P

Panel Data Analysis

Aitor Ciarreta, María Paz Espinosa and Ainhoa Zarraga
University of the Basque Country, UPV/EHU, Bilbao, Spain

Abstract

A panel data set is one that follows a given sample of units over time. Thus, panel data analysis refers to econometric tools that deal with the estimation of relationships that combine time series and cross-sectional data. Appropriate estimation methods are discussed depending on the characteristics of the data.

Denote by y_{it} an observation of the dependent variable for unit i at time t and x_{it} the set of K-independent variables observed for unit i at time t. Consider a linear regression model:

$$y_{it} = \alpha_{it} + \beta'_{it}x_{it} + u_{it}, \quad i = 1,\ldots,N, \quad t = 1,\ldots,T, \tag{1}$$

where the error term u_{it} is independently and identically distributed over i and t, and α_{it} and β'_{it} are (1×1) and $(K \times 1)$ vectors of parameters to be estimated.

This model cannot be estimated because the degree of freedom, NT, is less than the number of parameters to be estimated, $NT(K + 1)$. Therefore, a structure needs to be imposed. Assuming constant parameters over time, the basic initial test is the covariance test for homogeneity across units, where three hypotheses can be made (Greene 2008; Hsiao 2003):

$\mathbf{H}_1$: Both intercept and slope coefficients are the same: $\alpha_{it} = \alpha$ and $\beta'_{it} = \beta'$.

When H_1 is rejected, two different types of heterogeneity across units can be considered:

$\mathbf{H}_2$: Slope coefficients are the same and intercept coefficients are not: $\alpha_{it} = \alpha_i$ and $\beta'_{it} = \beta'$.

$\mathbf{H}_3$: Intercept coefficients are the same and slope coefficients are not: $\alpha_{it} = \alpha$ and $\beta'_{it} = \beta'_i$.

Behind the implied models underlies the assumption that the effects of the omitted variables are of two types: individual time-invariant and period individual-invariant. The simplest representation is to introduce dummy variables for specific cross-sectional units that stay constant over time and time-specific effects constant across units. There are several methods to estimate Eq. 1 depending on the structure of the model. First, assume lagged dependent variable does not enter in the equation as a regressor. Then, the fixed-effects (FE) and the random-effects (RE) models are commonly used.

The FE model removes the effect of time-invariant characteristics from the explanatory variables which are unique to the unit and not correlated with other individual characteristics. Thus,

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

we can assess the net effect on prediction. The equation to estimate is

$$y_{it} = \alpha_i + \beta' x_{it} + u_{it} \quad i = 1,\cdots,N, \quad t = 1, \ldots,T. \tag{2}$$

The least-squares dummy variable estimator, also called the within-group estimator, is obtained applying OLS to Eq. 2 in which variables have been previously transformed to subtract the corresponding time series means. This estimator is unbiased and consistent when either N or T or both tend to infinity.

The RE model, on the contrary, assumes that the variation across units is random and uncorrelated with the independent variables included in the model:

$$y_{it} = \alpha + \beta' x_{it} + v_i + u_{it} i = 1,\ldots,N, \quad t = 1,\ldots,T, \tag{3}$$

where v_i is the between-individual error term. It is assumed that v_i is not correlated with the explanatory variables. If you have any reason to believe that differences across units have some influence on the dependent variable, then random effects are more appropriate. An advantage of RE is that you can include time-invariant variables, whereas in the FE model, these variables are absorbed by the intercept.

In Eq. 3, OLS estimator is inefficient because the error term is correlated. In this case, GLS gives an efficient estimator.

The choice between FE and RE can be performed using the Hausman test where the null hypothesis is that the best model is RE.

There are other tests the analyst has to perform before choosing one model or the other. If the FE model is selected, then test whether time-fixed-specific effects, λ_t, are needed. If they are needed, then the model to estimate is

$$y_{it} = \alpha_i + \lambda_t + \beta' x_{it} + u_{it} i = 1,\ldots,N, \quad t = 1,\ldots,T. \tag{4}$$

The basic model can be extended in several directions. Assume that T is large enough, then the model is dynamic in nature and lagged dependent variables may be included:

$$y_{it} = \gamma y_{i,t-1} + \alpha_i + \lambda_t + \beta' x_{it} + u_{it} \text{ where}$$

$$i = 1,\ldots,N \text{ and } t = 1,\ldots,T. \tag{5}$$

Finally, another extension is to assume variable coefficient models:

$$y_{it} = \sum_{k=1}^{K} \beta'_{kit} x_{kit} + u_{it} \text{ where}$$

$$i = 1,\ldots,N \text{ and } t = 1,\ldots,T. \tag{6}$$

For advanced reading see Baltagi (2008).

References

Baltagi BH (2008) Econometric analysis of panel data. Wiley, New York

Greene WH (2008) Econometric analysis, 6th edn. Prentice Hall, Upper Saddle River

Hsiao C (2003) Analysis of panel data, Econometric society monographs no 34., 2nd edn Cambridge University Press, Cambridge/New York

Parallel Economy

▶ Informal Sector

Parenthood

▶ Adoption

Party Competition: Definitions, Sources, and Economic Effects

▶ Political Competition

Passed Money

► Counterfeit Money

Passive Minority Interests

Nikolaos E. Zevgolis and Panagiotis N. Fotis
Hellenic Competition Commission, Athens, Greece

Abstract

Passive minority interests consist of various types of non-controlling interests in competitors, such as minority shareholdings and interlocking directorships. Loans and other financial products involving competitors may also play a crucial role in creating mutual interactions among the parties. A minority interest is the portion of a consolidated entity that is not owned by the consolidating entity. A consolidated entity may be formed through management/shareholding control and also depends on the prevailing accounting/regulatory environment.

Synonyms

Non-controlling minority interests

Introduction

Passive or non-controlling minority interests (minority shareholdings and interlocking directorships) among competitors of oligopolistic markets are links that tend to raise the price and reduce the quantities sold in the market, even in cases where a cartel, tacit or explicit, is not detected. Such links may force competitors to compete less vigorously and adopt behavior more conducive to joint profit maximization by facilitating price and other strategic information sharing among them. This may be happening because there are cases where a passive minority shareholder may be in a position to receive direct information regarding the victim firm's main operations. In such a case, despite the fact that the practical ability of the owner of the passive minority interest to have access to this kind of information is not enough to enforce him influencing the strategic behavior of the victim firm, it is adequate to provide him with a sort of control over the victim firm (OFT 2010).

Acquisitions of non-controlling minority interests in a rival may also reduce the quality of the product. However, economically speaking, the said links are less likely to trigger unilateral and coordinated effects than full acquisitions or acquisitions of minority shareholdings that influence the strategic behavior of the target firm.

Passive minority interests consist of various types of non-controlling interests in competitors, such as minority shareholdings and interlocking directorships. Loans and other financial products involving competitors may also play a crucial role in creating mutual interactions among the parties. Generally speaking, a minority interest is the portion of a consolidated entity that is not owned by the consolidating entity. A consolidated entity may be formed through management/shareholding control and also depends on the prevailing accounting/regulatory environment.

P

Minority Shareholdings

Minority shareholdings among firms refer to situations where shareholders hold less than 50% of the voting rights of other firms' equity capital. That is, they are portions of a consolidated entity which do not confer control in the legal sense to their owners and are therefore not notifiable under the worldwide merger notification systems. Minority shareholdings are also called structural links (SWD 2014a, Annex 1 and 6).

Active or Passive Minority Shareholdings

Minority shareholdings may be either *active* (controlling) or *passive* (non-controlling) in nature. Active minority shareholdings refer to minority interests where their owner, a person or a firm, may exercise some form of control over the

victim firm(s) (i.e., the issuing firm of the asset or claim for which a minority interest is established). Passive minority shareholdings refer to situations where a minority shareholder cannot be represented in the decision-making bodies of the issuing firm. They constitute purely passive financial interests. However, there are circumstances where a passive minority shareholder may receive direct information about an issuing firm's main operations. Even though the practical ability of the passive minority shareholder to have access to this kind of information does not provide him with the potential to influence the strategic behavior of the issuing firm, it does provide a sort of control over this firm.

Pre-existing Minority Shareholdings

Pre-existing minority shareholdings exists in the context of a notified transaction and therefore may be analyzed regarding its anti- or/and pro-competitive effects within the notified transaction. For instance, a national competition authority or the Commission could order the divestiture of the pre-existing minority shareholding in the case that it is found to be anticompetitive. Where a minority shareholding is acquired after the examination of a notified transaction, the competition authority lacks the competence to review this acquisition.

Horizontal or Non-horizontal Minority Shareholdings

Horizontal minority shareholdings are minority shareholdings in horizontal competitors, while *non-horizontal* minority shareholdings are interests either between firms operating at different levels of the supply chain (*vertical* minority shareholdings) or between firms that are neither purely horizontal nor purely vertical (*conglomerate* minority shareholdings).

Direct or Indirect Minority Shareholdings

When a firm holds a *direct* minority shareholding in another firm it means that it holds it directly without the intervention of a third firm. However, an *indirect* minority shareholding exists when a firm holds a minority shareholding in another firm via the intervention of a third firm.

One-Way or Reciprocal Minority Shareholdings

If both firms hold shares in one another equity capital, then the minority shareholding is called *reciprocal* minority shareholding. The opposite of reciprocal minority shareholding is the *one-way* minority shareholding.

Interlocking Directorships

Interlocking directorships refer to situations where horizontal and nonhorizontal (e.g., vertical) competitors share on boards of directors one or more directors, top executives, nonexecutives, managers, and close relatives, e.g., spouses and parents. Competitive concerns may be stronger in cases where interlocking directorships involve managers rather than directors since the former may more easily understand the day-to-day organization of the issuing firm than the latter and therefore enjoy more decisive influence over its strategic behavior in the interest of their employers.

Types of Interlocking Directorships

As in the case of minority shareholdings, there exist various forms of interlocking directorships such as active or passive, horizontal or non-horizontal, direct or indirect, and one-way or reciprocal interlocking directorships. Except from these basic types, there exist other types of interlocking directorships between firms. For example, in some circumstances firms' officers may hold a seat on the board of directors of competitors. This type of directorship is called *management interlock* (Arreda and Turner 1978). However, there are competitors that share common officers without involving directorships. This type of interlock seems not to raise competition concerns and it is called a *manager to manager interlock* (O'Brien and Salop 2000).

Passive Minority Interests and Competition Law

In a Nutshell

As a general rule, active or controlling minority shareholdings fall under the scope of merger

notification systems, while passive or non-controlling minority shareholdings are not subject to the majority of control systems found worldwide. Against the backdrop of the Council Regulation (EC) No. 139/2004 (*The EC Merger Regulation*), the Commission cannot examine minority shareholdings in stand-alone investigations. Nevertheless, in some national merger laws, there are provisions regarding the acquisitions of minority shareholdings in competitors (see, for instance, Germany, United Kingdom, Austria, and Lithuania), but the majority of them do not incorporate such provisions in their merger notification systems. In that case, the regulation of passive minority shareholdings falls under the mechanism of antitrust law. Preexisting minority shareholdings constitute exceptions to this rule.

Passive Minority Interests and Mergers and Acquisitions

Active along with pre-existing minority shareholdings are the only forms of minority interests that fall under the scope of the Merger Regulation 139/2004. More specifically, active minority shareholdings fall directly under the scope of this Regulation; preexisting minority shareholdings fall indirectly under the scope of the same Regulation. A few merger cases in EU (see, *inter alia*, *IV/M.042 Alcatel/Telettra, IV/M.113 Courtaulds/SNIA, IV/M.833 Coca-Cola/Carlsberg, IV/M.890 Blokker/Toys "R" Us On 23, IV/M.1082 Allianz/AGF, IV/M.1378 Hoechst/Rh ô ne-Poulenc, IV/M.1383 Exxon/Mobil, IV/M.1453 AXA/GRE, IV/M.1673 VEBA/VIAG, IV/M.1940 Siemens/Framatome/Cogéma, COMP/M.1980 Volvo/Renault, COMP/M.2050 Vivendi/Seagram, COMP/M.2567 Nordbanken/Postgirot, COMP/M.3547 Banco Santander/Abbey National Transactions, COMP/M.3653 Siemens/VA Tech, IV/M.3696 E.ON/MOL, COMP/M.4150 Abbott/Guidant, COMP/M.4153 Toshiba/Westinghouse, COMP/M.4439 Ryanair/Aer Lingus, COMP/M.5096 RCA/MAV Cargo, COMP/M.5406 IPIC/MAN Ferrostaal, COMP/M.6541 Glencore/Xstrata, COMP/M.6662 Andritz/Schuler*) constitute examples where a party to a transaction has pre-existing structural links to competitors or other firms and where the parties quite often offered to divest these minority shareholdings with the intention of remedying the competition concerns raised by the transactions.

An active minority shareholding exists if the major prerequisites of the EU Merger Regulation are satisfied. In particular, the merger notification system in the EU is based on three pillars: (i) the notion of concentration, (ii) the existence of acquisition of control, and (iii) the dimension of the notified transaction. According to the Commission Consolidated Jurisdictional Notice under Council Regulation . . . (2008), para 7, "a concentration only covers operations where a change of control in the undertakings concerned occurs on a lasting basis." So, "the concept of concentration is intended to relate to operations which bring about a lasting change in the structure of the market." The notion of concentrations also applies to joint ventures, which perform "on a lasting basis all the functions of an autonomous economic entity."

In terms of European law the concept of concentration cannot be extended to cases in which control has not been obtained and the shareholding at issue does not, as such, confer the power of exercising decisive influence on the other undertaking. However, the European Commission seems to support the need to extend Regulation 139/2004 to the acquisition of non-controlling minority shareholdings (see SWD 2014a, 221, para 62).

Passive Minority Interests and Anticompetitive Practices

In General

Articles 101 and/or 102 of the Treaty on the Function of the European Union (TFEU) may apply to passive minority interests in situations where there is evidence of an anticompetitive agreement or concerted practice among the investigated firms or the firms that are engaged in the acquisition of non-controlling stakes and/or one or more firms have a dominant position. Nevertheless, according to the Commission Staff Working Document (SWD 2014a, 221), "the Commission's ability to use Article 101 and Article 102 TFEU to intervene against anti-competitive minority shareholdings may be limited."

More specifically, in its *White Paper* and the accompanying Commission Staff Working Paper of July 2014, the Commission states that an acquisition of a minority shareholding may not constitute an agreement which by object or effect restricts the effective competition. As a consequence, it may not be possible to intervene in cases that involve potentially anticompetitive minority shareholdings. Furthermore, regarding Article 102 TFEU, the Commission states the same argument unless the acquirer or the issuing firm or both of them have a dominant position in the markets under scrutiny (see White Paper (2014b) para 40; SWD (2013) 239, Annex II, p 6, and SWD (2014a) 221, para 62).

Application of Article 101 TFEU

In some cases, an agreement which by object or effect restricts the effective competition and violates Article 101 TFEU does not exist. For instance, acquisitions of minority interests via a stock exchange may not contain an agreement among the firms of the sellers and the buyers. In this case, a memorandum among shareholders and/or particular firms' articles covering their relationship may be viewed as an agreement under Article 101 TFEU. Generally speaking, stock market acquisitions of minority shareholdings involve an agreement between the firm, who acts as a purchaser, and the shareholder (s) of a firm, who act as the seller(s) or the issuing firm in the transaction. In this transaction, the difficulty which arises for assessing the potential anticompetitive effects of such an acquisition is that the firm is not (at least typically) a participant in the transaction. Therefore, an agreement between the two firms does not exist.

On the other hand, an agreement between the two firms does exist if the corporate bodies of the two firms are involved directly in the transaction. Furthermore, Gilo et al. (2006) have proved that a controlling shareholder (whether a person or a parent corporation) can facilitate tacit collusion further by making a direct passive investment in rival firms. Such investment particularly facilitates collusion if the controller has a relatively small stake in his own firm.

Potential Application of Article 101 TFEU in Combination with Application of Article 102 TFEU

The establishment of coordinated effects (see section "Coordinated Effects of Passive Minority Interests") as a consequence of an acquisition of passive minority interests implies that the firms engaged in the transaction coordinate their strategic behavior, i.e., they raise prices and harm effective competition. Coordination may occur by firms that prior to the acquisition did or did not coordinate their behavior. In the former case, a minority shareholding may make coordination more effective and easier.

An acquisition of a minority shareholding may serve as an instrument for influencing the commercial or strategic policy of a direct competitor (especially in the case where reciprocal minority shareholdings exist), or it may enable the acquiring firm to acquire control over the acquired firm at a later stage. In that case, even though the acquisition of minority shareholding does not lead by itself to a restriction of competition for the purposes of Article 101 TFEU, i.e., it has neither the object nor the effect of distorting competition by creating the premises for cooperation and/or exchange of commercially sensitive information among the engaged firms, its potential impact on competition in the long run can be seen as an agreement which violates Article 101 TFEU.

As a consequence, such structural links may eliminate the incentives of the competitors to compete with each other in the market(s) in which they carry out business (Levy 2014). As well as potential commercial cooperation among firms engaged in a minority shareholding transaction, there are other factors that may serve in this direction. For instance, providing that the market under scrutiny is an oligopolistic market, such factors are barriers to entry and the ability of the acquiring firm to achieve effective control (legal or de facto) over the commercial and strategic policy of the acquired firm.

In particular, the level of concentration in the market under scrutiny plays a critical role in the existence or not of coordinated behavior. The rule is that the higher the level of concentration, the lower the number of the effective firms in the

market and the higher the likelihood of cooperation among them. This depends crucially on the magnitude of the influence of the strategic behavior of the acquired firm, since the more the acquiring firm knows about the target firm, the easier it is to coordinate with it in order to enhance the profits of both firms. A concentrated market enhances coordination among its players without the need for an agreement that may violate Article 101 TFEU; the possibility of collective dominance – and perhaps abuse of it under Article 102 TFEU – in such a case is quite strong. However, two factors that independently contribute to such an enhancement must be considered: firstly, the existence of reciprocal and indirect minority shareholdings among the firms in the market; and secondly, the existence or not of a maverick firm. (The existence of a maverick undertaking would probably diminish the possibility of collective dominance.) More specifically, the existence of reciprocal and indirect minority shareholdings among the firms in the market enhances the ability of the acquiring firm to monitor the strategic behavior of the target firm and at the same time the strategic behavior of all the other firms that are connected with it via the existence of minority shareholdings.

Application of Article 102 TFEU

Article 102 TFEU applies only in the case where the engaged firms in the transaction under scrutiny hold a dominant position, either independently or collectively, in the concerned relevant market(s). Therefore, an acquisition of minority shareholding requires the existence of a dominant position between the parties that engaged in it. However, in theory it has been stated that an abuse of a dominant position requires that an acquisition of a minority shareholding may serve as an instrument for influencing the commercial or strategic policy of a competitor, especially in the case where there exist reciprocal minority shareholdings in the relevant market under scrutiny (Fotis and Zevgolis 2016).

As a matter of fact, this was the judgment of the Court of Justice in the *Phillip Morris* case. Furthermore, in the *Gillette* case the firm abused its dominant position in the relevant market of disposable razors by acquiring a minority shareholding over its direct competitor, Wilkinson Sword. Gillette had become the main shareholder and creditor of Eemland (the owner of Wilkinson Sword) and could have used its rights to prevent future concentration plans that Eemland would have had, and Gillette would not have approved of. Therefore, in the *Gillette* case the European Commission established an infringement of Article 102TFEU since the transaction had altered the structure of the market under investigation.

The Economics of Passive Minority Interests

The majority of relevant research in the literature indicates that non-controlling minority interests decrease the level of competition in the markets by enhancing cooperation among rivals or by increasing the probability that a dominant firm will abuse its dominant position. In any case, passive minority interests decrease consumer welfare by increasing the product price and reducing its quantity.

The Anticompetitive Effects of Passive Minority Interests

Economic literature has shown that passive minority interests among rivals (*horizontal passive minority interests*) harm competition either unilaterally or via coordinated effects. The magnitude of these effects depends on the share of the issuing firm's profits to which the acquiring firm is actually entitled as a result of the non-controlling acquisition, and the ability of the acquiring firm materially to influence the issuing firm's strategic behavior. If passive minority interests are accompanied by interlocking directorships, then such interests may alter the competitive behavior of both acquiring and target firm.

The anticompetitive effects of *nonhorizontal passive minority interests* have also been examined in the economic literature. *Vertical passive minority interests* impose input (upward) or customer (downstream) exclusion. *Conglomerate passive minority interests* enhance anticompetitive

issues if the acquiring firm acquires a non-controlling minority stake in the issuing firm which is active in a market related to that of the acquiring firm. A conglomerate non-controlling minority interest may promote acquiring firms' sales of one product contingent on the sales of other products via *bundling*, *tying*, *exclusive dealing*, or *full-line forcing*.

Non-controlling minority interests may block entry both for potential entrants that are not active in the market or for active firms in the market that are interested in expanding their product range. Also, acquiring firm hinders another firm from getting an access to the capital stock of the issuing firm. For instance, in the Aer Lingus/Ryanair case Aer Lingus argued that Ryanair's non-controlling minority interest on its capital stock had a material impact on Aer Lingus's shares, by making them less attractive to potential buyers (*Case T-411/07, "The Ryanair Decision"* and *Case T-342*07 Ryanair v Commission*). However, it should not be ignored that the attractiveness of Aer Lingus on the stock market is not based solely on Ryanair's non-controlling minority interest (SWD 2014a, 221).

Furthermore, the acquiring firm has an enhanced incentive to deter entry by credibly committing to the potential entrant. If the non-controlling minority interest is accompanied by corporate rights, then the commitment is stronger that further prevents a potential entrant from entering to the market. However, if non-controlling reciprocal minority interests are prohibited (allowed), then entry will be deterred (promoted) if the relevant products are complements (substitutes) (Clayton and Jorgensen 2005).

Unilateral Effects of Passive Minority Interests

The anticompetitive effects of non-controlling minority interests are stronger on the horizontal than on vertical level. On the former, they change firms' strategic behavior in favor of coordination among them. If the acquiring firm has an incentive to compete softly with its competitor, may raise its product price or restrict its output (see, *inter alia*, Reynolds and Snapp 1986; Bresnahan and Salop 1986; Farrell and Shapiro 1990; Shelegia and Spiegel 2012).

In a static environment, the acquiring firm abuse part of the issuing firm's diverted sales via its non-controlling minority stake. In this scenario, the degree of competition between the two firms and the amount of sales that diverted to the competitors are of high importance. The more the rivalry between firms, that is the case of close substitute products, the more the diverted sales that are captured by the issuing firm after an increase of acquiring firm's product. This is called *Diversion Ratio*. In duopolistic markets, the *Diversion Ratio*, in absolute terms, is the ratio of the cross-elasticity of demand for firm's A product when firm B raises its product price divided by the own-elasticity of demand for firm's B product multiplied by the ratio of unit sales by firm B divided by the unit sales by firm A (Fotis et al. 2017). Therefore, the acquiring firm earns profit, the magnitude of which depends on the non-controlling minority interest in the capital of the issuing firm.

Vertical passive minority interests harm competition by imposing input (upward) or customer (downstream) exclusion. If they are accompanied by corporate rights, then the acquiring firm gains the ability to exclude the issuing firm's competitors from its input and customer base. The more concentrated the markets are, the more the concerns for either type of exclusion. The acquiring firm's incentive to exclude competitors is higher through corporate rights than through solely financial interests. If the acquiring firm gains access to strategic data of the issuing firm, then the gains from exclusion are more than the gains accrued from a vertical merger. The acquiring firm internalizes the whole stream of profits that come from the other levels of supply chain and at the same time incurs only a small portion of the costs caused by the anticompetitive strategy (see, *inter alia*, Flath 1989; Greenlee and Raskovich 2006).

An example of vertical passive minority interest at the European Union (EU) level is the case *COMP/M.5406 IPIC/MAN Ferrostaal*. In this case, MAN Ferrostaal (the acquiring firm) held a non-controlling minority stake in Eurotecnica (the issuing firm), a provider of technology and engineering services. The EU Commission was

concerned that the acquiring firm would determine the technology licenses distribution by the issuing firm and thus excluded (potential or active) competitors from the market. The suggested remedy by the EU Commission in order to clear the case included the divestiture of the non-controlling minority interest held by MAN Ferrostaal in Eurotecnica (SWD 2014a, 221).

Coordinated Effects of Passive Minority Interests

Non-controlling minority interests also enable involved firms to coordinate their conduct. Through coordination the acquiring firm internalizes part of the issuing firm's profit. Moreover, the level of transparency is enhanced in the presence of corporate rights. On the one hand, in the case of one-way acquisition of information from the issuing to the acquiring firm (*unilateral passive minority interest*), the acquiring firm has the ability to monitor the commercial policy of the issuing firm. If, on the other hand, a reciprocal minority shareholding exists, then the level of transparency is enhanced more than the previous case (see, *inter alia*, Malueg 1992; Reitman 1994; Gilo et al. 2006; Brito et al. 2013).

