punishment are called "self-protection," and measures that diminish the magnitude of punishment are called "self-insurance.") But his observations did not attract the attention of economics of crime scholars and went largely unnoticed in the literature until the 1990s, beginning with Malik's (1990) important work.

Malik (1990) incorporated the costs of avoiding criminal punishment into the basic normative model of enforcement. In his model, sanctioning crime not only deters criminal activity but also induces individuals who engage in crime to invest real resources in reducing the probability of punishment. Therefore, to the extent that individuals engage in crime, their expected investment in avoidance, which is generally considered socially wasteful, should be weighed in the design of enforcement policy, that is, the punishment and enforcement effort. In particular, given avoidance, Becker's prescription of maximum fines is not necessarily socially desirable. Indeed, certain normative analyses of crime and avoidance after Malik also emphasized the required conditions for Becker's prescription under avoidance (Langlais 2008, 2009; Innes 2001). Thus, incorporating avoidance into crime enforcement policy expands its functions. Crime enforcement does not only aim at deterring crime, but is also

responsible for minimizing socially wasteful avoidance. Although more lenient fines are socially costly in terms of (under-) deterrence, they also reduce the incentive to engage in avoidance, which saves socially valuable resources.

After Malik, the economics of crime literature expanded and revised Malik's analysis in several ways, incorporating different modeling or assumptions. Avoidance is analyzed using normative models similarly to Malik (Langlais 2008, 2009; Stanley 1995a; Innes 2001; Tabbach 2009) or by adopting positive frameworks (Nussim and Tabbach 2008a, b, 2009; Friehe 2011; Baumann and Friehe 2013). The latter typically allow for a more accurate description of reality at the expense of having no rigorous normative conclusions. Various characteristics of avoidance have been studied. For example, Malik (1990), Langlais (2008), and Friehe (2011) assumed the existence of a self-protection measure of avoidance (or partial self-protection in Langlais (2009)), which is non-punishable. Nussim and Tabbach (2009) allowed for both self-protection and self-insurance measures of avoidance, which is again non-punishable. Stanley (1995b), Nussim and Tabbach (2008a), and Sanchirico (2006) assumed that avoidance is observable and punishable, whereas Nussim and Tabbach (2008b) and Baumann and Friehe (2013) assumed it is only partially observable. The implications of avoidance for various substitutable public tools have also been examined (Nussim and Tabbach 2009). Naturally, various assumptions reflect reality in different manners and imply different behavioral effects and predictions.

Nussim and Tabbach (2005, 2009) incorporated avoidance activities into the common positive modeling of crime, where individuals choose an allocation of time, effort, or wealth between legal and illegal activities (Ehrlich 1973), and showed that harsher punishment of crimes may encourage rather than deter criminal activity. (See also Stanley (1995b), who shows that under a specific structure of avoidance with increasing returns to scale, increasing the punishment of crime may induce more criminal acts.) The important observation inferred from their model is that crime and avoidance are complements in the sense that more crime increases the marginal benefit of avoidance and thus triggers more investment in avoidance; similarly, larger investment in avoidance reduces the marginal costs of criminal activity and thus further encourages crime. Therefore, although harsher punishment of offenses directly deters potential offenders, it also fosters investment in avoidance, which in turn, indirectly, encourages criminal activity. The final outcome of these two opposing forces may be of more, rather than less crime owing to the harsher punishment. The authors showed that this counterintuitive outcome holds true under various assumptions concerning avoidance and enforcement technology as well as the offender's personal characteristics. Although Nussim and Tabbach (2009) extended the basic positive, well-known model of crime, their counterintuitive results may have confused some scholars. For example, Sanchirico (2012), using a common normative economic model of crime, unfortunately reached erroneous conclusions. Note further that Nussim and Tabbach's (2005, 2009) theoretical analysis can provide an explanation for the mixed empirical results on the deterrent effect of punishment (Nagin 2013; Chalfin and McCrary 2017). Clearly, other explanations may apply as well, such as risk-preferring offenders (Ehrlich 1973), choice of leisure (Schmidt and Witte 1984), or marginal deterrence effects (Stigler 1970).

Friehe (2011) extended Nussim and Tabbach (2009) to include forfeiture of illegal gains. He showed that given avoidance activity by offenders, forfeiting illegal gains, similarly to increasing punishment, may increase crime rather than deter it. The same mechanism and rationale as in Nussim and Tabbach (2009) apply.

The complementarity between crime and avoidance is missing in Malik (1990) because of the way in which his model is constructed, and therefore no indirect effects of enforcement are considered. For example, under his model, deterrence improves with increased punishment. Relatedly, less than maximal punishment and under-deterrence are optimal. Nussim and Tabbach (2009) noted that if complementarity is acknowledged, social optimality may actually entail harsher punishments and over-deterrence.

The reason is that although harsher punishment directly encourages avoidance, it also directly reduces crime, which due to complementarity also discourages avoidance. If the indirect effect of punishment on avoidance is sufficiently strong, over-deterrence is optimal.

The possibility of optimal over-deterrence due to avoidance is also suggested in a normative framework by Langlais (2008) (see also Langlais 2009). Langlais reconsiders Malik's framework by assuming that avoidance and enforcement effort are complementary, whereas Malik assumed that they are independent. This assumption has several consequences. In particular, contrary to what Malik argues, optimal enforcement may produce over-deterrence. Langlais also shows that punishment and enforcement effort may become complements, rather than substitutes, because of avoidance. The reason is that larger punishment induces avoidance, which in turn, given complementarity, induces enforcement effort. (Note that although Langlais' complementarity assumption is not impossible in reality, it is far from being straightforward. It requires that any additional public investment in enforcement increases the marginal effect of an investment in avoidance on the probability of punishment. Unfortunately, Langlais provides no real-life examples to support this assumption. See also a short discussion in Nussim and Tabbach (2005, 2009).)

Nussim and Tabbach (2005, 2009) also showed that investment in avoidance provides a relative advantage to other policies over punishment in deterring crime; these policies are enforcement effort and subsidizing legal alternatives. Enforcement effort deters crime without affecting the complementarity between crime and avoidance, and as such also reduces avoidance. Moreover, certain enforcement techniques may reduce the marginal effectiveness of the avoidance effort, further reducing avoidance, which in turn further reduces crime. For example, investing in the quality of inspections or audits, rather than in their quantity, may reduce the marginal effectiveness of avoidance activities. Following Nussim and Tabbach (2005, 2009), Sanchirico (2006) reiterated the point of the relative advantage of enforcement effort and provided a legal viewpoint and valuable examples of legal procedures. Clearly, legal procedures are not the only relevant set of enforcement activities.

Another policy tool is subsidizing legal alternatives to crime, such as work subsidies, job training, education, and vocational programs. Subsidies to legal work increase the opportunity cost of crime, reducing crime, but have no direct influence on avoidance. Because crime and avoidance are complements, wasteful avoidance is indirectly reduced by such subsidies. Therefore, Nussim and Tabbach concluded that both enforcement effort and subsidies are socially costly, but they can deter crime and save on socially wasteful avoidance activities. Punishment generally requires negligible costs, but it may have a high social cost expressed in crime rate and wasteful avoidance.

The empirical literature on deterrence generally supports the conclusions of these avoidance-driven results, such as enforcement effort and incentives for legal alternatives are better deterrent tools than punishment. Empirical studies of crime deterrence show not only that the effect of punishment on crime is unclear or insignificant but that both enforcement effort and legal alternatives generate consistent deterrent effects.

Langlais (2009) constructed a "self-protection" model under which avoidance activities by offenders, upon apprehension, may only reduce potential forfeiture of benefits gained by crime, but do not affect fines. Based on this assumption, avoidance is clearly independent of fines, and therefore fines are not constrained by potential wasteful avoidance, as in Malik's model. Becker's result of maximum fines is thus restored. Note that Langlais abstracted from offenders wealth and therefore ignored the increase in the offenders' wealth and in the potential fines due to benefits gained from crime. Taking the offenders' wealth into account in this manner would change his results. Put differently, forfeiture of illegal gains is a part of punishment; therefore avoidance of forfeiture should not be different in general from avoidance of fines. See also Friehe (2011). Admittedly, the economic literature typically ignores benefits gained from

crime when considering the offenders' wealth, in particular because benefits added to the offenders' utility (they are consumed immediately) rather than to their wealth. This assumption is usually innocuous (Polinsky and Shavell 2007). But the case is different when benefits can be forfeited and only partially so, as in Langlais's model.