A market becomes more transparent when *indirect non-controlling minority interests* are accompanied by corporate rights. Such interests enlarge the scope of coordination and enable the coordinating firms to detect possible deviations from the common target which aims at maximizing the monopoly profits (*IV/M.1673 V EBA/VIAG*).

Pro-competitive Passive Minority Interests and Efficiencies Gains

Non-controlling minority interests raise arguments in favor of efficiencies. Clayton and Jorgensen (2005) analyze reciprocal minority interests in a duopoly Cournot market, and they state that consumer surplus and firms' profits with complements products are enhanced more when the said interests are allowed rather than when they are prohibited. If the products are substitutes, then firms' profits are enhanced when passively acquired interests are prohibited rather than when they are allowed. On the contrary, consumer surplus is enhanced when non-controlling minority interests are allowed.

Brito et al. (2014) argue that consumer surplus is enhanced by turning voting shares (shares with control rights) into nonvoting shares (shares with no control rights). Moreover, the sale of voting shares to a new large shareholder is better than the sale of voting shares to a series of small shareholders. The authors agree with the main results of O'Brien and Salop (2000), but also consider that a financial interest affects not only the incentives of the acquiring but also the incentives of the acquired firm (see also *COMP/M.2416 Tetra Laval/Sidel*).

Passive minority interests enable the transfer of technology and managerial skills and the creation of synergies among firms which positively influence firms' profits (Ono et al. 2004). Amundsen and Bergman (2002) argue that cost reduction through sales cooperation and learning, such as information about production process, are also, *inter alia*, motives which enable a firm to acquire non-controlling minority interests in their rivals' capital stock.

Cross-References

▶ Cartels and Collusion

References

Amundsen ES, Bergman L (2002) Will cross-ownership re-establish market power in the Nordic power market ? Energy J 23(2):73–95

Arreda P, Turner DF (1978) Antitrust law. Little Brown and Company, Boston

Bresnahan TF, Salop SC (1986) Quantifying the competitive effects of production joint ventures. Int J Ind Organ 4:155–175

Brito D, Ribeiro R, Vasconcelos H (2013) Quantifying the coordinated effects of partial horizontal acquisitions. CERP discussion papers no. 9536

Brito D, Ribeiro R, Vasconcelos H (2014) Measuring unilateral effects in partial horizontal acquisitions. Int J Ind Organ 33:22–36

Clayton MJ, Jorgensen BN (2005) Optimal cross holding with externalities and strategic interactions. *J Bus* 78(4):1505–1522

Commission Consolidated Jurisdictional Notice under Council Regulation (EC) No. 139/2004 on the control

of concentrations between undertakings (2008) OJ C95/01. Accessed 10 June 2010
Council Regulation (EC) No. 139/2004 of 20 January 2004 on the control of concentrations between undertakings (the EC Merger Regulation) OJ L 24, 29.01.2004, pp 1–22. Accessed 10 June 2010
Farrell J, Shapiro C (1990) Asset ownership and market structure in oligopoly. RAND J Econ 21(2):275–292
Flath D (1989) Vertical integration by means of shareholding interlocks. Int J Ind Organ 7:369–380
Fotis P, Zevgolis N (2016) The competitive effects of minority shareholdings legal and economic issues. Hart Publishing, Bloomsbury
Fotis P, Polemis M, Eleftheriou K (2017) Unilateral effects of partial acquisitions: consistent calculation of GUPPI under horizontal merger guidelines within the EU. Econ E Polit Ind. https://doi.org/10.1007/s40812-016-0053-6
Gilo D, Moshe Y, Spiegel Y (2006) Partial cross ownership and tacit collusion. RAND J Econ 37(1):81–99
Greenlee P, Raskovich A (2006) Partial vertical ownership. Eur Econ Rev 50:1017–1041
Levy N (2014) Expanding EU merger control to non-controlling minority shareholdings: a sledgehammer to crack a nut? CPI Antitrust Chron 12:1–16
Malueg AD (1992) Collusive behavior and partial ownership of rivals. Int J Ind Organ 10:27–34
O'Brien DP, Salop S (2000) Competitive effects of partial ownership: financial interest and corporate control. Antitrust Law J 67(3):559–614
OFT (2010) Minority interests in competitors. A research report prepared by DotEcon Ltd, OFT1218
Ono H, Nakazato T, Davis C, Alley W (2004) Partial ownership arrangements in the Japanese automobile industry; 1990–2000. J Appl Econ 7(2):355–367
Reitman D (1994) Partial ownership arrangements and the potential for collusion. J Ind Econ 42:313–322
Reynolds RJ, Snapp BR (1986) The competitive effects of partial equity interests and joint ventures. Int J Ind Organ 4(2):141–153
Shelegia S, Spiegel Y (2012) Bertrand competition when firms hold passive ownership stakes in one another. Econ Lett 114:136–138
SWD (2013) 239 Final, commission staff working document (Towards more effective EU merger control), Annex II. Accessed 5 June 2010
SWD (2014a) 221 Final, commission staff working document (Towards more effective EU merger control). Accessed 5 June 2010
SWD (2014b) White Paper (Towards more effective EU merger control) 449 final. Accessed 3 June 2010

Further Reading

Fotis P, Zevgolis N (2016) The competitive effects of minority shareholdings legal and economic issues. Hart Publishing, Bloomsbury

List of Cases

COMP/M.1980 Volvo/Renault (2000) OJ C 301
COMP/M.2050 Vivendi/Seagram (2000) C 261/04
COMP/M.2416 Tetra Laval/Sidel (2004) OJ L38, 1
COMP/M.2567 Nordbanken/Postgirot (2001) OJ C 347
COMP/M.3547 Banco Santander/Abbey National (2004) OJ C 255
COMP/M.3653 Siemens/VA Tech 2006/899/EC, OJ L 353
COMP/M.4150 Abbott/Guidant (2006) OJ C 256/10
COMP/M.4153 Toshiba/Westinghouse (2006) OJ C 184/02
COMP/M.4439 Ryanair/Aer Lingus (2007) OJ C 47/08 (decision of 20 December 2006, OJ C 10, 2007 (the Ryanair decision)
COMP/M.5096 RCA/MAV Cargo (2008) OJ C 29
COMP/M.5406 IPIC/MAN Ferrostaal (2009) OJ C 114
COMP/M.6541 Glencore/Xstrata (2014) OJ C 109/1
COMP/M.6662 Andritz/Schuler (2013) OJ C 14
IV/M.1673 V EBA/VIAG (2000) OJ L 188
IV/M.042 Alcatel/Telettra (1991) OJ L 122/48
IV/M.113 Courtaulds/SNIA (1991) OJ C 333
IV/M.833 Coca-Cola/Carlsberg (1997) OJ L 145
IV/M.890 Blokker/Toys 'R' Us (1998) OJ L 316/1
IV/M.1082 Allianz/AGF (1998) OJ C 246/4
IV/M.1378 Hoechst/Rh ô ne-Poulenc (1999) OJ C254/5 90
IV/M.1383 Exxon/Mobil (1999) OJ 2004 L103/1
IV/M.1453 AXA/GRE (1999) OJ C 030/6
IV/M.1673 VEBA/VIAG (2000) OJ L 188
IV/M.1940 Siemens/Framatome/Cogéma (2000) OJ L 289/8
IV/M.3696 E.ON/MOL 2006/622/EC, OJ L 253
Warner-Lambert/Gillette (1993) IV/33.440
T-342/07 Ryanair v Commission (2010) ECR II-3457
T-411/07 Aer Lingus v Commission (2010) ECR II-3691
The Philip Morris case, 142 and 156/84 (1987) ECR 4487 (1988) 4 CMLR 24

Internet Addresses

Commission Consolidated Jurisdictional Notice. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2008:095:0001:0048:EN:PDF
TFEU. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012E%2FTXT
The EC Merger Regulation. http://ec.europa.eu/competition/mergers/legislation/regulations.html
SWD (2013) 239 Final. http://ec.europa.eu/competition/consultations/2013_merger_control/merger_control_en.pdf
SWD (2014) 221 Final. http://ec.europa.eu/competition/consultations/2014_merger_control/staff_working_document_en.pdf
White Paper. http://ec.europa.eu/competition/consultations/2014_merger_control/mergers_white_paper_en.pdf

Patent Annulment

► Patent Litigation

Patent Infringement Lawsuit

► Patent Litigation

Patent Invalidation Challenge

► Patent Opposition

Patent Litigation

Elisabetta Ottoz
Department of Economics and Statistics Cognetti De Martiis, University of Turin, Turin, Italy

Abstract

Patent litigation refers to patent infringement lawsuits or revocation proceedings. Infringement is the act of making, using, selling, or offering to sell a patented invention without the permission of the patent owner. Revocation proceedings refer to the claim on patent validity before civil courts that may be carried out by firms interested not to be sued for infringing "wrongly" granted patents. Presently, national courts of the member states of the European Patent Convention are competent to pass judgment on the infringement and validity of European patents, with inevitable consequences in terms of duplication and inconsistencies. In December 2012, the European Parliament approved the EU unitary patent package, whose ratification by the individual member states will give rise to a European patent with unitary effects in all jurisdictions involved and to the creation of a Unified Patent Court (UPC) with exclusive jurisdiction to hear infringement and invalidity actions. The result will be a patent protection for all participating member states based on a single application and validation.

Synonyms

Patent annulment; Patent infringement lawsuit

Definition

Patent litigation refers to patent infringement lawsuits or revocation proceedings. Infringement is the act of making, using, and selling a patented invention without the permission of the patent owner. Revocation proceedings refer to the claim on patent validity before civil courts.

Introduction

A patent is the exclusive right granted by a government to an inventor to manufacture, use, or sell an invention for a certain number of years in exchange for detailed public disclosure of the invention itself so as to encourage research and development activities fostering knowledge dissemination.

In order to be patentable, an invention must satisfy some requirements concerning novelty, usefulness, and nonobviousness (WIPO 2008).

A patent is not a perfect protection against imitation, but it grants the patent holder the right to sue intruders once they have been identified. Conversely third parties have the right to challenge the validity of patents granted by the patent authority.

Patent litigation can thus take two distinct forms: infringement or revocation proceedings.

Patent infringement is the act of making, using, selling, or offering to sell a patented invention without the permission of the patent owner. The economic significance of a patent depends on its scope: the broader the scope, the larger the number of competing products and processes that will infringe the patent. The claims contained in the application are the basis of the extent of patent protection as they determine what third parties are legally allowed to do.

Revocation proceedings refer to the claim on patent validity before civil courts that may be carried out by firms interested not to be sued for infringing "wrongly" granted patents, either autonomously or as a counterclaim in a cause for patent infringement.

Proceedings before a national tribunal for revocation as part of patent litigation are not to be

confused with patent opposition, which is an administrative procedure contained in the article 99 of the European Patent Convention allowing third parties to question the validity of a patent granted by the European Patent Office (EPO) within 9 months from the its publication. Patent opposition applies to the European patent at the European-wide level, whereas revocation proceedings within litigation apply to national jurisdictions.

Infringement and validity of European patents are currently under the jurisdiction of national courts and authorities of the member states of the European Patent Organisation.

Despite the Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights, relevant differences still exist as patent litigation and court judgments on validity and infringement vary significantly from one country to another (EPO 2013a). In practice, this gives rise to a number of shortcomings: high costs, risk of diverging decisions, and lack of legal certainty (Luginbuehl 2011).

The European patent system is currently undergoing major reforms aimed at overcoming existing fragmentation, which becomes increasingly problematic as innovation and industrial R&D has assumed a global scope. In December 2012, the European Parliament approved the EU unitary patent package, whose ratification by the individual member states will give rise to a European patent with unitary effects in all jurisdictions involved and to the creation of a Unified Patent Court (UPC) with exclusive jurisdiction to hear infringement and invalidity actions.

Institutional Features

European inventors can obtain patent protection by filing several national applications or, alternatively, one patent application to the EPO in which several States adhering to the 1973 European Patent Convention (EPC) are designed. The granting of a European patent allows the applicant to achieve a bunch of patents treated as independent rights, each having a limited scope: whether or not the national parts of the European patent are infringed or invalid is then determined based on the national laws of the respective member states of the European Union.

Relevant institutional differences among the jurisdictions still exist concerning several aspects such as the existence of bifurcation, remedies for patent infringement, forum shopping, and the allocation of legal costs as illustrated by Graham and Van Zeebroeck (2014).

Bifurcation

In terms of institutional settings, a major difference relates to whether litigants are permitted or required to address infringement and invalidity claims within the same court and suit or, as in a bifurcated system, infringement is heard and determined separately from validity.

The German system is a bifurcated one: invalidity challenges, either standalone invalidation challenges or appeals of decisions rendered by the German Patent Office, can only be brought to the Federal Patent Court, whereas infringement actions can be lodged in any of the twelve competent district courts. In the French, British, Dutch, Italian, and Belgian systems, patent infringement and invalidity actions, either for national patents or for national validations of patents granted by the EPO, are brought to the same court.

The bifurcated patent system is positively considered because of quick decisions and low costs, but it might be potentially biased toward the patentee in two ways. The fast infringement proceedings in the regional courts mean, in fact, that it is possible to get to an injunction before the patent can be invalidated by the slower invalidity courts. In addition, a bifurcated system is subject to the so-called Angora cat problem: the patentee will argue for a narrow interpretation of his claim when defending the patent but an expansive interpretation when asserting infringement.

Injunctions

If a patent holder discovers that his/her patent is being infringed by products or services belonging to other parties, he/she faces the decision to file a

lawsuit and, following that, either to engage the alleged infringer in a pretrial settlement negotiation or to file suit in court claiming damages for the infringement.

The TRIPS Agreement generally provides for injunctions and damages as a remedy to patent infringement, but the specific procedures and standards for awarding these remedies are left to the member states, which apply different conditions and thresholds.

Courts in several member states allow in principle preliminary injunction, which is granted very early in a court action and restrains the defendant from infringing the patent during the pendency of litigation (Cotter 2011).

If a preliminary injunction is issued, the plaintiff will nearly always have to post a bond for securing any costs or damages caused to the defendant in case infringement is not proved. The amount of the bond is left to the court's discretion.

If the plaintiff wins at the trial, the preliminary injunction usually becomes permanent, but if the defendant wins, the preliminary injunction is removed.

Cross-border preliminary injunctions forbidding accused infringers from practicing the litigated patents both in the domestically and abroad have been applied in the Dutch litigation system with the result of a "forum shopping strategy," as patent holders may stop infringement of their patent throughout Europe.

Belgium too has become a preferred venue for patent owners who wish to quickly enforce their patent rights: Belgian courts assume, in fact, in their preliminary injunctions the prima facie validity of the patents, even when an opposition was pending or appealed at the EPO.

Damages

Most jurisdictions provide for three "standard methods" for damages assessment based on the calculation of lost profits, reasonable royalty, and infringers' profits (unjust enrichment) (Reitzig et al. 2007).

In the case of lost profits, the patentee shall be reinstated in a position where he/she would have been but for the infringement. The calculation method is accepted by all major jurisdictions (USA, Japan, Germany, UK, and France).

In the US jurisdiction, the patentee, in order to be awarded lost profits, has to show causation, establishing that "but for" the infringement, she would have made additional profits. In 1978, the Court of Appeals for the Sixth Circuit put forth a test designed to determine whether a patent holder is entitled to recover for lost profits by the infringement (Panduit Corp. v. Stalin Bros. Fibre Works, Inc., 575 F.2d 1152, 1156 (6th Cir. 1978)).

It is a four-step test requiring that the patentee establishes (1) the demand for the patented product, (2) the absence of noninfringing substitutes, (3) the manufacturing and marketing capability to exploit the demand, and (4) the amount of profit that would have been made absent the infringing product.

Assuming that the four Panduit conditions are met, lost profit evaluation requires to infer how the market would have evolved absent infringement and to compare the hypothetical behavior of both the patentee and the infringer with their actual behavior. The difference between the "but for" and the actual profit represents the patent holder's lost profit damage.

Proof of damages is simpler when the patentee and the infringer compete in a two-supplier market, notwithstanding a market share analysis can be used even in case of several firms in the market in order to determine a measure of lost profit award.

Once lost sales are determined, total lost profits can be calculated by measuring the incremental profits on lost sales plus profits lost on price erosion.

To the extent the patentee does not satisfy "but for" causations required by the Panduit steps, he is entitled to a reasonable royalty. The courts attempt to reconstruct the hypothetical bargain that the parties would have negotiated had they willingly tried to do so at the time infringement began. A hypothetical negotiation between a willing licensor and a willing licensee is imagined generally relying on fifteen factors set forth in Georgia-Pacific Corp. v. United States Plywood Corp. (1970).

The case law has established a few guiding principles that should be present in any reasonable royalty determination: first of all the best measure of reasonable royalty is an established royalty rate in the industry; second the hypothetical negotiation takes place at the time the infringement began, meaning that infringers' sunk cost are not part of the infringer's anticipated profits; and third, the patentee need not prove any actual harm to be entitled to a reasonable royalty (Frank and DeFranco 2000).

A damage award can be composed of lost profits and a reasonable royalty as, for instance, in the case where lost profits include lost sales for which the patentee had manufacturing capacity, while a reasonable royalty accounts for additional sales that exceeded the patentee's manufacturing capacity. In case the patent holder is a university or a research institute, the damage is entirely based on a reasonable royalty.

The third way of calculating damages relies on infringer's profits. In Germany the damage "is based on the legal fiction that in using another's patent, the infringer undertook a business on behalf of the rights-owner, who would thus be entitled to obtain all profits made from such business" (Reitzig et al. 2007). Granting "infringers' profits" is formally not allowed in France and the USA, even if the US term "unjust enrichment" may be interpreted as a rather close notion.

In case of willful infringement, in the USA, whether the damage award is in the form of lost profits or reasonable royalties, courts have discretionary authority to enhance the damage award by three times.

Courts have sometimes awarded inflated reasonable royalties that do not reflect the market ones, but rather imply a deterrent function against future infringements (Love 2009).

Choice of Fora

When a patent infringement suit involves a defendant domiciled in an EU member state, national courts of member states must exercise jurisdiction in accordance with Articles 2 and 5(3) of the Brussels Regulation. Under this regime, a plaintiff may bring an action in the courts of the defendant's domicile or in the state(s) in which the alleged infringing product was manufactured or commercialized in breach of a local patent. Dutch courts have even exercised jurisdiction over foreign defendants for violations of foreign patents.

This means that patent litigants in Europe are permitted a choice of fora, which gives rise to an opportunity to engage in "forum shopping," a strategic choice of court venues in order to obtain a favorable outcome leading to economic inefficiencies. Parties try to take advantage of differences in national courts' interpretation of European patent law and in procedural laws, as well as of differences in speed and in the amount of damages awarded.

"Forum shopping is a common practice in the USA too, as any civil action for patent infringement may be brought in the judicial district where the defendant resides or where the defendant has committed acts of infringement and has a regular and established place of business" 28 USC. § 1400 (Lemley 2010).

The Allocation of Legal Costs

Two main systems for allocating litigation costs are applied, namely, the "American system," where each party bears its own costs, and the "British system," where the loser incurs all costs. In an intermediate position lie the systems which allow a partial fee shifting.

The legal-cost allocation rule plays a role in favoring patent litigation or settlement and bears implications on the royalty-bargaining process.

The Unified Patent Court

In order to overcome duplication and inconsistencies, in December 2012 the European Parliament approved the EU unitary patent package, whose ratification by the individual member states will give rise to a European patent with unitary effects in all jurisdictions involved, that is to say subject to the same legal conditions in all member states. The result will be a patent protection for all participating member states, except Italy and Spain, based on a single application and validation putting an end to validations and litigation of the patent in each state.

A Unified Patent Court (UPC), with exclusive jurisdiction to hear infringement actions, invalidity actions and counterclaims, and actions for provisional and protective measures and injunctions for litigation relating to European patents and European patents with unitary effect (unitary patents), will be created considering a transitional period of 7 years during which actions of litigation may be brought before national courts (Agreement on a Unified Patent Court and Statute 2013b)

The UPC will comprise a court of first instance, a Court of Appeal, and a registry. The court of first instance will be composed of a central division (with seat in Paris and two sections in London and Munich) and by several local and regional divisions in the contracting member states to the Agreement. The Court of Appeal will be located in Luxembourg.

Generally, claimants will bring action for revocation before the central division and will bring actions for infringement before a local/regional division in a member state in which the infringement has occurred or where the defendant is domiciled.

The system allows for a choice between bifurcation – the separation of infringement and validity claims into separate court actions as in the German system –and an integrated process for hearing infringement and invalidity cases: the regional courts have, in fact, the discretion to refer counterclaims for revocation raised by the defendant to the central division.

Where a decision is taken finding an infringement of a patent, the Court may grant an injunction against the infringer aimed at prohibiting the continuation of the infringement.

As for the award of damages Art. 68 of the UPC states:

1. The Court shall, at the request of the injured party, order the infringer who knowingly, or with reasonable grounds to know, engaged in a patent infringing activity, to pay the injured party damages appropriate to the harm actually suffered by that party as a result of the infringement.
2. The injured party shall, to the extent possible, be placed in the position it would have been in if no infringement had taken place. The infringer shall not benefit from the infringement. However, damages shall not be punitive.
3. When the Court sets the damages:
   (a) It shall take into account all appropriate aspects, such as the negative economic consequences, including lost profits, which the injured party has suffered, any unfair profits made by the infringer and, in appropriate cases, elements other than economic factors, such as the moral prejudice caused to the injured party by the infringement; or
   (b) As an alternative to point (a), it may, in appropriate cases, set the damages as a lump sum on the basis of elements such as at least the amount of the royalties or fees which would have been due if the infringer had requested authorisation to use the patent in question.

Legal costs and other expenses incurred by the successful party shall, as a general rule, be borne by the unsuccessful party.

The Agreement was signed by 25 EU member states on 19 February 2013. It will need to be ratified by at least 13 states, including France, Germany, and the UK to enter into force: at the moment, August 2014, the Agreement has been ratified by five member states.

Empirical Evidence

Characteristics and changes in the US patent litigation system have been studied by a large number of authors (Lanjouw and Schankerman 2001; Atkinson et al. 2009; Henry and Turner 2006).

Empirical evidence on patent litigation in the largest and most judicially active countries of the European Union is provided by recent studies, which give some real insights based on recent patent suits.

Graham and Van Zeebroeck (2014) analyzing a dataset of European patent litigation during 2000–2010, comprising approximately 9000

judicial patent decisions from seven European countries, show that the incidence of litigation and the bases of judicial outcomes diverge radically across the different countries and technology sectors. Relevant differences are also detected in the likelihood of patent litigants raising patent validity and infringement claims.

Litigation rates are highest in Belgium and France, whereas Germany and the UK show a low rate. Patent litigation varies widely across technology sectors, with the majority of cases in Europe focusing on patents granted for industrial processes, civil engineering, consumer goods, machinery, and transport technology.

Litigation costs significantly differ in European jurisdictions ranging from 50.000 to 200.000 € in France, Germany, and the Netherlands, but being considerably higher in the UK, 150.000–1500.000 €, which may explain the lower number of cases brought to court in the UK. It is worth noticing that the average cost in the USA is much higher ranging from 1.000.0000 to 10.000.000 €.

On the assumption that the Unified Patent Court will offer litigation at roughly the same cost level as the three largest low-cost national systems, Harhoff (2009), by utilizing different data sources, estimates that the total savings from the creation of the unified Patent Court are considerably larger than the actual operating costs, even for the most conservative scenarios. For 2013, the benefit-cost ratios would range between 5.4 and 10.5: in other words, duplication of litigation combined with high costs of litigation, in some countries, costs firms about 5.4–10.5 times more than the establishment and annual operation of the Unified Patent Court.

A comparison of 8,323 patent litigation cases across Germany, France, the Netherlands, and the UK, covering cases filed during the period 2000–2008 (Cremers et al. 2013), highlights relevant differences in the four jurisdictions concerning the number of case loads, settlement rate, average time for judgment, outcomes, characteristics of the litigants, fragmentation, sector distribution of litigants, and value of patents.

Out of a total of 6,739 cases in Germany, 5,121 are infringement cases heard by the three regional courts covered by our study, whereas 1,618 are revocation cases. By far the number of infringement cases heard by German courts exceeds the combined number of cases in all three other jurisdictions. Depending on how cases are defined, Germany has between 12 and 29 times as many litigation cases as the UK; the difference is similar with regard to the Netherlands; compared to France, Germany has around six times as many cases.

The settlement of disputes reveals relevant differences across countries: settlement is more likely in Germany (60% of cases) as compared with UK (40%). As for cases decided by a judge, revocation is the most likely outcome regardless of whether the initial claim was for infringement or revocation in UK, whereas infringement is in Germany and the Netherlands. France is characterized by a large share of patents that is held not to be infringed, but valid.

The time lag between the filing for a claim for infringement and a first decision is less than 1 year in Germany, the Netherlands, and the UK, nearly double in France. Claims for invalidity are decided fastest in the UK (11.2 months), but take a lot longer to be decided in Germany (15 months) and the Netherlands (11.4 months). Again, invalidity cases in France take significantly longer (19.8 months) than in any other jurisdiction.

There are large differences across jurisdictions with regard to case outcomes. Infringement cases with court decision amount to about 22% in Germany, 36% in the Netherlands, 14.7% in the UK, but only 5.6% in France where most patents are held valid, but not infringed. In the UK the large share of revoked patents of cases that allege infringement, 26%, is due to the fact that, in about 60% of cases alleging infringement, the defendant counterclaims for revocation.

Also outcomes of invalidity actions differ considerably across jurisdictions. Whereas in the UK 42% of patents are revoked if the case is decided by the judge, less than half as many invalidity cases end with revocation in Germany and France. The risk of infringing a patent that forms the subject of a revocation action is very low in all jurisdictions (4% in the UK and 7% in Germany).

Fragmentation leading to parallel litigation of the same patent in multiple jurisdictions is low in Germany (2%) and France (6%), but more relevant in the Netherlands (15%) and in the UK (26%). However, as the number of cases in the UK and the Netherlands is considerably lower than in Germany, the upper bound for the share of duplicated cases lies in Germany.

As for the characteristics of the litigating parties involved in the patent cases, half of all cases involve only domestic claimants in Germany and France. The share of cases with only domestic claimants drops below 40% in the UK and the Netherlands. The data look similar for defendants, with the exception of Germany where the share of cases with only domestic defendants exceeds 60%.

By sorting litigants by type – companies, individuals, universities, public research institutes, government, as well as international institutions – the largest differences in the shares of companies and individuals involved in patent cases are found across jurisdictions rather than between claimants and defendants. France, where there are almost twice as many individuals as defendants than there are claimants, is an exception. Overall the share of companies as claimants or defendants is smallest in Germany; on the contrary the greatest share of litigants in the UK falls into the "large" category. In all other jurisdictions, micro- and small companies represent the largest share of litigants.

As for the sector distribution of litigating companies, the share of pharmaceutical companies in the UK is the highest of the four jurisdictions amounting to 30%. In Germany, in contrast, companies are concentrated in manufacturing, notably the machinery and engine industry. In the Netherlands, the share of companies in the services industry (especially finance, insurance, and real estate) stands out. France does not show a strong characterization.

The number of forward citations received worldwide, a proxy for patent value, is significantly higher for the litigated patents compared to the group of non-litigated patents.

Strategic Use of Rules

First economic analyses of patent litigation stressed the role of divergent expectations (Priest and Klein 1984) and the presence of asymmetric information (Spier and Spulber 1993) in fostering litigation.

Later models argue that patent litigation reveals important information for potential entrants (Choi 1998), analyze patent enforcement through litigation when firms have private information (Llobet 2003), and compare the two doctrines of damages, lost profit and unjust enrichment (Schankerman and Scotchmer 2001).

More recently several studies have pointed out that the litigation system is likely to play a crucial role when patents are used as assets or as legal threats. Patent litigation can then be abused to extort licensing payments by patent-assertion entities (also known as "trolls"). Trolls engage in deliberate strategies whose aim is to acquire patents of failed companies or independent innovators using them to threaten suit against alleged infringers, without having the intention of actively using the patents they assert. In particular, in component-driven industries, notably information technology, trolls engage in deliberate tactics allowing them to take product developers by surprise once they have made irreversible investments (Lemley and Shapiro 2007).

As trolling activity, though not illegally, seeks to exploit structural and procedural weaknesses of the patent and judicial system to earn rents, an optimally designed patent litigation system should minimize the room for such welfare-reducing behavior.

The significant occurrence of trolls in the USA may find its roots in the high costs of legal proceedings, cost allocation rules (each party bears its own costs), contingency fees, high damage awards, and injunctive reliefs, characterizing the US litigation procedure. Moreover, questionable examination quality in patent granting and broadly defined patentable subject matter also play a role.

The weaker presence of "trolls" in Europe is presumably because the patent systems in Europe deviate from the US system in several crucial

points. Generally, court proceedings are much less costly, cost allocation favors the winning party, damage awards are not excessive, most courts have sought a careful balance between the rights of the parties, injunctions are not issued automatically, and the quality of patent examination has been considerably better than in the USA.

However, one should not assume that the European system is troll-proof: recently patent funds have acquired patent portfolios consisting of several thousand patents, largely European ones, and may seek to enforce them (Harhoff 2009).

Conclusions

The birth of the UPC will represent a considerable progress for the management of intellectual property in the EU in order to overcome duplication and inconsistencies and to lower litigation costs. The result will be a patent protection for all participating member states based on a single application and validation with a Unified Patent Court (UPC) with exclusive jurisdiction to hear infringement and invalidity actions.

A relevant argument put forth by the proponents of the UPC is that it will also reduce strategic behavior in Europe. This may be true for forum shopping, that is to say a strategic choice of courts venue by litigants to obtain a favorable outcome: nevertheless a new form of forum shopping might be originated by the UPC if local divisions behave differently with respect of the willingness to grant EU-wide injunctions and with respect to the attitude toward bifurcation.

As for the consequences of the EU-wide injunction, it is worth stressing that it might represent an incentive for "trolling activities," so far not so common in the EU, suggesting a cautious use of injunctions.

References

Atkinson S, Marco A, Turner J (2009) The economics of a centralized judiciary: uniformity, forum shopping and the federal circuit. J Law Econ 52:411–443

Choi J (1998) Patent litigation as an information-transmission mechanism. Am Econ Rev 88:1249–1263

Cotter T (2011) A research agenda for the comparative law and economics of patent remedies. Minnesota legal studies research paper no. 11-10

Cremers K, ErnickeM, Gaessler F, Harhoff D, Helmers C, McDonagh L, Schliesser P, Van Zeebroeck N (2013) Patent litigation in Europe. ZEW discussion paper 13-072 http://www.econstor.eu/bitstream/10419/83473/1/769014895.pdf

Directive 2004/48/EC. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:195:0016:0025:en:PDF

EPO (2013a) Patent litigation in Europe, 3rd edn. European Patent Academy, Munich. http://www.eplit.eu/files/downloads/patent_litigation_in_europe_2013_en.pdf

EPO (2013b) Agreement on a unified patent court. http://documents.epo.org/projects/babylon/eponet.nsf/0/A1080B83447CB9DDC1257B36005AAAB8/$File/upc_agreement_en.pdf

Frank R, DeFranco D (2000) Patent infringement damages: a brief summary. Fed Cir B J 10:281–291

Graham S, Van Zeebroeck N (2014) Comparing patent litigation across Europe: a first look. Stan Tec L Rev 17:655–708

Harhoff D (2009) Economic cost-benefit analysis of a unified and integrated European patent litigation system. Tender No. MARKT/2008/06/D. http://ec.europa.eu/internal_market/indprop/docs/patent/studies/litigation_system_en.pdf

Henry M, Turner J (2006) The court of appeals for the federal circuit's impact on patent litigation. J Legal Stud 35:85–117

Lanjouw J, Schankerman M (2001) Characteristics of patent litigation: a window on competition. RAND J Econ 32:129–151

Lemley M (2010) Where to file your patent case. AIPLA Q J 38:1–35

Lemley M, Shapiro C (2007) Patent holdup and royalty stacking. Tex Law Rev 85:1991–2049

Llobet G (2003) Patent litigation when innovation is cumulative. Int J Ind Organ 21:1135–1157

Love B (2009) The misuse of reasonable royalty damage as a patent infringement deterrent. Mon Weather Rev 74:909–948

Luginbuehl S (2011) European patent law: towards a uniform interpretation. Edward Elgar, Northampton

Priest G, Klein B (1984) The selection of disputes for litigation. J Legal Stud 8:1–56

Reitzig M, Henkel J, Heath C (2007) On sharks, trolls, and their patent prey—unrealistic damage awards and firms' strategies of "being infringed". Res Policy 36:134–154

Schankerman M, Scotchmer S (2001) Damages and injunctions in protecting intellectual property. RAND J Econ 32:199–220

Spier K, Spulber D (1993) Suit settlement and trial: a theoretical analysis under alternative methods for the allocation of legal costs. J Legal Stud 11:55–81

WIPO (2008) Chapter 2: Fields of intellectual property protection. In: Intellectual property handbook: policy, law and use. http://www.wipo.int/export/sites/www/about-ip/en/iprm/pdf/ch2.pdf

Patent Opposition

Alessandro Sterlacchini
Department of Economics and Social Sciences,
Università Politecnica delle Marche,
Ancona, Italy

Abstract

A patent opposition allows third parties to question the validity of the patents granted by the European Patent Office (EPO) on the grounds that they do not meet patentability criteria, do not fully disclose the invention, or extend beyond the original application. These issues are debated before an Opposition Division and, eventually, a Board of Appeal of the EPO which decides whether opposed patents are upheld as granted, amended, or revoked. The evidence indicates that these three possible outcomes are equally probable. Since the EPO decision applies to all the states designed in the application, the patent opposition represents a unique opportunity for challenging a patent's validity at European-wide level. Along with their relatively lower costs, this explains why, in Europe, patent oppositions are used by far more frequently than patent litigation.

Synonyms

Patent invalidation challenge; Post-grant patent review

Definition

A patent opposition is an administrative procedure adopted by the European Patent Office which allows third parties to challenge the validity of a granted patent on the basis of specific grounds.

Introduction

A patent granted by a public authority is presumed to be valid. This means that the patented invention is novel, based upon an inventive step (or "nonobvious"), and susceptible of industrial application (useful). Moreover, the patent application must disclose detailed information on the invention so that a person "skilled in the art" should be able to replicate it. To control for these requirements, patent examiners should have a relatively easy and cheap access to the relevant information about the state of prior art in specific technological fields. This condition is difficult to meet for many reasons: among them, the emergence of new fields, such as those of bio- and nanotechnologies, rooted on variegated but complementary disciplines, a staggering increase of patent applications, and a growing number of claims per application (Archontopoulos et al. 2007). To be stressed is that each claim identifies a specific property right that the patent should protect and, as such, must be validated by patent examiners.

The number of patent examiners has not expanded in line with that of applications and claims, and this has determined a growing workload and backlog in patent offices. For having a granted patent at the EPO, an average of 3 years was necessary during the early 1980s, while in the early 2000s, the examination delay increased to 5 years (van Zeebroeck 2011). To avoid an excessive length of the examination process, the evaluation of each application has become less accurate, subject to possible errors, and, thus, likely to increase the number of low-quality patents susceptible of being invalidated.

A growing uncertainty about the validity of patents generates remarkable economic losses to society (Hall and Harhoff 2004). Patent holders could underinvest in some particular technologies, and their rivals could reduce investment in competing technological advances; if both subjects have already undertaken substantial investments on these activities, they will be prone to embark in costly litigation. If the cost and length of time for invalidating patents are too high, large companies with wide financial means have an incentive to inflate their patent portfolio with low-quality patents. In this way they create a strategic barrier to entry for small innovative

companies, and this will reduce the overall pace of innovation.

To avoid these negative consequences, including that faced by the companies that risk to be sued for infringing "wrongly" granted patents, an efficient and not too expensive procedure for reviewing their validity is necessary. In principle ex post litigation before civil courts can fix some important errors of patent offices. In practice, however, patent litigation appears to be a too costly mechanism, giving rise to extremely uneven incentives to challenge and defend issued patents (Farrell and Merges 2004): in particular, small firms and independent inventors are likely to be severely discriminated in favor of bigger and financially wealthier patent holders (Lanjouw and Schankerman 2001, 2004; Kingston 2004; Schettino and Sterlacchini 2009).

An effective and cheaper alternative to litigation is an administrative post-grant review in which informed persons or entities have the chance to disclose relevant pieces of information that were not available or not adequately taken into account during the patent's examination and that could undermine its validity. In the following, the procedure of patent opposition before the EPO is examined along with some evidence about the frequency, the determinants, and the outcomes of oppositions.

Institutional Features

European inventors can obtain patent protection by filing several national applications or, alternatively, one patent application to the EPO in which several states adhering to the European Patent Convention (EPC) are designed. Considering the relative costs, if patent protection is sought in at least four countries, the EPO-centralized route is more convenient. The granting of a European patent allows the applicant to achieve a bunch of patents that are valid in different countries (obviously, upon having paid the national fees and translating the documents).

According to Article 99 of the EPC, within 9 months from the publication of the mention of the grant of a European patent, any person can challenge its validity by filing an opposition against the granting decision of the EPO. Although most of the opponents are rivals of patent holders seeking to obtain the limitation or revocation of a patent, in principle, it is not necessary for the opponent to have or manifest a particular interest in the patented invention. The notice of opposition may be filed, jointly or separately, by more than one opponent.

Section 2 of the same article makes clear that "The opposition shall apply to the European patent in all the Contracting States in which the patent has effect." Thus, in Europe, the opposition at the EPO represents a unique opportunity to challenge the validity of a patent at the European-wide level rather than in individual national courts. It should be added that only in 13 out of the 38 states adhering to the EPC, there is a procedure for post-grant patent oppositions (EPO 2013). Among the largest countries, only in Germany the patent law provides for invalidity actions in a unique national court (the Federal Patent Court) separated from other civil courts specialized in litigation for patent infringements. Instead, post-grant oppositions are not allowed in France, the UK, and Italy, while only to a limited extent in Spain. Thus, multiple suits before the civil courts of different countries could be necessary to invalidate a European patent. Compared to the centralized procedure at the EPO, such an alternative implies not only more costs (see below) but also a non-negligible level of uncertainty: in fact, it cannot be taken for granted that different national courts will achieve the same decision.

The grounds for opposition, established in Article 100 of the EPC, are the following:

1. The subject matter of the patent is not patentable (lack of novelty, inventive step, and industrial applicability, according to Articles 55–57 of the EPC).
2. The patent does not disclose the invention in a manner sufficiently clear.
3. The subject matter of the patent extends beyond the content of the filed application.

The opposition process is overseen by an Opposition Division, composed of three technical examiners of the EPO, at least two of whom must not have taken part in the proceedings for granting the opposed patent (Article 19, EPC). If the complexity or specificity of the case so requires, the Opposition Division can decide to include an additional legally qualified examiner who has not taken part in the proceedings for grant. The same requirement holds for being the Chairman of the Opposition Division. In case of parity of votes, that of the Chairman is decisive.

After 2 months from the filing of an opposition, the opponent must present his/her arguments and evidence for asking a revision of the EPO decision. Then, the patent holder has up to 6 months to reply, and the same time is allowed to the opponent for his/her counterarguments. After the exchange of observations, an oral hearing of arguments (normally open to the public) takes place before the Opposition Division. Then, the Division communicates to the parties its decision. On average, all the process is accomplished in about 22 months (Harhoff 2005).

The three possible outcomes from an opposition proceeding are the following (Article 101, EPC):

1. The opposition is rejected and the patent is upheld as granted.
2. The patent is maintained with amendments based on reformulations or cancelations of claims.
3. The patent is revoked.

The parties adversely affected by the above decisions may appeal to the EPO's Boards of Appeal, and this additional procedure can last almost 26 months. If a Board of Appeal confirms the revocation decision, the patent is invalid in all the states designated in the application. If the patent is maintained as granted, the opponents can resort to invalidity actions in the national civil courts of the designated states. In principle, the same possibility is allowed for amended patents, but the modification or cancelation of claims should reduce the likelihood of further legal disputes.

An important element that differentiates the European post-grant review from that of other countries (including the USA) is that the opposition procedure before the EPO is an adversarial process in which the legal representatives of the parties (opponent and patent holder) have the possibility of airing and debating their arguments before an adjudicator (Rotstein and Dent 2009). In this sense, an opposition proceeding resembles a validity suit before a civil court. However, while a patent litigation can be settled "out of court" (and this is what occurs in many circumstances), an opposition procedure may be continued by the EPO of its own motion even when the opposition is withdrawn (Rule 84, EPC).

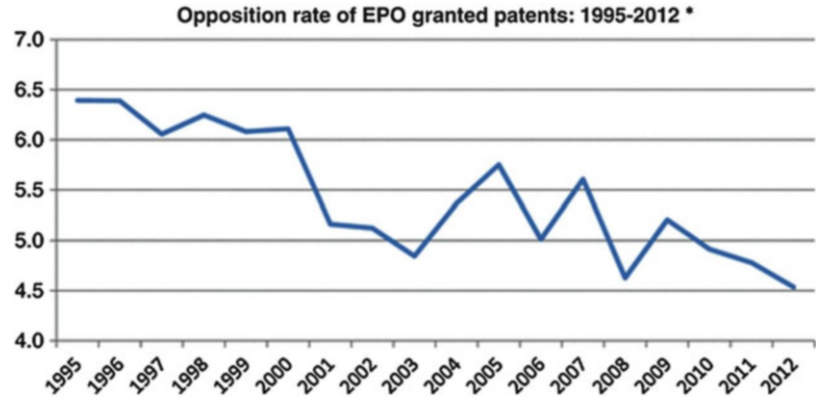
Empirical Evidence

The frequency of oppositions of the patents issued by the EPO was found particularly high during the first two decades of the office's life: the average opposition rate over the period 1980–1995 was about 8% (Harhoff and Reitzig 2004). Figure 1 shows that, during the subsequent years, the share of opposed patents has constantly decreased, a part from a slight recovery in the mid of 2000s.

From 1995 to 2000, the mean opposition rate was around 6%, while between 2000 and 2012, it reduced to 5%. The last annual report of the EPO concerned with the year 2013 documents a frequency of opposition equal to 4.7%.

The decrease of the opposition rate in the last decade is due to an almost constant number of opposition cases (about 2,900 per year), while the number of granted patents has increased from 60,000 in 2003 to a record of 66,700 in 2013. However, this does not mean that to obtain a patent from the EPO is becoming easier than it was in the past.

About 60% of all applications filed at the EPO between 1980 and 2002 ended with a grant (van Zeebroeck 2011). Instead, according to the last annual reports of the EPO, over the years 2006–2012, the share of granted patents on the processed applications (through the first search on prior art and, then, the proper examination) dropped to 47%. Almost 23% of the applications



Patent Opposition, Fig. 1 Opposition rate of EPO-granted patents: 1995–2012 (oppositions in year t on the average patents granted in year t and t − 1 (Source: EPO Annual Reports))

were withdrawn by the applicants after the search report, and another 30% were withdrawn or refused during or at the end of the examination process. The actual rate of refusals is not particularly high (around 5%), but it does not take into account that most of the withdrawals during the examination phase are induced by the "toughness" of EPO examiners who, by communicating detailed objections and remarks to the applicants, are able to discourage the less valuable applications (Lazaridis and van Pottelsberghe 2007).

Thus, being coupled with a remarkable drop of the share of granted patents, the observed decrease in the rate of oppositions could be a signal that the average quality of the patents granted by the EPO has improved over time giving less scope for invalidity challenges. This explanation is supported by the results of a survey jointly carried out in 2011 by Thomson Reuters and the *Intellectual Asset Management* magazine (issue of July/August 2011): the EPO was ranked first for patent quality among the world's five largest patent offices for the second consecutive year. Compared with the Japanese and, especially, the US patent office, the EPO was leading by a wide margin in terms of perceived patent quality and also improved its position with respect to previous years. Consistent findings are provided by de Saint-Georges and van Pottelsberghe (2013) who, by using a composite index of nine variables capturing the transparency of patent systems and the quality of examinations, show that the EPO ranks first among 32 patent offices. It should be added that such a high level of real or perceived quality was not achieved at the expense of the length in the granting process. On the contrary, in the years 2011–2012, the average delay to grant was lower than 4 years, a duration inferior to that recorded in the early 2000s (see above).

In spite of its declining trend, the rate of patent oppositions in Europe remains much higher than that of patent litigation before civil courts. Due to the presence of multiple and heterogeneous patent litigation systems among European countries, aggregate data on patent litigation at European-wide level are not available. However, some recent works have attempted to collect consistent data for some of the largest countries of the European Union.

Considering the judicial patent decisions from seven European countries published over the period 2000–2010, the litigation rate varies from a minimum of 0.1% in the UK to a maximum of 1.5% in the Netherlands (Graham and van Zeebroeck 2014). The two largest countries of the EU, Germany and France, record a frequency of patent litigation equal to 0.3% and 0.9%, respectively. These percentages underestimate the actual rate of patent disputes because those settled "out of courts" are neglected (for an analysis of patent litigation settlements in Germany, see Cremers and Schliessler 2015). However, it must be stressed that the majority of patent litigation refers to infringement rather than invalidity actions: recent data on patent litigation for France, Germany, the Netherlands, and the UK, collected from 2000 to 2008, show that only 22% of the patent disputes correspond to revocation cases (Cremers et al. 2013). As a consequence, the frequency of nullity actions before civil courts is

much lower than that of oppositions before the EPO.

The prevalence of patent oppositions as compared to invalidity litigation in Europe is mainly due to the different costs involved. Over the last decade, the total costs of an opposition before the EPO (including patent lawyers' fees) vary from €6,000 to €50,000 for each party (Mejer and van Pottelsberghe 2012). Instead, considering the proceedings before first-instance courts and patent cases of small and medium scales, the average cost of patent litigation ranges from €50,000 to €500,000, although in the UK, the maximum cost can be up to €1.5 million (EPO 2006). The country in which patent litigations are less expensive is Germany; much higher costs are documented in the UK, while France and the Netherlands record an intermediate level of litigation expenses (Cremers et al. 2013). It should be added that, in case of multiple litigation, the costs have to be cumulated across the national jurisdictions involved.

In terms of opposition rates, the differences among sectors or technological areas are remarkable. By considering the period 2000–2008, the opposition frequency of EPO-granted patent was found particularly high in the fields of chemicals and pharmaceuticals and lower in information and communication technologies (Caviggioli et al. 2013).

A large body of empirical evidence converges in showing that the opposition probability is significantly associated with the patent quality or value (Graham et al. 2003; Harhoff and Reitzig 2004; Cincera 2011; Schneider 2011). The latter can be approximated by different indicators: the most diffused and effective quality measures are the number of citations received by a patent (forward citations) and the size of patent families (given by the number of countries in which patent protection is sought for the same invention); other employed indicators are the number of backward citations (references to previous patents) and claims. The evidence suggesting that the most valuable patents (according to the above measures) are more likely to be opposed should be interpreted with some caution. In fact, an opposed patent that is revoked by the EPO cannot be considered of high quality. A similar consideration applies to patents with many claims that could be changed or canceled at the end of the opposition proceeding. In short, what can be safely said is that only the patents that survived an opposition have a higher quality and, as such, are more likely to be successfully enforced in subsequent legal disputes for infringement (Harhoff et al. 2003; van Zeebroeck 2011).

Another interesting issue that can be examined by using patent opposition data is whether the occurrence of an opposition could be due to strategic reasons. The relevance of this question is due to the fact that patent applications are significantly concentrated in a few hands: suffice it to notice that both in 2011 and 2012, the top ten applicants at the EPO accounted for almost 11% of the total applications filed. Are patent oppositions equally concentrated? Are some companies more exposed to the oppositions filed by direct competitors or more prone to challenge the patents held by industry rivals?

Although based on narrowly defined industries, the studies that employ data for individual companies (performing both the roles of attacked patent holders and challengers) suggest that the opposition procedure is essentially a game between the major industry players (Harhoff 2005; Schneider 2011). However, this does not mean that patent oppositions between the largest companies are undertaken for pure strategic motives, such as that of creating uncertainty and inducing the rivals to delay the commercialization of innovations.

In fact, on average, only about 30% of the oppositions before the EPO ended with a patent maintained as granted, that is, with a rejection of the opposition. The prevalent outcome of the opposition proceedings has been the revocation (34%) followed by the amendment (32%) of issued patents. The residual outcome (circa 4% of cases) corresponds to oppositions that were "closed" because the opposition was withdrawn or the patent lapsed because the patent holders stopped to pay the renewal fees. It should be stressed that the percentage of opposition rejected is quite stable over time, while in some years (such as in 2012 and 2103, according to the last

annual reports of the EPO), the share of patents upheld in amended form has been above that of revoked patents. In any case, the fact that around 66% of the opposed patents end with a more or less severe reduction of the property rights of patentees clearly indicates that the opponents are more successful than the defendants of granted patents.