Malik implicitly assumed that avoidance is non-observable and therefore cannot be controlled directly. Several studies, however, examine the effects of punishing or regulating observable avoidance activities. Stanley (1995a) assumed that avoidance activities are costlessly observable and therefore can be directly punished. He showed that it is both possible and socially desirable to completely eliminate avoidance by an appropriate sanction. Note that Stanley (1995a) assumed that offenders are identical and that their choice is binary (either engage or not engage in crime), with no intensive margin of decision.

Nussim and Tabbach (2008a) extended the positive model of crime and avoidance to allow for either ex ante or ex post punishment of avoidance. Ex ante punishment (or regulation) of avoidance can take the form of a tax levied on avoidance activity or of restrictions on its use (e.g., prohibiting its consumption or requiring a license). Ex ante punishment of avoidance is independent of its use in criminal activity or its enforcement. For example, taxing radar detectors is independent of their use in speeding or of any punishment imposed if speeding is detected. Ex post punishment of avoidance is imposed upon detection of avoidance activity, such as perjury, and can be either related or unrelated to enforcement of a "principal" crime. Hence, punishments for detected crimes and avoidance activities can be either independent or interdependent. Such and other descriptions of reality have been investigated by Nussim and Tabbach. They showed first that ex ante punishment or regulation of avoidance is superior to ex post punishment because it necessarily deters both avoidance and crime (through their complementarity, as shown in Nussim and Tabbach (2009)). Furthermore, they showed that punishing avoidance ex post is not trivial. Unless carefully designed, sanctioning avoidance can actually encourage investment in it because, although harsher punishment discourages it by increasing its price, it also increases the return on investment in avoidance if its punishment can be avoided. Moreover, crime deterrence may also be impaired by sanctioning avoidance, in particular, due to complementarity between avoidance and crime. Thus, Nussim and Tabbach delineated the critical features in the design of socially beneficial ex post punishment of avoidance.

In another study, Nussim and Tabbach (2008b) reconsidered the effect of sanctioning avoidance when certain avoidance activities are non-punishable. For example, although financial advice and legal litigation are generally used for legitimate purposes, they may also be used for avoidance purposes in a non-distinguishable manner and therefore are not punishable in practice. The authors showed that if offenders can engage in non-punishable avoidance activities, the control of punishable avoidance is not necessarily desirable. Based on the reasonable assumption that various avoidance activities are substitutes, (although Complementarity is not ruled out), ex ante punishment (i.e., taxation) of avoidance may improve deterrence and even reduce investment in non-punishable avoidance (due to its complementarity with crime). But it may also dilute deterrence and encourage investment in substitutable avoidance. Nussim and Tabbach carefully investigated the scenarios in which these different outcomes can be expected due to ex ante or ex post control of avoidance.

Baumann and Friehe (2013) examined measures that can provide legal consumption utility or facilitate avoidance (e.g., legal and illegal use of DVD writers). Unlike Nussim and Tabbach (2008b), who assumed that these measures are non-punishable, Baumann and Friehe allowed for their control either ex ante or ex post. They reached similar results to Nussim and Tabbach (2008a) on the difference between ex ante and ex post control of avoidance, but showed further that it may also be socially desirable to control/punish such measures. The intuition is that although taxing or regulating these measures distorts the individuals' legal behavior, it may also deter crime. Taxing (subsidizing) measures

that are complementary (substitutable) to crime improves deterrence.

In an informal discussion, Sanchirico (2006) noted that punishing avoidance is not different from punishing the "principal" offense in Malik's framework. In other words, punishing avoidance may induce investment in additional avoidance activities in order to avoid the detection and punishment of avoidance. Given multiple hierarchical avoidance activities, Sanchirico showed that punishing avoidance – and similarly, the principal crime – is not trivial. Although analytically correct, the application in reality of "avoidance of avoidance of avoidance" (and so on) is questionable.

Lastly, whereas the economics of crime literature assumes that avoidance is socially costly and therefore is always undesirable, Tabbach (2009) showed that avoidance may actually be desirable, and, counterintuitively, it should be encouraged in certain circumstances. It is not that avoidance is not socially wasteful – it is. But it may be less socially wasteful than the social costs of punishment. Tabbach showed that if punishment is costly to society (e.g., imprisonment), avoiding punishment saves the costs of imposing punishment, but may still function as a socially desirable deterrent, because avoidance activities impose costs on offenders. In other words, costly avoidance may serve as a substitute for costly punishment; both are costly to offenders and therefore act as a deterrent, but the latter is also costly to society.