Concluding Remarks and Future Directions

The above findings confirm that the opposition procedure adopted by the EPO is particularly effective in correcting the errors made in the first examination process, improving the quality of granted patents, and, then, reducing the chances of further litigation. In the absence of an effective post-grant review, the only way to fix the errors of patent offices would be that of challenging the patents' validity before national courts. However, as previously stressed, the cost of a patent lawsuit is much higher than that required to pursue an opposition case before the EPO. As a consequence, this kind of administrative patent review, by reducing the scope for further and more expensive litigation, is improving social welfare.

Especially on the basis of these considerations, many scholars have contended that also in the USA, a more effective post-grant patent review, resembling that adopted by the EPO, should be introduced (Graham et al. 2003; Hall and Harhoff 2004; Farrell and Merges 2004; Graham and Harhoff 2006). The Leahy-Smith America Invents Act, enacted into law in 2011, has introduced new procedures that have a limited duration and expand the bases for challenging the patents issued by the United States Patent and Trademark Office (USPTO). These new proceedings differ from the previous USPTO reexaminations which, contrary to the EPO oppositions, have been used by far less frequently than invalidity suits before civil courts. Also in Japan, in the light of the scanty use of the invalidation system (the only means, at present, for obtaining a patent revocation), the Japan Patent Office will probably reintroduce the procedure of post-grant opposition abolished in 2003.

With respect to the important changes that are coming in the European patent system, a final consideration is in order. The unitary patent, valid for the EU countries that signed the agreement, entered into force on 20 January 2013. However, it will apply only after the entry into force of a parallel agreement on a Unified Patent Court. For the purpose of our topic, the establishment of a unified court is important because it will have exclusive jurisdiction for litigation concerned not only with the new unitary patents but also with the current European patents. Both of them will be managed by the EPO, while users will be free to opt for one of the two systems. As a consequence, the availability of unitary patents will not change the procedures that the EPO currently adopts, including the opposition and appeal proceedings. However, the Unified Patent Court will reduce the excessive costs and, especially, the risk of divergent judicial decisions which, at present, make the recourse to patent litigation in Europe particularly burdensome and unlikely. Although not in the near future, this desirable institutional change could diminish the attitude to challenge the patents' validity by mainly resorting to the oppositions before the EPO.

References

Archontopoulos E, Guellec D, de la Van Pottelsberge PB, Van Zeebroeck N (2007) When small is beautiful: measuring the evolution and consequences of the voluminosity of patent applications at the EPO. Inf Econ Policy 19:103–132

Caviggioli F, Scellato G, Ughetto E (2013) International patent disputes: evidence from oppositions at the European Patent Office. Res Policy 42:1634–1646

Cincera M (2011) Déterminants des oppositions de brevets. Une analyse microéconomique au niveau belge. Rev Econ 62:87–99

Cremers K, Schliessler P (2015) Patent litigation settlement in Germany: why parties settle during trial. Eur J Law Econ. 40:185–208

Cremers K, Ernicke M, Gaessler F, Harhoff D, Helmers C, McDonagh L, Schliessler P, van Zeebroeck N (2013) Patent litigation in Europe. ZEW discussion paper no 13–072. http://ftp.zew.de/pub/zew-docs/dp/dp13072.pdf

de Saint-Georges M, de la van Pottelsberghe PB (2013) A quality index for patent systems. Res Policy 42:704–719

EPO (2006) Assessment of the impact of the European patent litigation agreement (EPLA) on litigation of European patents. Report of the European Patent Office acting as secretary of the Working Party on Litigation. http://www.eplaw.org/Downloads/EPLA_Impact_Assessment_2006_.pdf
EPO (2013) Patent litigation in Europe, 3rd edn. European Patent Academy, Munich
Farrell J, Merges P (2004) Incentives to challenge and defend patents: why litigation won't really fix patent office errors and why administrative patent review might help. Berkeley Technol Law J 19:1–28
Graham S, Harhoff D (2006) Can post-grant reviews improve patent system design? A twin study of US and European patents. CEPR discussion papers no 5680. http://www.cepr.org/pubs/dps/DP5680.asp
Graham S, van Zeebroeck N (2014) Comparing patent litigation across Europe: a first look. Stanf Technol Law Rev 17:655–708
Graham S, Hall B, Harhoff D, Mowery D (2003) Patent quality control: a comparative study of US patent reexaminations and European patent oppositions. In: Cohen W, Merril S (eds) Patents in the knowledge-based economy. The National Academic Press, Washington, DC, pp 74–119
Hall B, Harhoff D (2004) Post-grant reviews in the U.S. patent system – design, choices and expected impact. Berkeley Technol Law J 19:989–1015
Harhoff D (2005) The battle for patent rights. In: Peeters C, de la van Pottelsberghe PB (eds) Economic and management perspectives on intellectual property rights. Palgrave-Macmillan, London, pp 21–39
Harhoff D, Reitzig M (2004) Determinants of oppositions against EPO patent grants: the case of biotechnology and pharmaceuticals. Int J Ind Organ 22: 443–480
Harhoff D, Scherer F, Vopel K (2003) Citations, family size, opposition and value of patent rights. Res Policy 32:1343–1363
Kingston W (2004) Making patents useful to small firms. Intellect Prop Q 4:369–378
Lanjouw J, Schankerman M (2001) Characteristics of patent litigation: a window on competition. RAND J Econ 32:129–151
Lanjouw J, Schankerman M (2004) Protecting intellectual property rights: are small firms handicapped? J Law Econ 48:45–74
Lazaridis G, de la van Pottelsberghe PB (2007) The rigour of EPO's patentability criteria. World Patent Inf 29:317–326
Mejer M, de la van Pottelsberghe PB (2012) Economic incongruities in the European patent system. Eur J Law Econ 41:215–234
Rotstein F, Dent C (2009) Third-party patent challenges in Europe, the United States and Australia: a comparative analysis. J World Intellect Prop 12: 467–499
Schettino F, Sterlacchini A (2009) Reaping the benefits of patenting activities: does the size of patentees matters? Ind Innov 16:613–633
Schneider C (2011) The battle for patent rights in plant biotechnology: evidence from opposition filings. J Technol Transfer 36:565–579
van Zeebroeck N (2011) The puzzle of patent value indicators. Econ Innov New Technol 20:33–62

Path-Dependent Rule Evolution

Jan Schnellenbach
Brandenburgische Technische Universität Cottbus-Senftenberg Chair for Microeconomics, Cottbus, Germany

Definition

Path-dependent rule evolution occurs whenever the further change of formal or informal institutions is, at least to some degree, determined by the institutional history of a system.

How rules emerge and change

Different types of rules influence individual behavior. There are formal institutions, such as laws or self-adopted written rules of organizations (Furubotn and Richter 2005); there are informal institutions that are not captured in written form, such as social norms (Young 2008); and there are also habits or routines (Hodgson 2010; Vanberg 2002) that individuals themselves follow. A decision to implement and to follow such rules can be made consciously, but they can also evolve without any individual making a deliberate choice to change them. In any case, the evolution of rules is often path dependent.

Path dependence exists, simply put, when past events and decisions have an influence on and limit the scope of the future evolution of a system (David 2005). A simple example is the decision-making of individuals who have a preference for social approval and who attempt to infer from peer actions what the socially desired behavior is. At the initial stage, let there be a range of behaviors with roughly similar individual payoffs, so social

approval dominates the choice between them. In that case, it can happen that, given enough time, a vast majority of individuals coordinate on one type of behavior (Arthur 1994), which henceforth works like a social norm (Young 2008). It is important, however, that from the ex ante perspective, there were multiple possible equilibria, that is, various different kinds of social norms that individuals could have settled on. Path dependence implies that small differences in the early stages of the process, such as individuals randomly observing one kind of behavior rather than another, can have huge effects on the question which equilibrium is eventually chosen.

From an efficiency-oriented point of view, the biggest problem is that the equilibrium of a path-dependent selection process may not be efficient (David 1985), simply because chance and other individual motives than efficiency have driven the process. It may turn out that a different kind of norm would be, for example, associated with lower transaction costs or a more efficient utilization of technology. However, path dependence often leads to a lock-in (Arthur 1989) where, given the status quo social norm, there are little individual-level incentives to deviate from the norm. A change in the social norm therefore requires a widespread change of individual expectations regarding the desired behavior. This can occur as a result of political interventions, but also through decentralized processes such as social communication. If the latter occurs, the result often appears on the surface as a relatively sudden tipping from one social norm to another (Young 1998).

Some skepticism is however due with regard to the likelihood of efficient, deliberate changes of both informal social norms and formal laws through the political process. Path dependence in opinion-formation can lead to equilibria where publicly voiced political opinions, which are deemed false by a large majority of individuals, nevertheless dominate political discourse (Kuran 1995). Similarly, a majority of individuals can be easily locked-in believing factually false policy-related beliefs to be true and refusing to update them (Schnellenbach 2004). The fact that policy-making rests not on individual but shared beliefs (Denzau and North 1994; Bischoff and Siemers 2013), with very limited incentives for individuals to invest into holding factually correct beliefs, is therefore one factor that leads to a frequently observed persistence of inefficient rules.

The evolution of the rules that govern society is therefore to be understood as an interdependent process where informal institutions affect the evolution of formal institutions and vice versa (North 2005). In order to overcome inefficient lock-ins, a careful institutional policy-making is necessary, although not always sufficient (Eggertsson 2005).

References

Arthur WB (1989) Competing technologies, increasing returns and lock-in by historical events. Econ J 97:642–665

Arthur WB (1994) Increasing returns and path dependence in the economy. University of Michigan Press, Ann Arbor

Bischoff I, Siemers LHR (2013) Biased beliefs and retrospective voting. Pub Choice 156:163–180

David PA (1985) Clio and the economics of QWERTY. Am Econ Rev (P&P) 75:332–337

David PA (2005) Path dependence in economic processes: implications for policy analysis in dynamical system contexts. In: Dopfer K (ed) The evolutionary foundations of economics. Cambridge University Press, Cambridge, pp 151–194

Denzau AT, North DC (1994) Shared mental models: ideologies and institutions. Kyklos 47:3–31

Eggertsson T (2005) Imperfect institutions: possibilities and limits for reform. University of Michigan Press, Ann Arbor

Furubotn EG, Richter R (2005) Institutions and economic theory, 2nd edn. University of Michigan Press, Ann Arbor

Hodgson G (2010) Choice, habit and evolution. J Evol Econ 20:1–18

Kuran T (1995) Private truths, public lies. The social consequences of preference falsification. Harvard University Press, Cambridge, MA

North DC (2005) Understanding the process of economic change. Princeton University Press, Princeton

Schnellenbach J (2004) The evolution of a fiscal constitution when individuals are theoretically uncertain. Eur J Law Econ 17:97–115

Vanberg VJ (2002) Rational choice vs. program-based behavior. Ration Soc 14:7–54

Young HP (1998) Individual strategy and social structure. Princeton University Press, Princeton

Young HP (2008) Social norms. In: Durlauf SN, Blume LE (eds) The New Palgrave dictionary of economics online, 2nd edn. Palgrave Macmillan, Basingstoke

Permit Trading

► Emissions Trading

Piracy, Modern Maritime

Paul Hallwood and Thomas J. Miceli
Department of Economics, University of Connecticut, Storrs, CT, USA

Abstract

This entry provides an economic analysis of the problem of modern-day maritime piracy by first reviewing the current scope of the problem and then developing an economic model of piracy that emphasizes the strategic interaction between the efforts of pirates to locate potential targets and shippers to avoid contact. The model provides the basis for deriving an optimal enforcement policy, which is then compared to actual enforcement efforts that, for a variety of reasons, have largely been ineffectual. The entry concludes by reviewing the law of maritime piracy and by offering some proposals for improving enforcement.

Synonyms

Hijacking; Looting; Robbery

Definition

An act of robbery or criminal violence committed at sea.

This entry develops an economic approach to the problem of modern-day maritime piracy with the goal of assessing the effectiveness of remedies aimed at reducing the incidence of piracy. To date, these efforts have been largely ineffectual for several reasons, including gaps in domestic laws, reluctance of countries to bear the expense of imprisoning pirates, and the general lack of an effective international legal framework for coordinating and carrying out enforcement efforts. Indeed, it is the absence of such a framework that bedevils international public law as a whole, not just in the area of maritime piracy.

The theoretical framework is based on a standard Becker-type model of law enforcement (Becker 1968; Polinsky and Shavell 2000), extended to consider the effort level of pirates to locate and attack target vessels and of shippers to invest in precautions to avoid contact. The model provides the basis for prescribing an optimal enforcement policy whose goal is to minimize the cost of piracy to international shipping. It also serves as a benchmark for evaluating actual enforcement efforts within the context of international law (such as it exists). The entry concludes with several proposals aimed at improving enforcement.

Modern-Day Maritime Piracy

Modern-day maritime piracy is a worldwide phenomenon. Over 2,600 attacks, actual or attempted, were reported over the period 2004–2011, but with some recent decline due to the effort of naval task forces as well as a very large increase in the use of onboard armed guards. For example, the most recent data shows that in the first 11 months of 2013, there were 234 boardings worldwide, with 30 in Nigerian waters and 13 in Somali waters. Many incidents have also been reported in Southeast Asia, especially off Indonesia.

Somali pirates principally operate a capture-to-ransom model, with ransoms of up to $5.5 million per incident being collected. Elsewhere in the world, robbery is the main motive. The overall economic cost of maritime piracy in 2012 was estimated at $6 billion, down from $7 billion the year before and as much as $16 billion a few years earlier. Spending on onboard security equipment and armed guards increased from about $1 billion to $2 billion between 2011 and 2012. Other economic costs include additional travel days as a consequence of rerouting of ships; increased insurance costs of as much as $20,000 per trip; increased charter rates, as longer time at sea

reduces the availability of tankers; cost of faster steaming through pirate-affected seas; and greater inventory financing costs for cargoes that remain longer at sea (Bowden 2010). Also, according to, an additional 10 attacks are associated with an 11% decrease in exports between Asia and Europe at an estimated cost of $28 billion.

With regard to antipiracy efforts, Anderson (1995) notes that there are economies of scale in this activity. When trade on a given shipping route is sparse, individual merchant ships have to arm themselves, thereby duplicating investment. However, with greater amounts of trade, several shipping companies may reduce costs by hiring armed ships for their protection as they sail in convoy. And with still greater shipping traffic, the least cost protection method has turned out to be patrolling of large areas of ocean space by warships. Today all three methods are used.

Not surprisingly, the efficiency of the pirate organization contributes to its success, both historically and in modern times (Leeson 2007; Psarros et al. 2011). Accordingly, present-day Somali pirates have developed supportive "social" organizations that aid them on land and at sea (Bahadur 2011). Pirate leaders often require new recruits to swear allegiance to the organization and its leaders until death; many Somali pirates are ex-coast guardsmen or ex-militiamen and share a common background and training; there is a common belief that ransoms are like a tax on foreigners who are overfishing Somali waters; and there is even the use of stock exchanges to finance operations.

An Economic Model of Piracy

This section develops a simple model of maritime piracy that focuses on its harmful effects on shipping (Guha and Guha 2010; Hallwood and Miceli 2013). The model accounts for both the efforts of pirates to locate potential target ships and of shippers to avoid contact with pirates. In this sense, the model extends the standard economic model of crime to account for precautionary behavior of potential victims (Shavell 1991; Hylton 1996). After deriving the equilibrium of the model, we examine optimal enforcement policies.

The model focuses on a representative pirate and a representative shipper who traverse the same geographic area over a fixed period of time. The pirate devotes effort x (measured in dollars) to locate a target vessel, and the shipper invests precaution y (also in dollars) to avoid contact. The pirate's effort represents the amount of time at sea and/or the number of boats, while shipper's avoidance can represent the use of alternate (more expensive) routes, less frequent or fewer voyages, or the use of armed escorts. Let $q(x, y)$ denote the probability of a contact over a given time period, where $q_x > 0$, $q_{xx} < 0$, $q_y < 0$, and $q_{yy} > 0$. Thus, pirate effort increases the chances of an encounter, while shipper precaution reduces the chances, both at decreasing rates. The cross partial q_{xy} may be positive or negative, as discussed in more detail below. (A common formulation is $q(x,y) = x/(x + y)$.)

The benefit to the pirate from an encounter is the loot, which can take the form of confiscated cargo, ransom of passengers, or both. Let G be the gross expected gain from an encounter. The net gain, however, must account for the possibility of capture and punishment. Let p be the probability of capture and s the (dollar) sanction upon conviction, both of which the pirate takes as given. Thus, the net gain per encounter is $G–ps$, which we will assume is positive. (This will necessarily be true if $G > \bar{s}$, where $\bar{s}$ is the maximal sanction. We discuss the nature of s in greater detail in the section on enforcement below.) At the time it makes its choice of effort, the pirate's expected gain is therefore

$$q(x,y)(G - ps) - x. \tag{1}$$

The pirate chooses x to maximize this expression, taking as given y, p, and s. The resulting first-order condition is

$$q_x(G - ps) - 1 = 0, \tag{2}$$

which defines the pirate's reaction function, $\hat{x}(y,ps)$.

The shipper expects to earn gross profit of π, which will be reduced by any expected costs associated with the threat of piracy. These costs include the losses inflicted directly by the pirate, denoted h (including the loss of cargo as well as damage to the ship and harm to crew members), plus the cost of avoidance actions, y. The net expected return to the shipper is therefore

$$\pi - q(x,y)h - y. \tag{3}$$

The shipper chooses y to maximize Eq. 3, taking x as given. This yields the first-order condition

$$q_y h + 1 = 0, \tag{4}$$

which defines the shipper's reaction function, $\hat{y}(x)$.

The Nash equilibrium occurs at the point where the reaction functions intersect. Differentiating Eq. 2 yields the slope of the pirate's reaction function

$$\frac{\partial \hat{x}}{\partial y} = \frac{-q_{xy}}{q_{xx}}, \tag{5}$$

which has the sign of q_{xy} given $q_{xx} < 0$, while differentiating Eq. 4 yields the slope of the shipper's reaction function

$$\frac{\partial \hat{y}}{\partial x} = \frac{-q_{xy}}{q_{yy}}, \tag{6}$$

which has the opposite sign of q_{xy} given $q_{yy} > 0$. The equilibrium, which we assume exists and is unique, is shown graphically in Fig. 1.

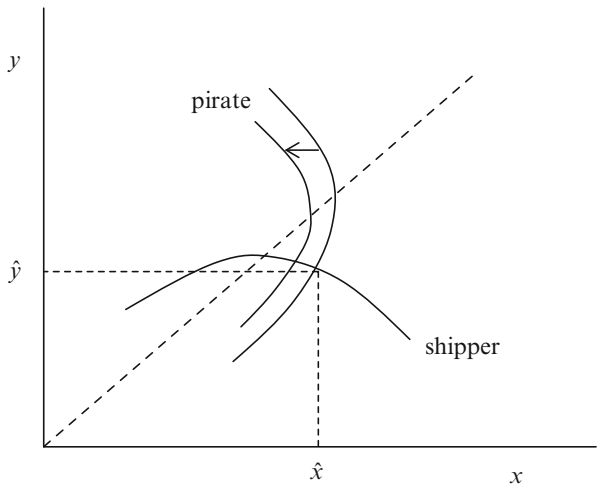
The Impact of Antipiracy Laws

Enforcement laws against piracy involve efforts to capture and punish pirates. Below we discuss the implementation of these laws in practice; here we examine their impact in theory, given the preceding equilibrium.

Law enforcement directly affects the behavior of pirates through the expected punishment term, ps, while it indirectly affects shipper behavior through their response to the resulting change in pirate behavior. Consider first the effect of changes in ps on the behavior of pirates. (Note that, given risk neutrality, it does not matter whether this is due to a change in p, s, or both.) Differentiating Eq. 2 yields

$$\frac{\partial \hat{x}}{\partial ps} = \frac{q_x}{q_{xx}(G - ps)} < 0, \tag{7}$$

given $G–ps > 0$. Thus, an increase in the expected sanction for piracy reduces the pirate's investment in effort for any y. In Fig. 1, this results in a leftward shift of the pirate's reaction curve. The new equilibrium involves an unambiguous reduction in the pirate's equilibrium level of



Piracy, Modern Maritime, Fig. 1 Equilibrium choices of pirate effort (x) and shipper avoidance (y)

effort, but the effect on the shipper's investment in avoidance is ambiguous. As drawn, $\hat{y}$ goes up, but it should be apparent that it could also go down, depending on the location of the initial equilibrium and the amount that the pirate's reaction function shifts. The intuitive reason for these effects is as follows.

The negative effect of greater enforcement on the pirate's effort reflects the standard deterrence argument – a higher expected sanction lowers the marginal benefit of criminal activity. The ambiguous effect of enforcement on the shipper's precaution hinges on the sign of q_{xy}. For the case shown in Fig. 1, $q_{xy} > 0$, so as the pirate's effort declines, the marginal benefit of shipper precaution increases (i.e., q_y becomes more negative), causing an increase in y. In this case, enforcement of laws against piracy and shipper precaution are complementary. However, the reverse would be true if $q_{xy} < 0$, for in that case, greater law enforcement efforts, by lowering x, would substitute for, or "crowd out," shipper precaution. The actual outcome is therefore an empirical question.

Given the preceding effects of increased enforcement, we now turn to the derivation of the optimal enforcement policy, which involves the enforcement authority (whoever that may be) choosing the probability of apprehension, p, and the sanction, s, to maximize social welfare. An important question here concerns whether or not to count the pirate's gains as part of welfare. The convention in the economics of crime literature has been to count the offender's gains, but there are differing views on this issue (Polinsky and Shavell 2000, p. 48). In the case of pure theft, the value of the loot is simply a transfer payment and thus would drop out of welfare if the thief's gains are counted (Shavell 1991). However, if the gains and losses differ, the possibility arises that the transfer could actually be value enhancing – an "efficient theft" – which most people would find objectionable, especially in the piracy context. Thus, although we will follow the standard convention and count the pirate's gains in welfare, we will assume that the loss suffered by the shipper exceeds the pirate's gains – that is, $h > G$. Consequently, any act of piracy is necessarily inefficient. This could reflect damages or harm to victims on top of the simple transfer of wealth, as well as any fear or "pain and suffering" incurred by victims of piracy and their sympathizers.

Based on these considerations, we write social welfare as

$$W = \pi - q(\hat{x}(ps),\hat{y}(ps))(h - G + p\beta s) - \hat{x}(ps) - \hat{y}(ps) - c(p), \tag{8}$$

where $\hat{x}(ps)$ and $\hat{y}(ps)$ are the equilibrium levels of pirate effort and shipper precaution, which depend on ps in the manner described above. The total expected enforcement costs are $c(p) + q(\hat{x}(ps),\hat{y}(ps))p\beta s$, where $c(p)$ is the cost of deploying more ships ($c' > 0$, $c'' \geq 0$) and β is the unit cost of increasing s. The enforcement problem is to choose p and s to maximize Eq. 8, subject to $p \in [0, 1]$ and $s \in [0,\bar{s}]$, where $\bar{s}$ is the maximal sanction. The possible interpretations of $\bar{s}$ are (i) the maximum prison term the offender could serve (e.g., life), (ii) a death sentence, or (iii) the harshest punishment that the country charged with carrying out the punishment is willing to impose (as discussed further below).

We begin by deriving a standard result in the law enforcement literature – namely, that $s* = \bar{s}$, or the optimal sanction is maximal (Polinsky and Shavell 2000). To see why, suppose that $s < \bar{s}$ and $p > 0$. Now raise s and lower p so as to hold ps fixed. As a result, all of the terms in Eq. 8 that depend on ps remain unaffected, but $c(p)$ falls, thus raising welfare. This proves that $s < \bar{s}$ could not have been optimal. The intuition for this result is that the cost of s is only incurred if a pirate is actually captured, so overall costs are lowered by capturing only a few offenders and punishing them harshly.

With $s^* = \bar{s}$, we differentiate Eq. 8 with respect to p and, after canceling terms using Eqs. 2 and 4, obtain the following first-order condition for p^*

$$-q_x \frac{\partial \hat{x}}{\partial p}(h + p\beta\bar{s}) + q_y \frac{\partial \hat{y}}{\partial p}(G - p\beta\bar{s}) = c' + q\beta\bar{s}. \tag{9}$$

The left-hand side is the marginal benefit of increased enforcement, while the right-hand side

is the marginal cost. The first term on the left-hand side, which is positive, is the saved costs (victim harm plus punishment costs) as pirates reduce their efforts in response to an increase in the probability of apprehension. This is the direct gain from deterrence. The second term, which is ambiguous in sign, reflects the uncertain effect of an increase in p on shipper effort. Suppose $\partial\hat{y}/\partial p > 0$ (as is the case in Fig. 1), indicating that shippers increase their precaution in response to greater p (i.e., public enforcement and private precaution are complements). Let us also suppose that the term $(G - p\beta\bar{s})$ is negative, as would be true if pirate gains are not counted in welfare. In that case, the overall term is positive (given $q_y < 0$), thus amplifying the marginal benefit of enforcement. Intuitively, when public enforcement elicits increased private precautions, p should be raised, all else equal to encourage such precaution. Conversely, if $\partial\hat{y}/\partial p < 0$, the case of crowding out, the second term on the left-hand side is negative, which works in the direction of reducing p so as not to overly discourage private precaution by potential victims.

Enforcement Problems

The preceding represents the optimal enforcement policy in an ideal setting where there exists a single enforcement authority (or a unified coalition of enforcers), possessing both the will and the resources to carry out the policies implied by Eq. 9. While this may represent a reasonable assumption in many law enforcement contexts, enforcement of international laws against piracy is undertaken by multiple countries with varying degrees of interest in devoting resources to the effort. As a result, enforcement involves a problem of collective action, which may lead to several departures from the prescribed policy.

First, the gains from deterring piracy are enjoyed by all countries who make use of the shipping lanes threatened by pirate attacks. Thus, each country has an interest in reducing piracy in proportion to its expected losses. At the same time, however, deterrence of pirate attacks is a public good in the sense that actions by any one country to invest in enforcement will benefit all countries. Thus, each country has an incentive to free ride on the enforcement effort of others. Absent some form of credible commitment, therefore, those countries with the largest stake (e.g., the highest value of shipping in the affected area) will undertake the bulk of the enforcement, and all other countries will free ride on that effort. Actual enforcement will therefore be less than the efficient level.

A second factor discouraging enforcement efforts concerns the expected cost of imposing punishment once a pirate is apprehended. If this cost is borne entirely by the country that first apprehends the pirate, then enforcers will likely underinvest in an effort to reduce their probability of incurring that cost. This represents a kind of "reverse rent-seeking" problem in which individual countries underinvest in order to lower the chances that they will be the first to catch the pirate. Note that both of the above problems, which arise from the collective nature of enforcement of piracy laws, will arise in any law enforcement context involving overlapping or undefined jurisdictional boundaries. For example, similar problems plague the enforcement of laws against international drug trafficking (Naranjo 2010) and prosecution of the global war on terror.

A third enforcement problem concerns the credibility of threats to actually impose any punishment at all on a band of pirates once they are captured. Since the pirates' harmful acts are sunk by the time they are apprehended, enforcers may lack adequate incentives to incur the high costs of detention, trial, and final punishment. Although there may be incapacitative benefits of detention, the probability of any particular pirate committing further harmful acts is small compared to the high cost of punishment. As a result, it may be optimal (in a time-consistent sense) simply to release him. This issue is largely ignored in the economics of crime literature, where it is generally assumed that threats to prosecute and punish criminals are taken as given. The issue is amplified in the piracy context because of the absence of a well-established international tribunal that can develop a reputation over time for carrying out threatened sanctions.

A final problem concerns the choice of the sanction s. As the model showed, the optimal

sanction is maximal, but countries may interpret this prescription differently based on constitutional or other considerations, or they may set *s* based on criteria that differ from that described above (which, if they sympathize with the pirates, could involve setting no punishment at all). As a result, pirates will not be able to predict with any accuracy the actual penalty upon conviction, thereby diluting the deterrent effect of greater enforcement. Countries may also differ in their criminal procedures and evidentiary standards. Although countries can theoretically agree by treaty to uniform standards on these matters, philosophical differences regarding appropriate measures (e.g., based on disagreements over the appropriateness of the death penalty or sympathies for pirates) will make this difficult in practice.

International Law Governing Maritime Piracy

This section evaluates the efficacy of international law in light of the preceding analysis. Piracy is a crime under customary international law and is codified as such in the United Nations *Convention on the Law of the Sea* (UNCLOS) (ratified in 1994). Under this *Convention* states parties agreed to "cooperate" in policing the oceans outside of territorial waters and to arrest, prosecute, and imprison persons suspected and ultimately found guilty of piracy (Articles 100–107). In fact, these articles were taken verbatim from Articles 14–21 of the *Convention on the High Seas*, which was put into force in 1962.

The evidence on actual enforcement of international laws against maritime piracy, as defined by UNCLOS, suggests that these laws have largely been ineffective. For example, over the period between August 2008 and September 2009, Combined Task Force 151 and other navies in the Horn of Africa region disarmed and released 343 pirates, while only 212 others were handed over for prosecution (Ungoed-Thomas and Woolf 2009). The UN Security Council likewise reports that 90% of apprehended Somali pirates were released (UN Security Council 2011).

The discussion in the previous section suggests why this is the case: policing and enforcement is a public good or at least a mixed good with external benefits for third parties. There are, however, some other considerations as well. The first simply concerns those acts that meet the definition of piracy under the *Convention*. Acts must be for "private ends," suggesting that they must be motivated by the desire for material gain rather than for political purposes. Thus, terrorist acts would not meet the definition of piracy (Bendall 2010, p. 182) nor would hijacking or acts involving "internal seizure" of a ship by its crew or passengers (mutiny) under the so-called two-vessel requirement for piracy (Hong and Ng 2010, pp. 54–55).

A second difficulty, as discussed above, is the overlapping jurisdiction problem. UNCLOS only applies to acts of piracy on the high seas and in the 200-mile exclusive economic zones, and enforcement relies on the cooperation of all member states. Enforcement in the 12-mile territorial waters is the responsibility of the coastal state, and states vary both in their definitions of piracy and in the availability of resources or the will to enforce anti-piracy laws (Hong and Ng 2010, p. 55; Dutton 2010). Pirates will therefore naturally gravitate toward those areas where enforcement efforts are low or where antipiracy laws are weak. Of course, shippers will also avoid those areas (though at a cost of rerouting), so states with weak laws will suffer economic costs. However, because shipping lanes inevitably cross jurisdictional boundaries, some of those costs will be externalized.

A third problem with enforcement of international law, mentioned earlier, is the problem of successfully prosecuting those pirates who have been apprehended. Article 105 permits the apprehending state to prosecute offenders, but this has often been difficult both politically and logistically. For example, Fawcett (2010) notes that problems of transporting defendants and evidence gathering are significant impediments.

However, in Southeast Asia there has been some success in cooperation against piracy under the *Regional Cooperation Agreement on Combating Piracy and Armed Robbery against Ships in Asia*, which has been functioning since 2006 and now has 17 contracting parties (Noakes 2009). Under this agreement, the parties share

information and perform antipiracy patrols, especially in the Straits of Malacca. What may have helped in this instance is that in this region, there is relatively little area of high seas (none in the Straits of Malacca), and so policing is largely restricted to waters over which sovereign rights exit. As a result, the benefits of enforcement against piracy are more concentrated on the enforcing country, and hence there is less of a public good problem.

Proposals to Improve Enforcement

This section offers three proposals to improve the enforcement of antipiracy laws. The first, suggested by Dutton (2010), involves putting suspected pirates on trial in the International Criminal Court (ICC) rather than in the national court of the apprehending party. The ICC was created by the Rome Statute, which was ratified in 2002 and has 110 signatories, all of whom share in its costs according to an agreed-upon formula (Romano and Ingadottir 2000). However, while the Rome Statute grants the ICC jurisdiction over war crimes, crimes against humanity, genocide, and aggression, at present the ICC has jurisdiction only over the first three of these and that it will not be until 2017 that it can exercise jurisdiction over the crime of "aggression," which still has to be defined in law but under which piracy could conceivably be classified. Another difficulty with using the ICC against piracy is that some signatories may decline to finance the court for this purpose; that is, a states party choosing to free ride under UNCLOS may also wish to do so under a revised Rome Statute.

The second proposal involves extending the *Convention for the Suppression of Unlawful Acts against the Safety of Maritime Navigation* (SUA Convention) to piracy as well as to terrorism. The SUA came into force in 1992 and by 2011 it had 156 signatories and ratifications. This *Convention* is targeted at policing the oceans against criminal activities, though it specifically targets terrorism rather than piracy. The word "terrorism" appears five times in SUA, but the term is never defined, leading some to believe that, with appropriate reinterpretation, SUA could be used against maritime pirates (Hong and Ng 2010).

Cognizant of these features Noakes (2009), the chief maritime security officer for the Baltic and International Maritime Council (BIMCO) argued before a US House of Representatives Committee that SUA 1988 can and should be used to combat piracy and that it is incorrect to view this *Convention* as applying only to maritime terrorism and not maritime piracy. It is certainly true that SUA 1988 uses the word "terrorism" sparingly, and this could give the impression that it could be used in the context of maritime piracy. However, SUA 1988 grew out of UN General Assembly Resolution 40/61 in 1985, itself being a response to terrorism on the *Achille Lauro*, and Resolution 40/61 is clearly aimed at terrorist acts at sea and not piracy.

Still, Article 3 of SUA defines seven offenses, the first three being described as follows:

> Any person commits an offence if that person unlawfully and intentionally: seizes or exercises control over a ship by force or threat thereof or any other form of intimidation; performs an act of violence against a person on board a ship if that act is likely to endanger the safe navigation of that ship; or destroys a ship or causes damage to a ship or to its cargo which is likely to endanger the safe navigation of that ship.

Although an authoritative legal opinion by each signatory states party regarding exactly what crimes at sea the SUA encompasses has yet to be given, the US legislative attorney, R. C. Mason, working for the Congressional Research Service, has suggested, based on Article 3, that the SUA is directed at piracy as well as terrorism at sea (Ploch et al. 2010). However, this is only "guidance" and at present the US position on SUA and piracy remains unresolved.

Kilpatrick (2011) offers a third proposal, arguing that the UN Hague Convention (1970) could be extended from international civil aviation to maritime piracy. This Convention has been widely adopted, with 185 signatories as of 2013, and it compels states to either extradite or prosecute airplane hijackers. It also requires signatory states to punish terrorist acts by "severe penalties" through domestic laws. However, it is questionable that countries will move to extend the Hague

Convention to maritime piracy. This is true for several reasons. First, while the United States is of central importance in global civil aviation, it is much less so in international shipping. In civil aviation, US legislation has a significant impact on global regulation because foreign airlines and flights from foreign airports to the United States that do not meet US security standards are effectively prohibited from accessing its lucrative market. Second, countries are probably more strongly motivated to move against aircraft hijackings because each single incident is likely to affect more people, say, 250 persons on an airplane versus 20 or so on a ship. Finally, aircraft hijackings seem to be given much more prominence in the media than maritime hijackings.

Cross-References

- ▶ Hijacking
- ▶ Ransom Kidnapping

References

Anderson J (1995) Piracy and world history: an economic perspective on maritime predation. J World Hist 6:175–195

Bahadur J (2011) The pirates of Somalia: inside their hidden world. Pantheon, New York

Becker G (1968) Crime and punishment: an economic approach. J Polit Econ 76:169–217

Bendall H (2010) Cost of piracy: a comparative voyage approach. Marit Econ Logist 12:178–195

Bowden A (2010) The economic cost of maritime piracy. Working paper, One Earth Future Foundation

Dutton YM (2010) Bringing pirates to justice: a case for including piracy within the jurisdiction of the international criminal court. Working paper, One Earth Future Foundation

Fawcett J (2010) Challenges to apprehension and prosecution of East African maritime pirates. Marit Policy Manag 37:753–765

Guha B, Guha A (2010) Pirates and traders: some economics of pirate-infested seas. Econ Lett 111:147–150

Hallwood P, Miceli T (2013) An economic analysis of maritime piracy and its control. Scott J Polit Econ 60:343–359

Hong N, Ng A (2010) The international legal instruments in addressing piracy and maritime terrorism: a critical review. Res Transp Econ 27:51–60

Hylton K (1996) Optimal law enforcement and victim precaution. Rand J Econ 27:197–206

Kilpatrick RL (2011) Borrowing from civil aviation security: does international law governing airline hijacking offer solutions to the modern maritime piracy epidemic off the coast of Somalia? Oceans beyond piracy working Paper

Leeson P (2007) An-arrgh-chy: the law and economics of pirate organization. J Polit Econ 115:1049–1094

Naranjo A (2010) Spillover effects of domestic law enforcement policies. Intl Rev Law Econ 30:265–275

Noakes G (2009) Statement on international piracy. Testimony before the US House of Representatives, Committee on Transportation and Infrastructure, Subcommittee on Coast Guard and Maritime Transportation. February. (www.marad.dot.gov/documents/HOA_Testimony-Giles%20Noakes-BIMCO.pdf)

Ploch L, Blanchard CM, Mason RC, King RO (2010) Piracy off the horn of Africa. Congressional Research Service, 7–5700, R40528, April. (http://assets.opencrs.com/rpts/R40528_20100419.pdf)

Polinsky AM, Shavell S (2000) The economic theory of public law enforcement. J Econ Lit 38:45–76

Psarros G, Christiansen AF, Skjong R, Gravir G (2011) On the success rates of maritime piracy attacks. J Transp Secur 4:309–335

Romano C, Ingadottir T (2000) The financing of the international criminal court: a discussion paper. Center for International Cooperation, Project on International Courts and Tribunals, June, NYU and School of Oriental and African Studies, University of London

Shavell S (1991) Individual precautions to prevent theft: private versus socially optimal behavior. Intl Rev Law Econ 11:123–132

Ungoed-Thomas J, Woolf M (2009) Navy releases Somali pirates caught red-handed: a legal loophole has helped scores of Somali gunmen escape justice. The Sunday Times, 29 Nov

United Nations Security Council (2011) Report of the special advisor to the secretary general on legal issues related to piracy off the Somali coast: a plan in 25 proposals. 30 Jan 2011

Piracy, Old Maritime

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

Abstract

Eighteenth-century pirates were profit-maximizing criminals. Their infamous practices reflect strategies pirates adopted to bolster their bottom line. The "pirate code" was a system of constitutional democracy that sea

dogs developed to govern themselves privately in the absence of government. The "Jolly Roger" – pirates' black flag of skull and bones – was a signaling device that sea dogs used to facilitate the seizure of prizes without costly conflict. Pirates used heinous torture to develop a reputation that incentivized captives to acquiesce to their demands. And pirates "pressed" – or pretended to conscript – willing recruits to reduce sailors' legal cost of joining their crews.

Synonyms

Caribbean piracy; Eighteenth-century piracy; Golden Age piracy

Definition

Old maritime piracy refers to the seafaring banditry of early eighteenth-century (c. 1715–1730) criminals. These pirates preyed on merchant ships traveling through the Atlantic Ocean, Indian Ocean, Caribbean Sea, and Gulf of Mexico. The pirate population – at its height, an estimated 2,400 sailors (Rediker 1987, p. 256) – was drawn predominantly from the merchant marine, consisting mostly of ordinary seamen who voluntarily joined a pirate crew's ranks after the vessel they were sailing on had been seized by its attackers. Old maritime piracy's most infamous members included Blackbeard (real name, Edward Teach), Bartholomew Roberts, "Calico Jack" Rackham, and at least a few women, such as Rackham's crewmates, Anne Bonny and Mary Read.

Pirates lived and worked together at sea and on land. When "on the account" – sea dogs' term for pirating – they operated in independent crews averaging 80 members (Rediker 1987, p. 256). When not at sea, many pirates made homes in New Providence Island, Bahamas – a neglected and thus rogue-infested British colony. Unlike privateers, who were also active in the early eighteenth century and plundered merchant ships for profit, pirates paid allegiance to no government and lacked the sanction or support of any state. Pirates were considered, and considered themselves, "enemies of all nations."

The Decision to Pirate

During peacetime, when privateering employment was unavailable and naval employment was difficult to find, most sailors had two employment choices. One possibility was to work in the merchant marine. The pay in this employment was low. Between 1689 and 1740 the average able seaman earned between £15 and £33 a year (Davis 1962, pp. 136–137) – $4,000–8,800 in current US dollars.

Sailors' second occupational option was piracy. Piracy was a capital offense throughout eighteenth-century Europe, and pirating was a risky activity. Nevertheless, the potential financial rewards of pirating were sufficient to attract an estimated 4,000 sailors to illicit maritime plunder between 1716 and 1726 (Rediker 1987, p. 30). No data exist that would permit a computation of the average pirate's wage. However, the evidence that is available suggests that piracy, unlike work as an able seaman, could be extremely lucrative.

In the early eighteenth century, Captain John Bowen's pirate crew plundered a prize that resulted in a payout of £500 per man (Johnson 1726–1728, p. 480). Several years later the members of Captain Thomas White's crew each earned £1,200 from its cruise (Johnson 1726–1728, p. 485). In 1720 Captain Christopher Condent's crew seized a prize that earned each pirate £3,000 (Marx 1996, p. 161). And in 1721 Captain John Taylor's and Oliver La Bouche's pirate consort earned an astonishing £4,000 for each crewmember in a single attack (Marx 1996, p. 163). Although more modest prizes were more common, a successful expedition could yield a pirate a sum sufficient to retire. The prospect of immense financial reward that piracy offered was an important lure for sailors who left merchant-marine employment for risky careers as pirates.

A second, and perhaps equally important, consideration for many sailors who chose piracy over

legitimate maritime employment was the work and living conditions available on pirate ships versus those available on merchant ships. For reasons described below, captain mistreatment of ordinary sailors on merchant vessels – financial and physical – was a relatively common feature of merchant-marine employment. In contrast, on pirate ships officer abuse was far less common, rendering life aboard such vessels superior in many sailors' eyes.

Pirates' Profit-Maximizing Strategies

Organization

As criminals, pirates could not rely on state-made institutions of law and order to produce social cooperation among them. Yet to live and work together at sea for months at a time – i.e., for pirating, a necessarily jointly produced activity, to be possible at all – piratical cooperation was essential. Indeed, as persons committed to theft and murder for personal gain, the need for institutions of social order among pirates was likely more pronounced than it was among the members of legitimate societies, most of whom, in contrast to pirates, were not so willing to resort to violence to benefit themselves.

The success many pirate crews enjoyed suggests that pirates overcame the problem of social cooperation they confronted without government. Central to the way pirates accomplished this was their institutional organization grounded in constitutional democracy (Leeson 2007, 2009a). Two chief officers led early eighteenth-century pirate ships – the captain and the quartermaster. The former wielded command in times of battle, when chasing and engaging prey. The latter wielded command at all other times: the quartermaster was in charge of distributing crewmembers' victuals and payment and enforcing pirates' private laws. Pirate crews elected both officers democratically and deposed their officers popularly too.

Pirates' democratic selection of officers contrasts starkly with the selection of officers on early eighteenth-century merchant ships and naval vessels, who were appointed by owners and governments externally. Likewise, the division of power on pirate ships via the captain and quartermaster contrasts starkly with the centralization of power on period merchant and navy ships, where captains wielded autocratic authority over their crews. Pirates' institutional organization also differed sharply from that of early eighteenth-century European states, whose autocratic forms closely resembled those of legitimate, period sea vessels.

Pirates created their own social rules to prevent conflict and maximize the potential for piratical cooperation, which in turn maximized their crews' illicit profits. Pirates developed a body of private law, sometimes called the "pirate code," but which pirates called "articles of agreement" (Leeson 2009c). Each crew forged its own articles democratically and required the consent of every would-be member to its provisions before he was permitted to join the company.

The specifics of pirate articles differed by crew, but their basic elements were uniform. These included rules prohibiting theft and violence; terms governing crewmembers' compensation; regulations on activities that could generate negative externalities for other crewmembers, such as drinking and smoking; and terms of piratical social insurance, which was provided for crewmembers injured on the job. Pirate articles also established punishments for article violations and provided explicitly for democracy as the governing mechanism for important crew decisions, such as the selection of officers.

Remarkably, pirates enshrined their articles in writing as constitutions. Below is the constitution that governed pirate captain Bartholomew Roberts' crew aboard their ship the *Royal Fortune* (Johnson 1726–1728, pp. 211–212):

(I) Every Man has a Vote in the Affairs of Moment; has equal Title to the fresh Provisions, or strong Liquors, at any Time seized, and may use them at Pleasure, unless a Scarcity make it necessary, for the Good of all, to vote a Retrenchment.

(II) Every Man to be called fairly in Turn, by List, on board of Prizes, because, (over and above their proper Share) they were

on these Occasions allowed a Shift of Cloaths: But if they defrauded the Company to the Value of a Dollar, in Plate, Jewels, or Money, Marooning was their Punishment. If the Robbery was only betwixt one another, they contented themselves with slitting the Ears and Nose of him that was Guilty, and set him on Shore, not in an uninhabited Place, but somewhere, where he was sure to encounter Hardships.

(III) No person to Game at Cards or Dice for Money.

(IV) The Lights and Candles to be put out at eight a-Clock at Night: If any of the Crew, after that Hour, still remained enclined for Drinking, they were to do it on the open Deck.

(V) To keep their Piece, Pistols, and Cutlash clean, and fit for Service.

(VI) No Boy or Woman to be allowed amongst them. If any Man were found seducing any of the latter Sex, and carry'd her to Sea, disguised, he was to suffer Death.

(VII) To Desert the Ship, or their Quarters in Battle, was punished with Death or Marooning.

(VIII) No striking one another on board, but every Man's Quarrels to be ended on Shore, at Sword and Pistol.

(IX) No Man to talk of breaking up their Way of Living, till each shared a 1,000 l. If in order to this, any Man should lose a Limb, or become a Cripple in their Service, he was to have 800 Dollars, out of the publick Stock, and for lesser Hurts, proportionately.

(X) The Captain and Quarter-Master to receive two Shares of a Prize; the Master, Boatswain, and Gunner, one Share and a half, and other Officers one and a Quarter.

(XI) The Musicians to have Rest on the Sabbath Day, but the other 6 Days and Nights, none without special Favour.

Constitutional democracy among pirates predates its adoption in the United States by more than half a century. Equally surprising, pirates' reason for relying on this system of governance echoes the reason given in the Federalist Papers – or rather the other way around, since the Federalists did not put pen to paper until decades after pirates put constitutional democracy into practice. According to pirates, their reason for using constitutional democracy was to prevent their officers from abusing the authority that officers' positions of power necessarily conferred on them. Guarding against such abuse was particularly important to pirates, they suggested, given the mistreatment they had suffered in their previous lives as legitimate sailors at the hands of merchant captains who enjoyed near-dictatorial authority over their crews.

Pirate democracy permitted crewmembers to control the behavior of captains through the carrot of election and the stick of deposition. Moreover, the creation of multiple, popularly elected offices aboard pirate ships fostered competition between officers, which encouraged both captains and quartermasters to more faithfully adhere to the will of their crews. Circumscribing officers' authorities in written constitutions helped pirates ensure that their leaders wielded the powers they enjoyed in their crews' interests rather than for leaders' private gain. By making behaviors that constituted transgressions of those authorities explicit and requiring that all crewmembers consent to their terms ex ante, pirate constitutions created common knowledge about legitimate and illegitimate officer behaviors within pirate crews and helped coordinate pirates' response to officer behaviors.

The contrasting institutional organization of eighteenth-century pirate ships and merchant ships has an economic foundation (Leeson 2007). Merchant ships confronted a classic owner-employee, principal-agent problem. These vessels and their cargoes were financed by external financiers – wealthy landlubbers who funded commercial voyages. Since merchant-ship owners did not sail on the vessels they invested in, their valuable vessels and cargoes were beyond their eyes or reach when away at sea.

This situation invited sailor opportunism, such as stealing cargo, shirking in work activities, or

even absconding with vessels. To control such opportunism, merchant-ship owners appointed captains who they gave small shares in their ships and endowed with autocratic authority to monitor and financially and physically punish sailors who did not act in their interests. Autocratic captains solved owners' principal-agent problem. But in doing so they enabled captain self-dealing, which, as merchant sailors frequently complained, many captains exploited at their sailors' expense.

Pirates, in contrast, confronted no such owner-employee, principal-agent problem. The reason for this is simple: pirates stole their ships. On pirate ships, crewmembers were both owners and employees – principals and agents. Because of this, pirates did not require autocratic captains to control crewmember opportunism. They still needed leaders who could command in times of battle and administer rules that prevented crewmember conflicts. But in the absence of a divergence between principal and agent interests, pirates could democratically elect their leaders and divide authorities among them without loss, and indeed with substantial benefit: restraining the potential for captain self-dealing. The economic context of eighteenth-century merchant shipping rendered autocracy the efficient institutional organization on merchant ships. The different economic context of eighteenth-century pirating rendered constitutional democracy the efficient institutional organization on pirate ships.