Avoidance and Self-Reporting

Strongly related to crime avoidance is the literature on self-reporting. Where law enforcement is required, eliciting self-reporting can be socially valuable for several reasons. First, self-reporting saves on enforcement costs for a given level of deterrence. With self-reporting, the government can monitor fewer individuals for the same probability of detection (Malik 1993; Kaplow and Shavell 1994). Second, self-reporting transforms uncertain punishment into a certain sanction and thus saves on risk-bearing costs (Kaplow and Shavell 1994). Third, self-reporting can improve social welfare where remediation is required after an offense. Self-reporting, then, induces higher incidence of remediation for the same level of deterrence (Innes 1999). Fourth, under heterogeneous probabilities of apprehension, self-reporting can improve deterrence (Innes 2000).

Innes (2001) studied the effects of incorporating self-reporting into Malik's model of avoidance and showed that with costless self-reporting, avoidance can be eliminated, restoring Becker's prescription for maximum punishment. Intuitively, raising punishment to the maximum increases investment in wasteful avoidance, as stressed by Malik, but offering an alternative equal punishment to self-reporters generates self-reporting and therefore no avoidance.

Although not directly discussed by Innes, it appears that there is a strong inherent connection between avoidance and self-reporting. (Stanley (1995b) also made the connection between self-reporting and avoidance, although he did not pursue it rigorously.) Self-reporting indicates a lack of avoidance. Additionally, whereas avoidance is commonly modeled as reducing the probability of punishment, self-reporting increases this probability. Only that self-reporting is typically costless, and it is modeled as a corner outcome in which the probability of punishment equals one, and enforcement costs are zero.

Avoidance in Private Law

There is no reason to restrict the study of avoidance to public law (i.e., taxation and regulation). The rules of private law can be avoided in much the same way as those of public law, the difference being that the interaction is between private entities rather than opposite the government. (Innes (2001) also discussed the application of avoidance and self-reporting to tort law.)

Friehe (2009) incorporated avoidance activities into a unilateral care tort model and showed that injurers' potential investment in avoidance activities makes the negligence regime superior

to strict liability. The intuition is that under an optimal negligence regime, the injurer pays no damages, and therefore he invests either in optimal precautions or in avoidance. Given a low effectiveness of avoidance, the injurer prefers optimal behavior. Under an optimal strict liability regime, the injurer must pay compensation and therefore may invest in both low-effectiveness avoidance measures and lower-than-first-best precautions. (Friehe's (2009) intuition is reminiscent of Buchanan and Tullock (1975).) Friehe showed further that even for high-effectiveness avoidance, a second-best design of a negligence regime can outperform strict liability. The intuition is similar: there is always a care standard that makes precautions cheaper than the consequences of investing in avoidance.

Friehe (2010) further investigated the effects of avoidance on various aspects of an optimal negligence regime. He showed that punitive damages should also account for their expected effect on avoidance, in a manner similar to that of punishment in Malik's crime enforcement model. He further showed that uncertain care standards reduce the incentive to avoid and thus may be superior to certainty. Lastly, Friehe examined the effect of avoidance on the choice of compensation under negligence: full compensation for harm or only harm due to negligence (Grady 1983; Kahan 1989). He showed that full compensation induces avoidance less frequently, but when it does, a full compensation regime produces a larger investment in avoidance and lower levels of care.

Conclusion

Avoidance is ubiquitous in reality. Indeed, we should expect utility maximizing individuals to circumvent legal rules for their own benefit. Economic analyses of avoidance attempt to explain various kinds of avoidance behavior, to predict avoidance reactions, and to suggest normative responses by a social planner. But the avoidance literature in public law and, in particular in private law, is still rather scarce, and further research is indispensable.