Other Profit-Maximizing Strategies

Pirates employed several other practices to maximize the profitability of their criminal enterprise. Each of these practices is an infamous part of popular pirate lore. One example of this is pirates' black flag of skull and bones, the "Jolly Roger," which pirates used as a signaling device to distinguish themselves from less ominous maritime belligerents who merchant ships might confront and whose advances merchant ships were more likely to resist (Leeson 2009a). A second example is pirates' use of heinous torture against recalcitrant captives and their popular image among contemporaries as fiendish "hair-triggers," both products of pirates' manipulation of their public image, which pirates used to develop a reputation that encouraged prizes to surrender to them peacefully (Leeson 2010a). A third example is pirates' practice of "pressing" – or pretending to compel crewmembers into their service – a practice pirates used to manipulate laws against piracy that reduced merchant sailors' expected cost of joining pirates' ranks (Leeson 2009b, 2010b).

References

Davis R (1962) The rise of the English shipping industry in the seventeenth and eighteenth centuries. Macmillan, London

Johnson C (1726–1728) [1999] A general history of the pyrates . . .. Dover, Mineola

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

Leeson PT (2009a) The invisible hook: the hidden economics of pirates. Princeton University Press, Princeton

Leeson PT (2009b) The invisible hook: the law and economics of pirate tolerance. NY Univ J Law Lib 4:139–171

Leeson PT (2009c) The myth of the myth of social contract: the calculus of piratical consent. Public Choice 139:443–459

Leeson PT (2010a) Pirational choice: the economics of infamous pirate practices. J Econ Behav Organ 76:497–510

Leeson PT (2010b) Rationality, pirates, and the law: a retrospective. Am Univ Law Rev 59:1219–1230

Marx JG (1996) The pirate round. In: Cordingly D - (ed) Pirates: terror on the high seas – from the Caribbean to the South China Sea. Turner Publishing, Atlanta

Rediker M (1987) Between the devil and the deep-blue sea: merchant seamen, pirates and the Anglo-American maritime world, 1700–1750. Cambridge University Press, Cambridge

Political Competition

Anna Laura Baraldi

Department of Economics, Second University of Naples, Capua, Italy

Abstract

In this entry, we tried to explain the role of political competition in the economy. Political science provided various definitions of

political competition. After analyzing in detail all of those, we mentioned the actors political competition involves. In economic terms, competition in political market has seen as competition in good market. Therefore, economic science studied if this kind of competition is good or bad in terms of economic variables as, for example, growth.

Synonyms

Party competition: definitions, sources, and economic effects

Political Competition: Definitions, Sources, and Economic Effects

Political competition is a complex process at the heart of representative democracy. *It* can be defined as competition for political power, that is, for the right to shape and control the content and direction of public policy.

Broadly speaking, this process involves the interaction of a set of ***citizens***, each one with views about the relative desirability of conceivable states of the world; a set of ***constitutional rules***; a group of ***politicians*** who compete in the selection for constitutionally privileged positions within the decision-making process; ***political parties*** who choose to coordinate their behavior that they see as mutually beneficial; a group of ***political representatives*** in charge of representing citizens in the decision-making process; a group of ***office-holders,*** politicians selected under constitutional rules to hold privileged positions in the decision-making process; and a ***government***, defined as a unique coalition of office-holders in the most privileged positions in the decision-making process.

The unequal distribution of benefits and costs of living among different individuals and groups is the primary source of political competition in a society. Indeed, the decisions and actions of governments benefit, reward, and advantage only certain areas of the society, while others are compelled to bear more of the costs and burdens. Therefore, political competition is the continuing controversy over who controls the government and who officially decides how its formal-legal power is to be used.

Under a free and competitive, constitutional democratic political regime, different groups and organizations within the society engage in political competition. This can be for political authority and for political influence.

Competing for political authority means contending for the legitimate right to govern the entire political community (as carried out by the Labour and Conservative Parties in Britain, the Democrats and Republicans in the United States, and the Liberal and Progressive Conservative Parties in Canada); they compete for direct control of the government and for the legitimate right to officially decide how and for what purposes governmental authority is to be used. Each political party seeks to place its own political leaders in the majority of the government, thereby giving them the right to decide for and act in the name of the entire political society, exercising formal-legal authority to make and enforce authoritative, and binding decisions on public policy. The individual candidate, by seeking his/her party's nomination and then election by the voters to a higher government office (such as member of the legislature), is competing for the right to be a formal-legal participant in the processes of authoritative decision-making and action carried out by the government.

Competition for political influence means that *political interest groups* or *pressure groups,* whose members have common interests and views in a single area of public policy, seek to obtain and exercise political influence therein. Their objective is to influence the policy decisions and actions of the government in one or more areas of public policy. An interest group focuses on influencing and shaping public policy in their chosen policy area or areas, rather than attempting to have an impact on public policy in general.

Political competition overlaps two disciplines: economics and political science. The concept of political competition has been developed in the literature of political science and is directly related to the outcome of elections. It is commonly

asserted that the more competitive the parties, the more responsive the political system will be to the desires of the majority. Different interpretations of political competition have been given (Bardhan and Yang 2004) in the relevant literature. According to these interpretations, political competition generates costs and benefits. The definitions of political competition are the following:

1. Political competition as *accountability for incumbents*. This interpretation focuses on the process of political turnover. The intensity of political competition increases when the incumbent leaders can be removed more easily by the public and replaced by their competitors. Political competition affects the behavior of incumbent leaders today via tomorrow's threat of dismissal. The benefits of *accountability for incumbents* are due to the "accountability" of politicians: if the incumbent politician wants to maintain power, their incentive to respond to the public's wishes is stronger when their position is vulnerable. Therefore, more intense political competition makes the incumbent more accountable for his/her actions. This interpretation of political competition also generates costs linked to the strength of the threat of dismissal for the incumbent politician. Actually, when a degree of political competition is too high, the probability of reelection becomes sufficiently unrealistic; thus, the incumbent politician may abandon any hope of reelection in order to extract the maximum rents for himself/herself during his remaining time in office. This then shifts political incentives towards the short term. In an economic sense, this concept of political competition may be an important determinant of corruption of politicians.

   Political competition, in the sense of political turnover, may generate another cost. The cost is evaluated in terms of the incumbent politician's incentives to undertake public investments. Although public investments are growth enhancing, they are politically destabilizing: to keep his position secure, the incumbent may decide not to invest, given that investment must be financed by taxes. In this situation, when political competition is high, the destabilizing effects of public investment weigh more heavily on the incumbent, in many cases leading to a non-investment outcome (Acemoglu and Robinson 2000, 2002).
2. Political competition as *decentralization of political authority*. According to this interpretation, political competition is more intense when political authority is in the hands of a larger number of agents at any given point in time. This kind of political competitiveness has been largely studied in the literature on fiscal federalism. Competition between authorities representing distinct political jurisdictions induces each of them to become more politically efficient. In fact, authorities representing "efficient" political jurisdictions, in terms of limited corruption level or sound economic policy, have the opportunity of attracting mobile resources away from authorities representing inefficient jurisdictions. In these situations, competition between authorities is similar to competition between firms. As in the case for firms, competition between political authorities can generate economic costs whenever externalities are present. The model of fiscal policy is the theoretical framework used to illustrate this point. In this model, decentralized political authorities decide on public spending; they compare the marginal benefits of public spending to only a fraction of the social marginal costs. The reason is that the benefits of public spending are generally concentrated within a particular jurisdiction, or a particular interest group, while the costs are spread out across the whole of society. This leads, in equilibrium, to a depleted pool of public savings (Persson and Tabellini 2000; Drazen 2001).

   If we consider the competition in attracting mobile resources, inter-jurisdictional competition may generate potential costs and benefits (Besley and Case 1995). The model considers a multi-jurisdictional framework with heterogeneous elected officials; agents are *immobile,* but they can deduce information about their "types" of local officials by observing the behavior of officials in neighboring

jurisdictions. This leads local officials to be engaged in "yardstick competition," whereby each recognizes that his performance will be judged in relation to the performance of others. This form of competition can yield benefits to the people because, in order to maximize their probability of reelection, politicians have an incentive to "hide their true colors" (Besley and Case 2003), for example, by restraining taxation. However, this form of competition can also produce costs to the people in situations where this behavior is expected to prove too expensive; politicians may decide to abandon their hope of reelection in order to "seize all they can" from the public then and there.

It is possible that the "market for policy" view is not complete. This can be a result of inter-jurisdictional competition aimed at gaining mobile resources, which are also inclined to distinctive frictions. For example, agreements with coalitions of less mobile constituents are made instead of facing electoral incentives to disregard the intimidations of the mobile (Rodden and Rose-Ackerman (1997)). The distortionary effects can be quite serious when tighter complicity rather than greater numbers are reflected in the political power of the immobile; this is common for oligarchic landed interests.

For a model of the trade-off between local informational advantages of decentralization and the possibility of capture by the local elite, see Bardhan and Mookherjee (2006).

3. Political competition as *electoral competition*. This interpretation of political competitiveness focuses on the conflict between parties and elites to win public support. In this sense, political competition is the competition between essentially identical agents to *acquire* political power or between those already in power. Therefore, it is clearly related to both the first and second interpretations of political competition, which associate political competition with opportunities for political turnover, and decentralization of political authority, respectively.

Now we can highlight some key points characterizing political competition. Firstly, an increase in political competition (i.e., the number of political parties engaged in electoral competition) can be seen as a sign of the democratization of a society. Secondly, a major part of political literature states that the degree of political competition among political party systems is largely determined by the choice of the electoral system (Cox 1997; Duverger 1954; Lijphart 1994, 1999; Sartori 1976; Taagepera and Shugart 1989). Thirdly, the party system determines the degree of bargaining complexity that may affect government formation and maintenance (De Winter and Dumont 2003; Lijphart 1999; Müller and Strøm 2003; Van Roozendaal 1997) and feature among the determinants of public policy. In this entry, we will also treat political competition from an economic perspective. Since Marshall's "Principles of Economics" (1890), economists have been trained to believe that market competition maximizes the welfare of the consumers, whereas monopoly and market power create economic rents that make producers better off and consumers worse off. The literature of public choice, followed by political economy scholars, attempted to import this notion into the political market, using economic performance as a benchmark to evaluate the welfare properties of political equilibria (Persson and Tabellini 2000). Stigler (1972) studied analogies between economic and political competition. Political competition, as economic competition, is supposed to be welfare enhancing for the citizens. This is because more vigorous political competition implies the selection of politicians who are better at resisting interest group pressures to obtain transfers financed by distortionary taxation. On the contrary, electoral competition with few parties reduces the degree to which the political system represents the heterogeneous preferences of the electorate (Lipset and Rokkan 1967) and may also reduce the pool of available political talent (Becker 1983). Since modern times, it can be said that political competition enhances economic efficiency and makes a faster income growth possible.

Nevertheless, theoretical literature is not conclusive in stating whether political competition is or is not growth enhancing. In order to analyze how it affects economic performance, such

literature refers to political competition as electoral competition: competition among parties and candidates in elections to obtain public support through votes of citizens (Bardhan and Yang 2004). This concept of political competition seems close to that of accountability for incumbents (Persson et al. 1997): if political competition is intense, the incumbent politician is more accountable for he/she actions in office and he has an incentive to perform well; otherwise, he can be easily removed by the public and replaced with opponents.

If the votes market is considered as a goods market (with politicians competing with each other to win the elections – Becker 1983), the transmission mechanism linking political competition to growth implies that more intense political competition induces incumbent politicians to act in the public interest, maximizing their probability of being reelected (Mulligan and Tsui 2006). Otherwise, when political competition is intense, the electoral base of each party tends to be smaller. In order to cater to their narrow support base, politicians find it expedient to promise pork-barrel policies rather than policies that benefit the electorate as a whole. The resulting policies benefit the supporters of the winning politician, but do not necessarily maximize aggregate welfare (Lizzeri and Persico 2005).

Public Choice literature concentrates on the effect of political competition on government size, according to two interpretations: the interest group theories (Mueller and Murrel 1986) and the fiscal illusion view (Buchanan and Wagner 1977), both assessing a negative relation between political competition and government size. Lipford and Yandle (1990), taking an industrial organization view of the political market, find that the state share of total state and local tax revenues rises with party dominance.

A well-consolidated body of the relevant literature analyzes how political competition affects corruption which, in turn, affects economic performance. The theoretical literature in the field of the relationship between corruption and economic growth is split, for example, in the literature on political competition and corruption. Polo (1998) shows that intense political competition may alter the form of corrupt behavior. Policy distortions resulting from lobbying activities are likely to be greater when there is little electoral competition. However, when politicians use discretion over the way in which political contributions are spent, greater electoral competition increases the incentive to divert funds for personal use (Damania and Yalcin 2005). Persson et al. (1997) state that intense political competition implies that the incumbent politician is more accountable for his actions in office: either the incumbent has an incentive for good performance or he can be easily removed and replaced (Mulligan and Tsui 2006). Intense political competition may also lead to a low probability of reelection for the incumbent, as for a firm that may lose a share of the market if the latter becomes more competitive; in this case, an incumbent can act in a myopic manner, maximizing rents during his remaining time in office. To sum up, the overall effect of political competition on corruption is complex and difficult to define.

The empirical literature on the effects of political competition on economic growth is still poor and lacks in giving a unanimous answer. Recently, the analysis focusing on the different sources of growth (Pinto and Timmons 2005) has made the effect of political competition on growth unpredictable.

Besley et al. (2010), analyzing the relationship between political competition, economic policy, and economic performance in the United States, show that political competition may positively affect policy and economic growth via the "quality of politicians". Padovano and Ricciuti (2009) analyze the effect of an institutional reform (i.e., the change in the regional electoral system) on the competitiveness of Italian regional politics. They find evidence of a positive correlation among them for the 15 Italian regions with Ordinary Statutes. Alfano and Baraldi (2012) find an inverted U relationship between the degree of political competition and the growth rate; this result shows the possibility that an "optimal" level of political competition, in terms of growth, may exist. This optimal level resolves the trade-off between political accountability and government instability.

The complexity in the empirical literature on that topic is how to measure political competitiveness among political parties, and a number of measures of political competition have been proposed. Vanhanen (2000, 2003) measures competition by the share of votes captured by "minor" parties in parliamentarian elections, while Holbrook and Van Dunk (1993) and Rogers and Rogers (2000) use the "win margin" of the incumbent governor as a measure of competition, and Skilling and Zeckhauser (2002) focus on the length of time a party has been in office. Alfano and Baraldi (2012) construct a "political competition index" equal to one minus a (normalized) Herfindahl index for the political parties in a country, taking the percentage of votes for each political party at elections. All these measures have merits, but they often start from the presumption that some basic democratic structures are in place and they are not invariant to the choice of election rule and do not account well for party structure (Vanhanen 2000).

In order to overcome the drawback that measures of political competition have, Grofman and Selb (2009) have recently identified six properties that any index of competition should have:

1. The measure should be party specific, i.e., it should allow for the possibility that voters of different parties might have different incentives to turn out to vote.
2. For each party, the measure should run from 0 to 1, with 0 indicating situations where voter incentives to turn out are the least and 1 indicating situations where voters incentives to turn out are the greatest.
3. The measure should be summable over all parties to give a weighted average of overall incentives for turnout in a given district. This aggregate measure should, when appropriately normalized, still run from 0 to 1, with 0 again indicating situations where, in the aggregate, voter incentives to turn out are the least and 1 indicating situations where, in the aggregate, voter incentives to turn out are the greatest. The weights should reflect the vote shares of the parties, and aggregation to the legislature as a whole should not be distorted by variation in district population size as a function of district magnitude.
4. For each party, the maximum value should be obtained if the votes required to win its last seat (s) are such that a vote loss of one vote would convert a win for that seat (those seats) into a loss. The minimum value should be obtained if one candidate/party receives all the votes.
5. The measure should be sensitive to the nature of the voting rule being used. We propose that it should vary with the threshold of exclusion of that rule.
6. For two-candidate plurality elections, the measure should be reduced to a simple function of the difference in vote share between the winner and the loser.

To sum up, neither theoretical nor empirical literature clarifies whether an intense political competition is beneficial for economic growth or not. Nevertheless, the tendency of many democracies is to reduce the number of political parties engaged in electoral competition. A lot of electoral systems are set up to counteract the tendency of parties to multiply. Many parliamentary democracies have thresholds of exclusion that deny representation to parties with a vote share below the threshold (e.g., Germany has a threshold of 5%). With a threshold of exclusion, the number of parties is reduced because parties that anticipate a small vote share do not field candidates. The general point is that several features of electoral systems act to discourage the proliferation of parties. The main conjecture in support of that is that having a large number of parties creates government instability, which is thought to harm economic growth. In fact, when there are many competing parties, the electoral base of each party tends to be smaller. To cater to their reduced support base, politicians find it expedient to promise pork-barrel policies with narrow appeal, rather than policies that benefit the electorate at large. The resulting policies benefit the supporters of the winning politician, but do not necessarily maximize aggregate welfare. For instance, politicians must choose between creating a large or small amount of bureaucracy. The former entails an aggregate deadweight loss but generates some

benefits (jobs, etc.) that can be targeted to supporters; the latter, instead, is efficient but does not allow the politician to target largesse to supporters. Consequently, we should expect a large amount of bureaucracy. The idea is that projects which diffuse benefits (the minimum bureaucracy) are less appealing to office-motivated politicians because these benefits are less targetable and so may be under-provided by the political system. This distortion becomes worse as the number of competing parties increases. The reason is that the smaller the fraction of the electorate as the support base of each party, the higher the gain from targeting a smaller subset of the electorate, and therefore, the temptation for politicians to engage in special interest politics becomes greater.

References

Acemoglu D, Robinson J (2000) Political losers as a barrier to economic development. Am Econ Rev Pap Proc 90:126–130

Acemoglu D, Robinson J (2002) Economic backwardness in political perspective. NBER working paper no. 8831

Alfano MR, Baraldi A L (2012) Is there an optimal level of political competition in terms of economic growth? Evidence from Italy. Eur J Law Econ, on line first. https://doi.org/10.1007/s10657-012-9340-5

Bardhan P, Mookherjee D (2006) Corruption and decentralization of infrastructure delivery in developing countries. Econ J 116:101–127

Bardhan P, Yang T (2004) Political competition in economic perspective, Department of economics, UCB. Department of Economics, UCB, UC Berkeley. Retrieved from: http://escholarship.org/uc/item/1907c39n

Becker G (1983) A theory of competition among pressure groups for political influence. Q J Econ 98:371–400

Besley T, Case A (1995) Incumbent behavior: vote-seeking, tax- setting, and yardstick competition. Am Econ Rev 85(1):25–45

Besley T, Case A (2003) Political institutions and policy choices: evidence from the United States. J Econ Lit 41(1):7–73

Besley T, Persson T, Sturm D (2010) Political competition and economic performance: theory and evidence from the United States. Rev Econ Stud 77:1329–1352

Buchanan J, Wagner R (1977) Democracy in deficit: the political legacy of Lord Keynes. Academic, New York

Cox GW (1997) Making votes count: strategic coordination in the world's electoral systems. Cambridge University Press, Cambridge

Damania R, Yalcin E (2005) Corruption and political competition. World Bank, Mimeo

De Winter L, Dumont P (2003) Luxembourg: stable coalitions in a pivotal party system. In: Strøm K, Müller WC, Bergman T (eds) Coalition governance in western Europe. Oxford University Press, Oxford

Drazen A (2001) The political economy of macroeconomics. Princeton University Press, Princeton

Duverger M (1954) Political parties: their organization and activity in the modern state. Wiley, London/Methuen/New York

Grofman B, Selb P (2009) A fully general index of political competition. Elect Stud 28(2):291–296

Holbrook TM, Van Dunk E (1993) Electoral competition in the American States. Am Polit Sci Rev 87(4):955–962

Lijphart A (1994) Electoral systems and party systems: a study of twenty- seven democracies, 1945–1990. Oxford University Press, Oxford

Lijphart A (1999) Patterns of democracy. Yale University Press, New Haven

Lipford J, Yandle B (1990) Exploring dominant state governments. J Inst Theor Econ 146(4):561–575

Lipset S, Rokkan S (1967) Party systems and voter alignments: cross-national perspectives. The Free Press, Toronto

Lizzeri A, Persico N (2005) A drawback of electoral competition. J Eur Econ Assoc 3:1318–1348

Marshall A (1890) Principles of economics, vol 1, 1st edn. London, Macmillan. Retrieved 07 Dec 2012

Mueller D, Murrel P (1986) Interest groups and the size of the government. Public Choice 48:125–145

Müller WC, Strøm K (2003) Coalition government in western Europe. Oxford University Press, Oxford

Mulligan C, Tsui K (2006) Political competitiveness. NBER working paper no. 12653, Oct

Padovano F, Ricciuti R (2009) Political competition and economic performance: evidence from the Italian regions. Public Choice 138:263–277

Persson T, Tabellini G (2000) Political economics: explaining economic policy. MIT Press, Cambridge, MA

Persson T, Roland G, Tabellini G (1997) Separation of powers and political accountability. Q J Econ 112(4):1163–1202

Pinto PM, Timmons JF (2005) The political determinants of economic performance. Political competition and the sources of growth. Comp Polit Stud 38:26–50

Polo M (1998) Electoral competition and political rents. IGIER, Bocconi University, Mimeo

Rodden J, Rose-Ackerman S (1997) Does federalism preserve markets? Virginia Law Rev 83(7): 1521–1572

Rogers D, Rogers J (2000) Political competition and state government size: do tighter elections produce looser budgets? Public Choice 105:1–21

Sartori G (1976) Parties and party systems: a framework for analysis. Cambridge University Press, Cambridge

Skilling D, Zeckhauser K (2002) Political competition and debt trajectories in Japan and the OECD. Jpn World Econ 14:121–135

Stigler G (1972) Economic competition and political competition. Public Choice 13:91–106

Taagepera R, Shugart MS (1989) Seats and votes: the effects and determinants of electoral systems. Yale University Press, New Haven
Van Roozendaal P (1997) Government survival in western multi-party democracies: the effect of credible threats via dominance. Eur J Polit Res 32(1):71–92
Vanhanen T (2000) A new dataset for measuring democracy, 1810–1998. J Peace Res 37(2):251–265
Vanhanen T (2003) Democratization: a comparative analysis of 170 countries. Routledge, London

Political Corruption

Roy Cerqueti and Raffaella Coppier
Department of Economics and Law, University of Macerata, Macerata, Italy

Abstract

Political corruption represents a specific type of public-to-public corruption which implies that one participant of corrupt transaction belongs on the State and the other to the private sector: in fact, public corruption is a particular (and illegal) State-society relationship. Political corruption occurs when politicians, who are delegated to make laws and enforce them by the citizens, act themselves in a corrupt way. More precisely, it appears when policymakers exploit their political strength to pursue their own economic benefits and/or maintain their powerful position.

Introduction

To begin, it is important to define the word "corruption." According to Jain (2001), "Corruption is an act in which the power of public office is used for personal gain in a manner that contravenes the rules of the game." Here we are only referring to public-to-public interaction and not private-to-private. While public-to-public corruption usually involves a private party on one side and public official or member of State on the other, private-to-private corruption is any form of illegal transaction between associations or corporations within the private sector. When, for example, a director or employee gets involved in practices that are not part of his responsibilities and that are detrimental to the company he is working for but advantageous for himself or others, this can be defined as private-to-private corruption: "the type of corruption that occurs when a manager or employee exercises a certain power or influence over the performance of a function, task, or responsibility within a private organization or corporation" (Argandoña 2003).

For many years different political scientists have approached the subject of political corruption and the difficulty in defining the term. In some cases, certain practices may be considered legal (e.g., certain forms of bribery), consequently making corruption appear less prevalent or occasionally also having the adverse effect (Olken and Pande 2012). After Peters and Welch (1978) explained that "the systematic study of corruption is hampered by the lack of an adequate definition" and "What may be 'corrupt' to one citizen, scholars, or public official is 'just politics' to another, or 'indiscretion' to a third," many have tried to come up with different ways of solving the problem.

As we can deduct from Jain's definition, corruption usually stems from three different circumstances – weak institutions, discretionary power, and economic rents.

Firstly, regarding weak institutions, the rewards gained from corruption will always have to be more significant than the foreseen payoff from the process of being discovered and punished. Secondly, for corruption to occur, the State member in question must be able to control the designation of resources in a discretionary manner. Lastly, it must be possible to extract rents or create extractable rents on behalf of the discretionary power.

Therefore with regard to public-to-public corruption, as mentioned above, there are two sides – the State (politicians, bureaucrats, civil servants, officials, and anyone who has discretionary power over society) and the private party. When politicians (or any State members) take advantage of their position to unlawfully obtain resources to their advantage (or in favor of others), this is

considered to be a corrupt practice which is carried out on the boundary between public and private environments.

Business and financial corruption which are widespread and indeed any transactions which happen exclusively in the private environment, away from the public sector, will not be considered in this voice.

It is worth identifying what actually constitutes corruption. Some practices are illegal but are not actually defined as corruption because they are not linked to political power. These include money laundering, fraud, drug smuggling, and black market transactions.