References

Allingham MG, Sandmo A (1972) Income tax evasion: a theoretical analysis. J Public Econ 1(6):323–338

Alm J, Jackson BR, McKee M (1992) Estimating the determinants of taxpayer compliance with experimental data. Natl Tax J 45:107–114

Alm J, Bahl R, Murray MN (1993) Audit selection and income tax underreporting in the tax compliance game. J Dev Econ 42(1):1–33

Andreoni J, Erard B, Feinstein J (1998) Tax compliance. J Econ Lit 36:818–860

Baldry JC (1987) Income tax evasion and the tax schedule: some experimental results. Public Finance 42(3): 357–383

Bankman J (2000) The economic substance doctrine. Calif Law Rev 74:5, 13

Baumann F, Friehe T (2013) Regulating harmless activity to fight crime. J Econ 113(1):1–17

Bebchuk LA, Kaplow L (1993) Optimal sanctions and differences in individuals likelihood of avoiding detection. Int Rev Law Econ 13:217–224

Beccaria C (1764) On crimes and punishments. In: Manzoni A (ed) The column of infamy prefaced by Cesare Beccaria's of crimes and punishments (trans: H. Paolucci, 1963). Prentice Hall, Oxford University Press, 1963

Becker GS (1968) Crime and punishment: an economic approach. J Polit Econ 76:169–217

Bentham J (1789/1948) An introduction to the principles of morals and legislation (JH Burns, HLA Hart eds., 1970). (Oxford: Blackwell, 1948 ed.), Athlone Press, pp 158–159

Besley T (2004) Principled agents? Motivation and incentives in government. The Lindahl Lectures (Oxford University Press, 2005)

Boadway R, Sato M (2000) The optimality of punishing only the innocent. Int Tax Public Financ 7:641–664

Border KC, Sobel J (1987) Samurai accountant: a theory of auditing and plunder. Rev Econ Stud 54(4): 525–540

Buchanan JM, Tullock G (1975) Polluters' profits and political response: direct controls versus taxes. Am Econ Rev 65:139–147

Chalfin A, McCrary J (2017) Criminal deterrence: a review of the literature. J Econ Lit 55(1):5–48

Chander P, Wilde LL (1998) A general characterization of optimal income tax enforcement. Rev Econ Stud 65:165–183

Chase CR (2003) To shred or not to shred: document retention policies and federal obstruction of justice statutes. Fordham J Corp Financ Law 8:721

Christian CW (1994) Voluntary compliance with the individual income tax: results from the 1988 TCMP study. In: The IRS Research Bulletin, 1993/1994, Publication 1500 (Rev. 9–94). Internal Revenue Service, Washington, DC, pp 35–42

Clotfelter CT (1983) Tax evasion and tax rates: an analysis of individual returns. Rev Econ Stat 65(3):363–373

Cowell FA (1981) Taxation and labour supply with risky activities. Economica 48(192):365–379
Cowell FA (1990a) Tax sheltering and the cost of evasion. Oxf Econ Pap 42:231–243
Cowell FA (1990b) Cheating The Government: The Economics of Evasion
Cremer H, Gahvari F (1993) Tax evasion and optimal commodity taxation. J Public Econ 50(2):261–275
Cremer H, Gahvari F (1994) Tax evasion, concealment and the optimal linear income tax. Scand J Econ 96(2): 219–239
Cremer H, Gahvari F (1996) Tax evasion and the optimum general income tax. J Polit Econ 60(2):235–249
Cross RB, Shaw GK (1981) The evasion-avoidance choice: a suggested approach. Natl Tax J 34:489–491
Downing PB, Watson WD Jr (1974) The economics of enforcing air pollution controls. J Environ Econ Manag 1(3):219–236
Ehrlich I (1972) The deterrent effect of criminal law enforcement. J Leg Stud 1:259–276
Ehrlich I (1973) Participation in illegitimate activities: a theoretical and empirical investigation. J Polit Econ 81:521–565
Ehrlich I (1975) On the relation between education and crime. In: Juster FT (ed) Education, income and human behavior. McGraw-Hill, New York, pp 313–337
Ehrlich I, Becker GS (1972) Market insurance, self-insurance, and self-protection. J Polit Econ 80(4): 623–648
Farrell J (1986) Voluntary disclosure: robustness of the unraveling result, and comments on its importance. In: Grieson RE (ed) Antitrust and regulation. Lexington Books, Lexington, pp 91–103
Feinstein JS (1991) An econometric analysis of income tax evasion and its detection. Rand J Econ 22(1): 14–35
Feldman NE, Slemrod J (2007) Estimating tax non-compliance with evidence from unaudited tax returns. Econ J 117(518):327–352
Friedland N, Maital S, Rutenberg A (1978) A simulation study of income tax evasion. J Public Econ 10(1): 107–116
Friehe T (2009) Precaution v. avoidance: a comparison of liability rules. Econ Lett 105:214–216
Friehe T (2010) On avoidance activities after accidents. Rev Law Econ 6:3
Friehe T (2011) A note on the deterrence effects of the forfeiture of illegal gains. Rev Law Econ 7(1):118–124
Garvie D, Keeler A (1994) Incomplete enforcement with endogenous regulatory choice. J Public Econ 55(1): 141–162
Grady MF (1983) A new positive economic theory of negligence. Yale Law J 92:799–829
Grossman GM, Helpman E (2001) Special interest politics. MIT Press, Cambridge, MA
Gunn A (1978) Tax avoidance. Mich Law Rev 76:733–767
Harford JD (1978) Firm behavior under imperfectly enforceable pollution standards and taxes. J Environ Econ Manag 5:26–43
Harford JD (1987) Self-reporting of pollution and the firms behavior under imperfectly enforceable regulations. J Environ Econ Manag 14(3):293–303
Heyes AG (1994) Environmental enforcement when 'inspectability' is endogenous: a model with overshooting properties. Environ Resour Econ 4(5):479–494
Huang C-H (1996) Effectiveness of environmental regulations under imperfect enforcement and the firm's avoidance behavior. Environ Resour Econ 8(2):183–204
Innes R (1999) Remediation and self-reporting in optimal law enforcement. J Public Econ 72:379–393
Innes R (2000) Self-reporting in optimal law enforcement when violators have heterogeneous probabilities of apprehension. J Leg Stud XXIX:287–300
Innes R (2001) Violator avoidance activities and self-reporting in optimal law enforcement. J Law Econ Org 17:239–256
Isenbergh J (1982) Musings on form and substance in taxation. Univ Chic Law Rev 49:859–884
Jones CA (1989) Standard setting with incomplete enforcement revisited. J Policy Anal Manage 8: 72–87
Joulfaian D, Rider M (1996) Tax evasion in the presence of negative income tax rates. Natl Tax J 49(4): 553–570
Kahan M (1989) Causation and incentives to take care under the negligence rule. J Leg Stud 18:427–447
Kaplow L (1990) Optimal taxation with costly enforcement and evasion. J Public Econ 43(2):221–236
Kaplow L, Shavell S (1994) Optimal law enforcement with selfreporting of behavior. J Polit Econ 102(3):583–605
Khambu J (1989) Regulatory standards, compliance and enforcement. J Regul Econ 1(2):103–114
Khambu J (1990) Direct controls and incentives systems of regulation. J Environ Econ Manag 18(2):72–85
Kolm A-S, Larsen B (2004) Does tax evasion affect unemployment and educational choice? SSRN papers
Kronman A (1978) Mistake, disclosure, information, and the law of contracts. J Leg Stud 7(1):1–34
Langlais E (2008) Detection avoidance and deterrence: some paradoxical arithmetic. J Public Econ Theory 10:371–382
Langlais E (2009) On the ambiguous effects of repression. Ann Econ Stat 93/94:349–361
Lee DR (1984) Monitoring and budget maximisation in pollution control. Econ Inq 18:565–575
Linder SH, McBride ME (1984) Enforcement costs and regulatory reform: the agency and firm response. J Environ Econ Manag 11(2):327–346
Malik AS (1990) Avoidance, screening and regulatory enforcement. Rand J Econ 21(3):341–353
Malik AS (1993) Self-reporting and the design of policies for regulating stochastic pollution. J Environ Econ Manag 24:241–257
Mayshar J (1991) Taxation with costly administration, The Scandinavian Journal of Economics 93:75–88
Nagin DS (2013) Deterrence: a review of the evidence by a criminologist for economists. Annu Rev Econ 5(1): 83–105