Another important point is the close link between politics and organized crime. Furthermore, there is even a new type of "business politician" or "entrepreneur" who "combines mediation in (licit or illicit) business transactions, first-hand participation in economic activity, and political mediation in the traditional sense" as stated by della Porta and Pizzorno (1996).

There is a specific analysis of corruption, known as the "principal-agent analysis" (Rose-Ackerman 1978), which involves considering three main agents or "actors": the principal (the State), the private agent (citizen or company), and an agent (a public official) who in theory follows the interests of the principal. If the private party wants to obtain an advantage, he pays a bribe to the public official to help him instead of carrying out the interests of the principal. This is considered to be corruption.

Corruption can happen on two levels – bureaucratic and political. Bureaucratic or petty corruption is corruption in the public administration where bureaucrats (known as the "actors") use their public position for financial or nonfinancial benefits. Political or grand corruption, on the other hand, is when politicians are corrupt and use their power for illegitimate financial gain and to maintain their political position. They are given the right to make decisions and to create laws, but they abuse this right in order to satisfy their own requirements.

Since political corruption has many different aspects, it is a difficult task to come up with a conclusive definition of the term. The characteristics and objectives of political corruption depend on the polity under discussion. Nye (1967) gives this statement about political corruption as "behavior which deviates from the formal duties of a public role (elective or appointive) because of private-regarding (personal, close family, private clique) wealth or status gains: or violates rules against the exercise of certain types of private-regarding influence." The various forms of political corruption can be recognized and analyzed by using an overall definition as a starting point.

It is worth noting that political and bureaucratic corruption are in fact inextricably linked as in the two different spheres their areas of competence can become interchangeable. However, for our objectives, the two different forms of corruption should be analyzed in practical terms.

In this encyclopedic voice, the focus is on political corruption, as opposed to general corruption; therefore one can concentrate on the political actors and the public environment in which they perform. Political corruption has a complex nature with a wide range of aspects which should be outlined. Although an extensive classification of the subject could reduce the analytical study, it is necessary in order to fully understand the different forms and characteristics.

Different Types of Political Corruption

Political corruption is not only a question of breach of formal legislations, code of conducts, and court rulings, but it also has a deeper impact on the entire political system. It can have a negative influence on decision-making, leading to the mishandling of procedures and finally the breakdown of political institutions.

While political corruption involves politicians carrying out illegal practices in order to maintain their position and improve their personal situation, it does not include police misconduct or suppression of political rivals. It is also important to clarify if the political corruption is taking place in democratic nations or dictatorial countries, due to the fact that there is lack of accountability between those in power and "the people." There

are problems with political corruption in dictatorial countries as the legal framework is already fragile, making it difficult to assess corruption and often leading to infringement by the leaders. Therefore the legal infrastructure of the State cannot be used as a point of reference to determine the extent of political corruption.

While bureaucratic corruption can often be approached by using administrative means (auditing, legislation, and institutional organization), political corruption with its detrimental consequences needs a much stronger solution. It is necessary to use legislative, moral, ethical, and political reference points first and foremost to differentiate between legality and legitimacy regarding political corruption.

Major political reforms are essential for fighting local political corruption. In the majority of dictatorial countries, political corruption is a routine event. As bureaucratic corruption is often rife in these regimes, this encourages political corruption, which obviously varies according to the different types of authoritarianism. Rulers of these countries use corrupt means in order to empower themselves and maintain their political status.

With regard to democracies, political corruption is more of a sporadic and incidental event and can therefore often be resolved by means of reforms and reinforcing the political framework and institutions of checks and balances.

Political corruption can be divided into different categories, mainly "personal" and "institutional" or "collective" corruption. The difference depends on what extent the financial (or other) rewards from the corruption are "privatized." While some may share the extraction with his associates, others will keep most if not all of the rewards for himself. "Personal" corruption is when politicians abuse their status to illegally obtain benefits for "personal" gain. A very common example is the acceptance of bribes by politicians and giving out benefits in return. On the other hand, "collective" corruption is a form of corruption which involves extracting resources for the benefit of a bigger group or organization, not just the private individual. Public power is abused for "private" gain, which does not necessarily mean for the personal benefits of one person but also for "collective" gain, for private groups. It is often the ruling parties and/or potential ruling parties, national governments, and administrative authorities that become involved in this form of corruption, using the resources at hand to their own advantage.

There is a significant distinction to be made between political corruption and "lobbying" when talking about "collective corruption." According to some (Damania et al. 2004; Harstad and Svensonn 2011; Campos ad Giovannoni 2007), the definition of "corruption" only refers to petty or bureaucratic corruption, whereas any form of practice done to influence policy makers should be referred to as "lobbying." However, this definition cannot be applied to all political systems as it does not distinguish between what is legal and what is illegal when influencing policy makers and does not take into account the regulations in all countries. While in the United States, where it is completely legal to provide pecuniary payments to policy makers (lobbying) and the definitions of petty corruption and lobbying may be applied, in other nations, the same payments are regarded as illegal and therefore politically corrupt.

It is worth looking at the procedures used to obtain private gains within the political environment in order to distinguish between the different forms of political influence (including lobbying) and political corruption. Since political corruption occurs when politicians obtain private benefits by means of nontransparent, unofficial procedures, one could say that lobbying is a particular form of corruption as it is still a way of gaining benefits from public entities or legislative organizations, with favors being given in return. However, lobbying and political corruption do have their differences which need to be noted. Lobbying is not always a case of bribes and campaign contributions. According to Austen-Smith and Wright (1994), lobbyists are capable of influencing politicians as they share competences and know-how which some politicians lack. Grossman and Helpman (1999, 2001) mention how some lobbyists even threaten politicians by saying they will give voters detrimental information about their policies, while others sway politicians with endorsements.

Other two types of corruption are "transactive" and "extortive" corruption. "Transactive" corruption is an illegal exchange between a donor and a recipient, where both actors benefit. "Extortive" corruption usually involves coercion of some sort so that the donor and people he is close to are not hurt in any way. Transactive practices are in theory based on reciprocity. It is a transaction between public officials (the State) and private actors (society) where both parties benefit from the exchange. In this corrupt exchange, there are two further distinctions to be made between "redistributive corruption" and "extractive corruption." With "redistributive corruption," it is the State who provides the private citizen or businessman with the resources, while "extractive corruption" is from the private actor to the State. Even though in theory the relationship is a reciprocal one, in reality this is rarely the case. In some nations or States in the world, the State plays the weaker role in transactions with society. Here redistributive corruption might take place as the mafia is rife and prevalent clientelism, which means that the State loses power as these private individuals or groups form a corrupt relationship with the State but end up with more benefits than the State itself.

"Extractive corruption," on the other hand, is when the State has the upper hand in the relationship. This is often seen in the political framework of various authoritarian States. The corrupting group or individual plays a passive role as the State (the corrupted) has the advantage in the corrupt exchange.

The Principal Causes of Political Corruption

It is a complex task to try and find the reasons behind political corruption. Rose-Ackerman (1996) states "Many officials remain honest in the face of considerable temptation, and others accept payoffs that seem small relative to the benefits under their control. Others, however, amass fortunes. The level of malfeasance depends not only on the volume of potential benefits, but also on the riskiness of corrupt deals and on the participants' moral scruples and bargaining power."

Various theories explain the widespread phenomenon of political corruption through cultural and moral factors typical of a country. When considering the factors that contribute to the onset and spread of corruption, an important role should be assigned to the cultural, ethnic, and social framework that fix in the "moral" of a country the seriousness of an act of corruption. Cultural norms, which vary from State to State, pose a boundary between a gift and a bribe or favoritism: therefore the definition of what is corrupt or not is a cultural thing (Rose-Ackerman 1999) and what can be defined by a corrupt external observer can be considered as an acceptable gift within a country.

However, the economic analysis of the causes and the extent of political corruption cannot and must not be limited to cultural factors. In fact, in addition to these, one needs to consider structural and institutional (economic and political) aspects that may override cultural interpretations (the latter is sometimes used as a way of justifying corrupt practices in some countries).

It is worth noting that the framework or "structure" of the administration and how well the institutions of a country work can influence the spread of corruption. For example, in Britain where the administrative organization is separate from the political system and the political parties were institutionalized at an early stage, there is notably less local political corruption. If there is a clear boundary between the two different systems (social and political procedures from political and economic ones), it is more difficult for politicians to enter into the bureaucratic environment.

In some countries, the presence of so-called regional brokers or different forms of mafia has meant that the social order of these States has been weakened. Due to the particularistic and fragile financial and administrative network in these countries, the State has had to depend on these "brokers" to function locally.

It is worth mentioning that each State develops in different ways, so it is perhaps incorrect to conclude that the infiltration of these groups

between the bureaucratic authorities and political system is the main cause of political corruption. That said, widespread clientelism may indeed be largely to blame for the corruption problem. Della Porta (1997) mentions how the links between clientelism and corruption and weak administration structures and corrupt practices have become "vicious circles" from which it is incredibly difficult to break free.

Another cause of political corruption identified in literature is government intervention in the economy. This intervention creates income managed on a discretionary basis by public officials, which in turn can generate corruption. Due to the growth of the welfare state and public sector, the political elite now have the monopoly over many financial resources.

It is easier for members of State and private individuals to sidestep rules and regulations as decisions are not influenced by the workings of the market but linked to these rules.

While the focus so far has mainly been on the reasons and opportunities for one to enter into corrupt practices, it is also important to look at "institutional" and "political" factors.

One significant factor is in what political context the corruption takes place. There is a widespread belief that corruption is measured by the level of democracy in a country. The more democratic a State, the less corrupt it should be. According to Friedrich (1989) the more lawful the country, the less likely corruption will evolve.

However, in some authoritarian countries, dictators are able to control the extent of corruption. In other words he will be able to decide who benefits from the State resources. For this reason, in some countries under powerful dictatorships, corruption is actually less rife. The citizens will consider these States a legitimate entity as they have the power and ability to carry out social changes and promote economic production as well as to manage law and order.

Regarding the functioning of political parties, we must consider the different methods of financing of political parties and the funding of elections. In countries such as the United States and Japan, this is now becoming a key matter with regard to political corruption.

Due to the high costs of elections and the political process in general, parties often have to look elsewhere for financial resources. They have few funds available and therefore attempt to find other means to keep their political status and to appear the most efficient in the eyes of the public.

Another important factor which may lead to political corruption is "longevity in power."

This term refers to how governments can stay in power without being challenged by other rivals. In authoritarian States, the absence of a formal autonomous structure means that bureaucrats or public officials can take advantage and use power in corrupt ways. On the contrary, in democratic countries public accountability is guaranteed by the institutions of the "State of law." If governments believe that their position of power is infallible, many begin to confuse the difference between State and government and as a result begin to think they are exempt from the laws that apply to everyone else.

Among the causes of greater or lesser spread of corruption are certainly the control system and the extent of the punishment related to corrupt acts. In deciding whether to engage in a corrupt transaction, an individual (be they public officials or firms), by comparing the expected benefits with the expected cost, decides to be corrupt only if the total benefit is at least equal to the total cost. The reforms designed to increase the risk of discovery and the extent of the punishment may reduce the demand for and supply of bribes. With regard to the likelihood of being discovered, it is often endogenous compared to the level of corruption, in that "the more widespread the corruption, [. . .] the lower the risk of being denounced" (della Porta and Vannucci 2016).

Consequences of Political Corruption

While most analysts concur that political corruption is morally condemnable, its economic and political consequences are less discussed. This is perhaps due to the fact that political corruption – complex as a concept – is difficult to both define and measure.

Corruption, in the past, has even been considered as having a positive role in some developing

countries. This was the opinion of analysts who were working within the scheme of the "modernization theory" in the 1960s. More precisely, the debate on the economic effects of corruption started with the two pioneering works of Leff (1964) and Huntington (1968), who said that corruption would stimulate economic growth mainly through the operation of two mechanisms:

- Corruption if "speedy money" can ensure that individuals are able to circumvent red tape.
- If the bureaucrats are paid directly for their work through bribery, this should make the bureaucrats work better and faster.

Therefore corruption is seen as something constructive in promoting economic growth and investment as well as helping social development in countries with weak bureaucracy, making up for the lack of adequate formal practices. However, these statements have been widely opposed by recent literature that says that corruption can have positive effects only in limited areas: they refer to the benefits arising from the assumption that State intervention creates inefficiency and corruption, shrinking them or eliminating them for the best economic growth of the system. Therefore, only in cases where the political system is inefficient, may political corruption be an improvement (see della Porta and Vannucci 1997).

With the spread of corruption, government costs will increase, public services will no longer be guaranteed as resources will begin to dwindle, revenues may be exposed, and it will no longer be possible to make effective decisions. Corruption can also have a negative influence on the purchase or even production of goods for the public. The whole process of purchasing, tendering, giving contracts, the realization of work, and payment at the end may all be affected.

A key moment in the political process of a democracy is when one needs to respond to public demand. The administration network is faced with the demands and requirements of different social groups. The function of political mediation according to Pizzorno (1992) is "to identify and interpret the needs and desires of the population; select and generalize those which can be expressed in political terms; propose, justify and criticize policies and measures to achieve these ends or, when necessary, to explain why they cannot be satisfied." It is the means through which interests are collected and expressed.

The efficiency of this service is then jeopardized by corruption as the demand becomes "internal" not "public." Public demand is no longer protected as the formal restrictions of the institutions are weakened by corrupters, who are able to make decisions and take the profits from corrupt practices creating the disintegration of public demand to satisfy specific needs.

Public spending is often inextricably linked with corruption and clientelism. Public officials aim to draw as many resources as possible into their domains of power so they can receive payment for mediation as a bribe. In this way, these administrators also try to win public support as a consequence of the influence of public investment on employment. For this reason, public spending is moved into the areas where there are the most benefits from corruption and where there are fewer risks due to the discretional characteristics of the practices. Citizen needs are not taken into account when these operations and services are carried out and many works remain incomplete or discarded altogether. In addition, in the case where politicians are not content with the sum given through bribes, public demand often remains unfulfilled. It is worth noting, however, that during purchasing transactions, there are competitive rules which mean that those participating have equal treatment once the cost of public spending has been decided.

In fact, in a widespread corrupt society, it can therefore be expected that those who formulate public demand will award the provision of public goods or a "carte" of all the companies participating in tendering. The bids are then able to be correlated and the income obtained then redistributed. Politicians or bureaucrats who have been involved in corruption in the past will then make sure that these individual companies or group of companies will definitely (or most likely) win a tender.

As a consequence, public demand will then be administered in such a way that these "protected"

companies will have an advantage in promoting their own products and will be able to benefit in the future from their "special" connections with politicians. Costs become inflated as public tenders are allocated to companies who lack the required competences, equipment, and qualifications.

Another element which politicians use to manipulate public demand is when they have real (or false) "urgent" interests. Usually the amount of public spending and its administration will be relatively high so corrupt politicians need to create a contrived emergency.

Lack of planning expertise of public administration and convoluted bureaucracy slows down the time taken to complete public operations, which in turn paves the way for corruption as discretionary proceedings in the contracting process become legitimate.

Corruption therefore creates a redistributive process from which politicians benefit: "A corrupt system of government services has the distributional disadvantage of benefiting unscrupulous people at the expense of law-abiding citizens who would be willing to purchase the services legally" (Rose-Ackerman 1978). Corruption attributes property rights to agents who then breach the rights of public interest and consequently further the tax burden which has a detrimental impact on the life of citizens.

Cross-References

- ▶ Administrative Corruption
- ▶ Corruption

References

Argandoña A (2003) Private-to-private corruption. J Bus Ethics 47:253–267

Austen-Smith D, Wright JR (1994) Counteractive lobbying. Am J Polit Sci 38:25–44

Campos N, Giovannoni F (2007) Lobbying, corruption and political influence. Public Choice 131:1–21

Damania R, Fredricksson PG, Mani M (2004) The persistence of corruption and regulatory compliance failures: theory and evidence. Public Choice 121:363–390

della Porta D (1997) The vicious circles of corruption in Italy. In: della Porta D, Mény Y (eds) Democracy and corruption. Printer, London, pp 35–49

della Porta D, Pizzorno A (1996) The business politicians: reflections from a study of political corruption. J Law Soc 23:73–94

della Porta D, Vannucci A (1997) The "perverse effects" of political corruption. Political Stud XLV:516–538

della Porta D, Vannucci A (2016) The hidden order of corruption: An institutional approach. Routledge, London.

Friedrich CJ (1989) Corruption concepts in historical perspective. In: Heidenheimer AJ, Johnston M, LeVine VT (eds) Political corruption. A handbook. Transaction Publishers, New Brunswick. (third printing 1993)

Grossman G, Helpman E (1999) Competing for endorsements. Am Econ Rev 89:501–524

Grossman G, Helpman E (2001) Special interest politics. MIT Press, Cambridge, MA

Harstad B, Svensson J (2011) Bribes, lobbying and development. Am Polit Sci Rev 105:46–63

Huntington SP (1968) Political order in changing societies. Yale University Press, New Haven

Jain A (2001) The political economy of corruption. Routledge, London

Leff NH (1964) Economic development through bureaucratic corruption. Am Behav Sci 8:8–14

Nye JS (1967) Corruption and political development: a cost-benefit analysis. Am Polit Sci Rev 61:417–427

Olken BA, Pande R (2012) Corruption in developing countries. Annu Rev Econ 4(1):479–509

Peters JG, Welch S (1978) Political corruption in America: a search for definitions and a theory, or if political corruption is in the mainstream of American politics why is it not in the mainstream of American politics research? Am Polit Sci Rev 72:974–984

Pizzorno A (1992) La corruzione nel sistema politico. Introduction to Della Porta D, Lo scambio occulto. Il Mulino, Bologna

Rose-Ackerman S (1978) Corruption. A study in political economy. Academic, New York

Rose-Ackerman S (1996) Democracy and 'grand' corruption. Int Soc Sci J 48(149):365–380

Rose-Ackerman S (1999) Corruption and government. Cambridge University Press, New York

Political Economy

Karsten Mause
Department of Political Science (IfPol), University of Münster, Münster, Germany

Definition

The term "political economy" (PE) is mainly used in two related contexts. First, it is used to denote a multidisciplinary research field in which political

scientists, economists, legal scholars, and other social scientists investigate the relationship between the political sphere (most notably "the state") and the economic system of different societies on Earth at different points in time. Second, social scientists, journalists, and other observers sometimes use the term PE to refer to the observable interaction of politics and business in real-world societies. Focusing on the first context, this entry gives an overview of the research field of political economy (PE) and discusses its relationship to law and economics as a research program.

Positive Political Economy: Analyzing What Is

Within the toolkit of PE, there are basically two different approaches which are currently used to analyze the relationship between politics and the economy: positive PE (explained in this section) and normative PE (see next section). Using a positive approach, researchers conduct empirical analyses in the form that they describe and explain the relationship between the political and economic sphere (i.e., what is) – but they usually do not make value judgments in the form of normative statements as to what public and private sector actors *should* do or not do: for instance, positive politico-economic analyses usually do not contain policy recommendations regarding the "right" way for the government to intervene in certain sectors or markets in the economy (i.e., what ought to be). In other words, scholars doing positive PE research first of all *describe* as precisely as possible the extent to which the state or other political actors intervene in the economic system of a society (or different societies, if a comparative perspective is taken) in a certain investigation period.

Political interventions in different sectors of the economy or particular markets may take various forms, such as government subsidies, taxation, state-owned enterprises, or regulations (Den Hertog 2000; Boettke and Leeson 2015), and may be done for various reasons, such as eliminating market failures/allocative inefficiencies, redistributing resources from the rich to the poor, or stimulating economic growth and employment (see below). As it is often the case that firms, business associations, trade unions, and other actors within the economic system try to influence the process of economic policymaking via lobbying and other forms of leverage, in many contexts we can observe a mutual interference of the political and the economic sphere. This may also include a possible correlation between (a) the economic situation in a society and (b) the popularity and election results of government, opposition parties, or individual politicians (Lewis-Beck and Stegmaier 2013).

Moreover, it has to be taken into account that external factors such as global issues (poverty, climate change, war refugees, etc.), international organizations (e.g., World Trade Organization, European Union, International Monetary Fund, World Bank), developments in international markets, or the activities of foreign governments (e.g., tariff policy, international tax competition, sovereign debt, sovereign defaults) may influence a (sub)national PE understood as the interaction of the political and economic system in a real-world society. The fact that nation states are these days embedded, in various respects, in an international system is analyzed in the literature on "international PE" and "global PE" (see, e.g., Ravenhill 2016). In this context it should also be mentioned that there is a strand of PE research which focuses explicitly on the differences between the national economic systems of the countries in the world including different "Varieties of Capitalism" (Hall and Soskice 2001) as well as the remaining more or less socialist "command economies" or "centrally planned economies" such as Cuba and North Korea (Fine and Saad-Filho 2012, chap. 7).

However, many politico-economic analyses within positive PE do not content themselves with describing the relationship between politics and the economy but, moreover, try to *explain* the observations made in the descriptive phase of research. Which factors can explain why the state intervenes in a particular way in a society's economy? Which explanatory factors may have driven the transformation of the "interventionist state" over time? Why do some countries show a better macroeconomic performance (economic growth,

employment, price stability, etc.) than other countries? In this spirit, for example, numerous politico-economic studies have empirically analyzed whether factors such as government ideology, powerful interest groups, fiscal pressure, socioeconomic problems (e.g., de-industrialization, unemployment, economic slump), path dependence, or globalization help explain the observable differences across the member states of the European Union (EU) or the Organization for Economic Co-operation and Development (OECD) with respect to the use of policy instruments such as public entrepreneurship, regulation, taxation, or subsidization in the decades after World War II (e.g., Leibfried et al. 2015; Obinger et al. 2016).

Similar studies exist for the less-developed world and/or for country groups including countries with "not-so-democratic" political systems. There is, for example, a politico-economic literature that describes and explains different aspects within the relationship between politics and the economy in autocratic regimes (e.g., Acemoglu and Robinson 2012; Leibfried et al. 2015, parts IV/V). Moreover, there are many studies entitled "The Political Economy of XY" which means that the particular study analyzes the interplay between political and economic factors in the specific context under investigation, for example, the political economy of migration, foreign aid, higher education, terrorism, and so on.

Normative Political Economy: Recommending What Ought to Be

PE research may be done not only in the form of empirical or "positive" analyses (as defined above) but also in the form of a *normative* analysis. This means that a specific area in the economic system is analyzed in order to come to conclusions as to what "the state" should (not) do in the area under investigation. Should the government intervene in a particular sector of the economy or a particular market by means of regulations or other policy tools? Should public bureaucrats be allowed to control certain activities of private sector firms and households? Should regulatory agencies be mandated to supervise competition in particular sectors and markets? Such questions are addressed in the fundamental and ongoing PE debate over the proper role of the state in the economy (see Boettke and Leeson 2015, for a survey). Contributions to this debate are based, more or less explicitly, on the following major schools of thought.

Varieties of Economic Liberalism

Political economists in the tradition of Adam Smith (1723–1790), whose seminal book on "the Wealth of Nations" (Smith 1776/1981) is the "bible" for those advocating "economic liberalism" and "market liberalism," basically argue that the state should leave the economy alone. It is assumed that there is some kind of natural tendency to equilibria in markets. That is, if there is an excess demand or excess supply, then such disequilibrium will only persist for a short period of time. According to the economic laws of demand and supply, markets will find a "market-clearing price" at which demand equals supply. In other words, direct governmental interventions into markets are perceived to be unnecessary (or even harmful) as specific markets and the economy as a whole possess "self-healing powers" in the form of the "market forces": that is, the interplay of demand and supply coordinated via the price mechanism.

However, it should be mentioned that Smith (1776/1981) and other advocates of economic/market liberalism such as Friedrich August von Hayek (1899–1992) and Milton Friedman (1912–2006) acknowledge that society may not be left to markets alone – but that the state has to perform at least some tasks to make markets and society work. For example, political economists in the tradition of Smith (1776/1981), Hayek (1960), and Friedman (1962) consider it to be a government task to ensure that there is a functioning legal system (rule of law, laws, courts, judges, etc.) that can be used, among other things, for enforcing (i) property rights and (ii) the contracts signed by market participants. By contrast, *libertarian* political economists, who consider the possibility of a stateless society, go a step further: they argue that private governance mechanisms

(reputation, nongovernmental courts, etc.) are sufficient to enforce property rights and contractual agreements (see, e.g., Friedman 2014; Leeson 2014; Stringham 2015).