Neck R, Wachter JU, Schneider F (2012) Tax avoidance versus tax evasion: on some determinants of the shadow economy. Int Tax Public Financ 19(1):104–117
Nowell C, Shogren JF (1994) Challenging the enforcement of environmental regulation. J Regul Econ 6:265–282
Nussim J, Tabbach AD (2005) Avoidance and deterrence, (October 20, 2005). Available at SSRN: http://ssrn.com/abstract=927216
Nussim J, Tabbach A (2008a) Controlling avoidance: ex-ante regulation v. ex-post punishment. Rev Law Econ 4:Article 4
Nussim J, Tabbach A (2008b) (Non)regulable avoidance and the perils of punishment. Eur J Law Econ 25: 191–208
Nussim J, Tabbach A (2009) Deterrence and avoidance. Int Rev Law Econ 29:314–323
Parker SC (2003) Does tax evasion affect occupational choice? Oxf Bull Econ Stat 65:379–394
Pencavel JH (1979) A note on income tax evasion, labor supply, and nonlinear tax schedules. J Public Econ 12(1):115–124
Persson T, Tabellini G (2000) Political economics: explaining economic policy. MIT Press, Cambridge
Pestieau P, Possen UM (1991) Tax evasion and occupational choice. J Public Econ 45(1):107–125
Piketty T, Saez E (2013) Optimal labor income taxation. In: Handbook of public economics, vol 5. (Elsevier) pp 391–474
Polinsky AM, Shavell S (1979) The optimal tradeoff between the probability and magnitude of fines. Am Econ Rev 69:880–891
Polinsky AM, Shavell S (2007) The theory of public enforcement of law. In: Handbook of law and economics, vol 1. (Elsevier), pp 403–454
Reinganum JF, Wilde LL (1985) Income tax compliance in a principal–agent framework. J Public Econ 26(1):1–18
Reinganum JF, Wilde LL (1986) Equilibrium verification and reporting policies in a model of tax compliance. Int Econ Rev 27(3):739–760
Sanchirico CW (2006) Detection avoidance. N Y Univ Law Rev 81:1331–1399
Sanchirico CW (2012) Detection avoidance and enforcement theory. Proced Law Econ 8:145
Sandmo A (1981) Income tax evasion, labor supply and the equity-efficiency tradeoff. J Public Econ 16:265–288
Sandmo A (2005) The theory of tax evasion: a retrospective view. Natl Tax J 58:643–663
Schmidt P, Witte AD (1984) An economic analysis of crime and justice. Academic, Orlando, pp 142–216
Shaffer S (1990) Regulatory compliance with non-linear penalties. J Regul Econ 2(1):99–103
Shavell S (2011) Corrective taxation versus liability as a solution to the problem of harmful externalities. J Law Econ 54(4):S249–S266
Slemrod J (1994) Fixing the leak in Okun's bucket: optimal tax progressivity when avoidance can be controlled. J Public Econ 55(1):41–51
Slemrod J (2001) A general model of the behavioral response to taxation, International Tax and Public Finance 8:119–128
Slemrod J, Yitzhaki S (1987) The optimal size of a tax collection agency. Scand J Econ 89(2):183–192
Slemrod J, Yitzhaki S (2002) Tax avoidance, evasion, and administration. In: Handbook of public economics, vol 3. (Elsevier), pp 1423–1470
Srinivasan TN (1973) Tax evasion: a model. J Public Econ 2:339–346
Stanley TJ (1995a) Radar detectors, fixed and variable costs of crime. Unpublished working paper
Stanley TJ (1995b) Optimal penalties for concealment of crime. Unpublished working paper
Stigler GJ (1970) The optimum enforcement of laws. J Polit Econ 78:526–536
Stigler GJ (1971) The theory of economic regulation. Bell J Econ Manag Sci 2:3–21
Tabbach A (2009) The social desirability of punishment avoidance. J Law Econ Org 26:265–289
Usher D (1986) Tax evasion and the marginal cost of public funds. Econ Inq 24:563–586
Viscusi WK, Zeckhauser R (1979) Optimal standards with incomplete enforcement. Public Policy 27: 437–456
Weisbach DA (2002) An economic analysis of anti-tax-avoidance doctrines. Am Law Econ Rev 4(1): 88–115
Weiss L (1976) The desirability of cheating incentives and randomness in the optimal in-come tax. J Polit Econ 84(6):1343–1352
Yitzhaki S (1974) A note on 'income tax evasion: a theoretical analysis'. J Public Econ 3(2):201–202