Furthermore, there are two specific variants of economic liberalism which were developed some decades ago but are still influential in the current politico-economic discourse: ordoliberalism and constitutional political economy. Ordoliberals in the tradition of the German economist Walter Eucken (1891–1950) criticize that Smith (1776/1981) and other advocates of classical economic liberalism and its laissez-faire approach have neglected that a market economy does not automatically increase the wealth of a nation. For example, individual markets or whole sectors of the economy may suffer from anticompetitive practices by private and/or public companies (market-entry barriers, cartelization, price collusion, competition-distorting state aid, government monopoly, and so on). Consequently, for ordoliberals it is essential that the state creates and enforces a legal order and institutions (e.g., a politically independent competition authority) that try to prevent private and governmental restraints of competition and market forces as far as possible (Eucken 1952/2004; Vanberg 2015).

In a similar vein, constitutional political economists in the tradition of James M. Buchanan (1919–2013) emphasize that markets, competition, and the economy as a whole need "rules of the game" – a "constitution" – which channel the individual self-interests of consumers, firms, and other actors (for more details, see Buchanan 1987; Vanberg 2005). Moreover, constitutional PE points out that creating and enforcing such "rules of the game" is far from trivial. Reading Eucken (1952/2004) one gets the impression that he takes it for granted that there is a benevolent government which realizes that the rules recommended by ordoliberals are beneficial for society and implements these rules. In contrast, constitutional PE assumes that not only firms and consumers but also politicians and public bureaucrats are self-interested actors. Under these conditions, not only powerful firms and interest groups but also politicians and public bureaucrats may impede the implementation of rules which could be beneficial for citizen-consumers and society as a whole (e.g., the abolition of government monopolies, the abolition of special privileges for state-owned enterprises, better regulations for public utilities, and so on). However, what ordoliberalism and constitutional PE have in common is that both prefer a rule-based economic policy over discretionary government interventions in the economy and market processes.

For the sake of completeness, it should be mentioned that constitutional PE is part of the broader research program entitled "economic theories of politics" or "public choice theory" established by Downs (1957) and others (see Mueller 2003, for a survey). Public choice theory breaks with the welfare-economic assumption of benevolent governments working in the public interest. Instead, it is assumed that politicians and public bureaucrats (i) are primarily interested in maximizing their individual utility and (ii) "act solely in order to attain the income, prestige, and power which come from being in office" (Downs 1957, p. 28). In other words, it is theoretically assumed that political decision-makers are self-interested not only when they make private choices (as consumers, investors, landlords, and so on) but also when they make public choices in government, parliamentary committees, and other political contexts. As "older" schools of PE (e.g., classical economic liberalism in the tradition of Smith) did not pay much attention to the motivations of public sector actors and implicitly assumed that the government is primarily interested in maximizing the wealth of a nation, in the politico-economic literature, public choice theory is often denoted as the "New Political Economy" (Frey 1999).

The next step in this area has been taken by scholars working in the field of "behavioral political economy." Therein, it is taken into account that real-world actors often do not behave as rationally and with the self-interest that economists' traditional homo-economicus model predicts; this may lead to other policy implications regarding the "optimal" design of the incentive structures under which certain types of consumers, investors, policymakers, and other individuals make their more or less informed and more or less

selfish decisions (see Schnellenbach and Schubert 2015, for a survey).

Keynesianism and Other State-Interventionist Approaches

If the economy slips into recession, then hard-core economic liberals may argue that such an economic crisis may have painful consequences for firms and individuals (a drop in orders, bankruptcy, unemployment, poverty, and so on) but does not require government intervention – because thanks to its "self-healing powers" the economy will recover on its own after some time. By contrast, political economists in the tradition of John Maynard Keynes (1883–1946), whose seminal book "The General Theory of Employment, Interest and Money" (Keynes 1936) belongs to the most influential critiques of the laissez-faire approach of economic liberalism, consider it to be a government responsibility to stimulate the economy in times of economic slump (i.e., increasing government spending, reducing taxes, and so on). If the government lacks the necessary financial resources to implement an economic stimulus package, then Keynesians recommend (a) government borrowing (so-called deficit spending) and (b) repaying the debts after the crisis when government tax revenues increase due to economic growth and rising employment.

Critics of deficit-spending object that step (b) is often not conducted by government, which is one reason for the high levels of public debt observable in many countries these days. Keynesians usually respond to such criticism by arguing that costly state interventions to "stimulate," "stabilize," and "steer" the economy are necessary and legitimate as long as there is unemployment in a society (Krugman 2013). Complementary to fiscal stimulus packages, Keynesians propose measures of monetary policy to stimulate the economy (e.g., lower interest rates). If there is a politically independent central bank, then monetary policy is not a tool of government (i.e., politicians have no access to the tools of monetary policy).

While Keynesianism offers a macroeconomic justification for state intervention in the economy, the so-called market failure theory (for a survey, see Stiglitz and Rosengard 2015) has demonstrated that the laissez-faire approach of economic liberalism ignores the fact that different types of market failures offer a potential justification for government action. For example, the behavior of certain firms and consumers (e.g., environmental pollution by coal-fired power plants) may create *negative externalities* for other society members. The government may implement measures (law, regulations, etc.) that force polluters to reduce or even stop producing negative externalities. Moreover, it can be expected that many society members will not pay for certain goods and services if they can consume these goods and services free of charge. However, if free riding is possible, then private actors have a low or no incentive to supply such goods and services. To secure the provision of *public goods* in the sense that no one in society can be excluded from consuming such goods, the government may step in: for example, the public good argument offers an economic argument to justify the national defense being provided by the government and financed by taxes (i.e., society members, as potential free riders, are forced to pay for national defense).

Informational asymmetries constitute another type of potential market failure. If suppliers are better informed about certain characteristics of products and services (e.g., the quality of used cars) than potential buyers, then the markets for these products and services may not function well: because it can be expected that many consumers under these circumstances would hesitate to enter a market transaction as they would fear being exploited by the better-informed sellers (e.g., low-quality, high-price products). It is also possible that buyers are the better-informed market party. Imagine, for example, insurance companies that do not know the true health status of people seeking to buy a health insurance. In situations with informational asymmetries, the government may implement measures (governmental provision of quality information, disclosure laws, governmental regulation of product quality, etc.) to mitigate these informational problems and facilitate market transactions. Moreover, the market-failure framework considers *natural monopolies* to be a potential justification for government

action. Such monopolies occur if for efficiency reasons in certain sectors or markets of the economy only one firm is doing business (e.g., the provider of a rail network, a power supply line or a water line). To avoid allowing this provider to exploit its *monopoly power* (high prices, bad quality, and so on), the government may regulate this natural monopoly (price regulation, quality regulation, etc.). And, as mentioned above in the context of ordoliberalism, the government may also intervene in some way to tackle the problem that markets and competition do not work properly due to "ordinary" monopolies and other anticompetitive practices.

It should be emphasized that the existence of a market failure does not automatically imply that the government has to solve the problem. For instance, there may be private third parties (e.g., private certification agencies) and market-based mechanisms (e.g., reputation, brand-name capital) that help market participants to overcome their informational problems, so that buyers and sellers are able to enter into mutually beneficial market transactions. In other words, in the politico-economic literature, it is not only discussed (a) whether a certain market or sector of the economy really suffers from "market failures" and "allocative inefficiencies" but also (b) which governmental or private governance mechanisms (or a mixture of both) seem to be the most suitable to solve the problem at hand (Ostrom 2010). Moreover, political economists stress that all of these mechanisms are imperfect solutions that work more or less well depending on the specific real-world context in which they are used (Wolf 1993). And it may be the case that government action to solve a market-failure problem may create new problems (for an overview of the politico-economic debate on "government failure," see Keech and Munger 2015).

A normative yardstick that is often used by economic liberals to assess whether state activity is necessary to mitigate a certain type of market failure is the so-called subsidiarity principle. According to this principle, government action is only necessary if private market solutions and private governance mechanisms fail. A brief and oft-cited summary of this principle can be found in the book "Principles of Economic Policy" by Eucken (1952/2004, p. 348): "The structure of society should follow a bottom-up approach. What the individuals or the groups can autonomously accomplish should be done on their own initiative and to the best of their abilities. And the state should only intervene in those cases in which its assistance is indispensable" (own translation, K.M.). By contrast, political economists that have a less individualistic and more state-centered view of economy and society may start from the paternalistic, state-interventionist assumption that the state is automatically responsible for solving market-failure problems (for a survey of the politico-economic literature on paternalistic government, see Le Grand and New 2015). In democratic societies, the ultimate decision-maker in this context is the government in power – and this decision-maker is certainly free to ignore the normative (and often conflicting) policy recommendations made by political economists and other experts.

Marxism and the Social Question

Economic liberalism and its belief in markets and competition have always been the target of criticism. Karl Marx (1818–1883) and Friedrich Engels (1820–1895) have argued that it is a basic feature of capitalist market economies that the "working class" (the so-called proletariat) is exploited by business firms and their owners (the "capitalists"; see, e.g., Marx and Engels 1848/2002). The state is seen as an agent of the so-called bourgeoisie (including the capitalists) which constitutes the ruling class in society. Marx and Engels predicted that capitalism will be overthrown through a "proletarian revolution" that leads to socialism and, eventually, to communism (including a classless society). It is beyond the scope of this paper to critically review everything that has been written by Marx, Engels, and their followers under the label "Marxian Political Economy" about imagined and real existing types of capitalism, socialism, and communism (for a survey, see Fine and Saad-Filho 2012). Nor do we discuss the many problems of "command economies" or "centrally planned economies." However, Marx and his followers have repeatedly

pointed out a serious problem which many capitalist market economies still have to cope with: it may be the case that an economy consists of a system of well-functioning, efficient markets, but this system produces social problems.

For example, in many countries we observe income and wealth inequality among society members. While hard-core economic liberals may argue that such inequalities have to be accepted and simply reflect individual differences in performance and success on markets, other political economists argue that the state in the name of "social justice" has to tackle distributional problems via redistribution (see Piketty 2014, for an overview of this debate). And even economic liberals that are skeptical of state interventions, such as Hayek (1960) and Friedman (1962), take it for granted that those society members who, for whatever reason (e.g., disease, disability), are not able to earn money in the labor market should receive publicly financed welfare benefit payments ensuring a minimum income needed to exist. Whatever political economists from different schools may think about social problems and their solution – in the end, however, in democratic societies the scope and structure of the welfare state are determined in the political process.

Political Economy Meets Law and Economics

Last but not the least, we have to address the question of what law and economics (LE) as a research program has to do with PE as a multidisciplinary endeavor. First of all, we can observe that terms, concepts, and tools from the toolkit of PE (market failure, externalities, public goods, efficiency, utility, welfare, constitutional PE, behavioral PE, and so on) are used by LE scholars and in LE textbooks as well (see, e.g., Towfigh and Petersen 2015). In this context, it should also be mentioned that economists belonging to the LE movement have made important contributions to the research field of PE as sketched above. See, for example, the studies on externalities and public goods by Ronald H. Coase or the contributions by George J. Stigler, Richard A. Posner, and Samuel Peltzman to the economic theory of regulation (see the bibliographies in Den Hertog 2000; Boettke and Leeson 2015). In other words, many of the concepts presented above under the label PE, which is mainly used by political scientists, economists, and "political economists," are presented in LE publications under the label LE, which is mainly used by legal scholars, economists, and supporters of the LE movement. And while some may classify Coase, Stigler, Posner, and Peltzman as economists or LE scholars, others may classify them as political economists.

Likewise, Persson and Tabellini (2003), La Porta et al. (2008), and similar studies investigating the interplay of legal institutions and the economy are oft-cited in the PE as well as in the LE literature. In any case, it should be clear now that political economists and LE scholars who are interested in analyzing different aspects of the interplay between the political and economic sphere of society share a common terminology. This offers opportunities for research cooperation and interdisciplinary research – but does not mean that the disciplines participating in the "joint ventures" labeled PE and LE would have lost their idiosyncrasies and specific strengths. For example, as noted above, in the politico-economic works of Hayek, Friedman, Eucken, and Buchanan it is argued that "the state" should provide a legal framework which ensures that markets and competition work well; however, these thinkers do not say much about the fundamental issue of how exactly the specific legal framework for a specific market or economic sector in a particular real-world society should be designed and enforced (contract law, competition law, capital market law, energy law, environmental law, and so on). That is, legal experts are necessary to bridge the gap between normative politico-economic theories of the proper role of the state and practical public policy.

Moreover, it can be observed that in the positive, empirical branch of PE and LE, there seems to be a methodological convergence in the sense that scholars doing empirical research in this area basically use the same toolkit, consisting of various quantitative and qualitative methods (statistical techniques, document analysis, field

research, etc.; Towfigh and Petersen 2015). As we have seen above, however, such consensus cannot be observed in the normative branch of PE. Looking through the theoretical – some would say "ideological" – lenses of different schools of PE in many cases brings us to different conclusions regarding the question of what the state should do (or not do) in the particular area of the economy under investigation.

Cross-References

- ▶ Capitalism
- ▶ Constitutional Political Economy
- ▶ Government Failure
- ▶ Market Failure: Analysis
- ▶ Market Failure: History
- ▶ Ordoliberalism
- ▶ Public Choice: The Virginia School

References

Acemoglu D, Robinson J (2012) Why nations fail: the origins of power, prosperity, and poverty. Crown Business, New York

Boettke PJ, Leeson PT (2015) The economic role of the state. Edward Elgar, Cheltenham

Buchanan JM (1987) The constitution of economic policy. Am Econ Rev 77:243–250

Den Hertog JA (2000) General theories of regulation. In: Bouckaert B, De Geest G (eds) Encyclopedia of law and economics, The regulation of contracts, vol III. Edward Elgar, Cheltenham, pp 223–270

Downs A (1957) An economic theory of democracy. Harper & Row, New York

Eucken W (1952/2004) Grundsätze der Wirtschaftspolitik, 7th edn. Mohr Siebeck, Tübingen

Fine B, Saad-Filho A (eds) (2012) The Elgar companion to Marxist economics. Edward Elgar, Cheltenham

Frey BS (1999) Economics as a science of human behaviour: towards a new social science paradigm, 2nd edn. Springer, Dordrecht

Friedman M (1962) Capitalism and freedom. University of Chicago Press, Chicago

Friedman DD (2014) The machinery of freedom: guide to a radical capitalism, 3rd edn. Open Court, Chicago

Hall PA, Soskice D (eds) (2001) Varieties of capitalism: the institutional foundations of comparative advantage. Oxford University Press, Oxford

Hayek FA (1960) The constitution of liberty. University of Chicago Press, Chicago

Keech WR, Munger MC (2015) The anatomy of government failure. Public Choice 164:1–42

Keynes JM (1936) The general theory of employment, interest and money. Macmillan, London

Krugman P (2013) End this depression now! W.W. Norton & Company, New York

La Porta R, Lopez-de-Silanes F, Shleifer A (2008) The economic consequences of legal origins. J Econ Lit 46:285–332

Le Grand J, New B (2015) Government paternalism: nanny state or helpful friend? Princeton University Press, Princeton

Leeson PT (2014) Pirates, prisoners, and preliterates: anarchic context and the private enforcement of law. Eur J Law Econ 37:365–379

Leibfried S, Huber E, Lange M, Levy JD, Nullmeier F, Stephens JD (eds) (2015) The Oxford handbook of transformations of the state. Oxford University Press, Oxford

Lewis-Beck MS, Stegmaier M (2013) The VP-function revisited: a survey of the literature on vote and popularity functions after over 40 years. Public Choice 157:367–385

Marx K, Engels F (1848/2002) The communist manifesto. Penguin Books, London

Mueller DC (2003) Public choice III. Cambridge University Press, Cambridge

Obinger H, Schmitt C, Traub S (2016) The political economy of privatization in rich democracies. Oxford University Press, Oxford

Ostrom E (2010) Beyond markets and states: polycentric governance of complex economic systems. Am Econ Rev 100:641–672

Persson T, Tabellini G (2003) The economic effects of constitutions. MIT Press, Cambridge

Piketty T (2014) Capital in the twenty-first century. Harvard University Press, Cambridge

Ravenhill J (ed) (2016) Global political economy, 5th edn. Oxford University Press, Oxford

Schnellenbach J, Schubert C (2015) Behavioral political economy: a survey. Eur J Polit Econ 40:395–417

Smith A (1776/1981) An inquiry into the nature and causes of the wealth of nations. Liberty Fund, Indianapolis

Stiglitz JE, Rosengard JK (2015) Economics of the public sector, 4th edn. W.W. Norton & Company, New York

Stringham EP (2015) Private governance: creating order in economic and social life. Oxford University Press, Oxford

Towfigh EV, Petersen N (eds) (2015) Economic methods for lawyers. Edward Elgar, Cheltenham

Vanberg VJ (2005) Market and state: the perspective of constitutional political economy. J Inst Econ 1: 23–49

Vanberg VJ (2015) Ordoliberalism, Ordnungspolitik, and the reason of rules. Eur Rev Int Stud 2:27–36

Wolf C (1993) Markets or governments: choosing between imperfect alternatives, 2nd edn. MIT Press, Cambridge, MA

Political Violence

► Terrorism

Politicians

Benny Geys[1,2] and Karsten Mause[3]
[1]Department of Economics, Norwegian Business School BI, Oslo, Norway
[2]Vrije Universiteit Brussel, Ixelles, Belgium
[3]Department of Political Science (IfPol), University of Münster, Münster, Germany

Definition

Politicians are persons involved in the process of public policymaking in their role as members of governments, parliaments, political parties, and other political bodies at the (sub)national level (e.g., local government, state legislature, national parliament, etc.) as well as within the supranational political arena (e.g., United Nations Security Council, European Union institutions, and so on). Many politicians get into office through a democratic election, while others are selected or appointed to a public office. For many politicians, politics is a full-time job, while for others, it remains an activity in addition to a main job as a lawyer, teacher, or entrepreneur. This entry takes stock of what law and economics (LE) scholars have contributed to the large social science literature on politicians.

The Law and Economics of Politicians

Analyzing what politicians do belongs to the "core business" of political scientists. However, the law and economics discipline has extensively contributed to this multidisciplinary research field as well. Given the large number of articles and books, it is beyond the scope of this entry to review everything that LE scholars have written about "politicians." Rather, we focus on the following aspects: (1) the motivations of politicians, (2) politicians as lawmakers, (3) politicians and interest groups, and (4) the design and effects of the (legal) rules under which politicians act.

What Do Politicians Maximize?

LE scholars including, for instance, George J. Stigler and Richard A. Posner are also active contributors to the field of "public choice." Public choice theory in the tradition of Downs (1957) breaks with the welfare-economic assumption of benevolent governments working in the public interest and maximizing the wealth of the nation. Instead, it assumes that politicians and public bureaucrats (i) are primarily interested in maximizing their individual utility and (ii) "act solely in order to attain the income, prestige, and power which come from being in office" (Downs 1957, p. 28). In other words, it is theoretically assumed that political decision-makers are self-interested not only when they make private choices (as consumers, investors, landlords, and so on) but also when they make choices in government, parliamentary committees, and other political contexts (Mueller 2003; Posner 2014, chap. 20).

Whether real-world politicians really behave in line with predictions derived from the *homo-oeconomicus* assumption used in economic theories is, of course, an empirical question – and presumably depends on the specific context within which the individual politician acts. In democracies, an individual politician's private utility maximization is usually constrained by a number of mechanisms that may channel his/her self-interest and act as a disciplining device, e.g., his/her political party, political competition, press freedom, an independent judiciary, nongovernmental watchdog organizations, a politically interested electorate, and so on. However, in the absence of such constraints, a public choice perspective would lead to the expectation that politicians could have a low or even no incentive to pursue the public interest (Mueller 2003; Bovens et al. 2016).

Within the vast LE literature, there are a number of empirical studies indicating that many politicians "shirk" their duties in their last period

in office, that is, once citizen-voters are no longer able to punish them at the next election (see Geys and Mause 2016, for a survey of the "shirking" literature). Moreover, there is a large literature on "political corruption" demonstrating that politicians at times illegally abuse their power for their private gain (money, privileges, jobs, and so on). LE researchers have made numerous contributions to this literature, discussing in particular the question of whether changes in the incentive structures under which politicians act (e.g., more severe punishment, transparency laws, better pay, etc.) help to reduce the probability of corruptive practices occurring (Rose-Ackerman and Palifka 2016).

Politicians as Lawmakers

Otto von Bismarck (1815–1898), the first Chancellor of the German Empire lasting from 1871 to 1918, once said: "Laws are like sausages. It is better not to see them being made." Obviously many social scientists have ignored Bismarck's advice since there is a large literature analyzing different aspects of the role of politicians as "makers" of public policy and "producers" of constitutions, laws, regulations, and other outputs of the political process (for surveys, see McCubbins et al. 2007; Parisi 2011; Posner 2014, chap. 20). LE scholarship has contributed to this multidisciplinary research field especially by analyzing and rethinking the system of checks and balances in which policymakers in different political systems are embedded. This literature suggests that it makes a difference whether policymakers "produce" laws and regulations in a representative democracy, a parliamentary democracy with a strong second chamber, a presidential democracy, a direct democracy with a powerful electorate, and so on (Mueller 2003; Persson and Tabellini 2004). Moreover, supranational actors – such as the European Union (EU), the North Atlantic Treaty Organization (NATO), and the United Nations (UN) – have been found to influence (sub)national policymakers' decisions (Dreher and Lang 2016).

An interesting phenomenon in this context is that some political systems allow politicians to make laws and regulations with direct implications for themselves. For example, politicians active as lawyers occasionally become involved in making laws that favor the legal profession (Matter and Stutzer 2015). Likewise, the desire of political parties to establish and maintain electoral thresholds may be interpreted as an anticompetitive instrument used to erect a barrier to entry for new parties (Wohlgemuth 1999). A closely related issue is whether politicians should legally be allowed to set their own salaries and other parts of their compensation package – as is currently often the case (Mause 2014). As noted above, problems might occur if politicians' self-interest and the public interest collide. If citizens get the impression that politics is some kind of "self-service store" for the "political class," this might have an effect on voter turnout, election results, and trust in politicians and the political system as a whole.

Politicians and Interest Groups

Within any democratic society, interest groups (such as firms, trade unions, employers' associations, taxpayers' associations, churches, rabbit breeders associations, rifle associations, and so on) have a strong incentive to lobby for their interests and achieve favorable regulations, government subsidies, protection of monopolies, and other things. Imagine, for example, the case of a powerful association of business enterprises that would prefer to keep its market closed to competitors. Any market entry barriers benefit the incumbent firm(s) as they reduce the level of competition in that particular sector of the economy. This is likely to result in higher prices (and lower consumer surplus) compared to a situation with more competition, but in higher profits for the firms active in this market. Clearly, this business association would have a strong incentive to ensure – via various lobbying activities – that policymakers do not open the market. In the public choice and LE literatures, the phenomenon sketched above is denoted as "rent-seeking" (see Congleton and Hillman 2015, for a survey of the voluminous theoretical and empirical rent-seeking literature).

Politicians are a natural objective for lobbying activities, which can take the form of, for example, information campaigns, dinner invitations, or political (campaign) donations. While certain lobbying activities can be beneficial by improving the information available to policymakers, they can also become a serious societal problem if they influence legislators, legislation, and/or the allocation of public funds in favor of specific groups (Hodler and Raschky 2014). Such political favoritism indeed distorts spending allocations away from the normative principles that ideally drive them, which from a theoretical perspective "lowers aggregate social welfare, [and] creates inequality across social groups" (Bramoullé and Goyal 2016, p. 23). As a result, a vast literature has investigated the link between lobbying activities and political outcomes. For example, studies linking firms' political donations to congressional voting patterns and public procurement contracts (see, e.g., Gherghina and Volintiru 2017, including a literature review).

As noted above, it is not only business firms and other special interest groups that may try to influence policymakers to make decisions favorable to their particular firm or industry. In this context, a specific type of "rent-seeking" by politicians occurs when the latter use their influence and power to obtain and defend special privileges that "ordinary citizens" do not have (Mause 2014).

Getting Incentives Right: Regulating Politicians

Scholars working at the intersection of LE and constitutional political economy in the tradition of Buchanan (2008) and others tend to maintain a strong focus on the "rules of the game" for politicians. There is, for instance, a substantial literature analyzing how the design of the working conditions in the political system (e.g., wages, term limits, etc.) incentivize individuals to stand for election – and drive politicians to make policies in the public interest. For example, the fact that candidates for a public office have to disclose parts of their income and/or wealth to the public is argued to have an effect on the incentive to run for office (Djankov et al. 2010; van Aaken and Voigt 2011; Braendle 2016). This also touches on the issue of accountability regarding "shirking" politicians and how to deal appropriately with such behavior (Geys and Mause 2016). Given the potential (self-) selection effects this induces, should there be compulsory attendance for politicians at parliamentary debates, roll-call votes, or committee meetings (possibly with penalties for nonattendance)?

Likewise, there remains an intense debate as to whether existing regulations with respect to donations to politicians and political parties, politicians' sideline jobs, partisan patronage and favoritism, and political corruption are sufficient (see Rose-Ackerman and Palifka 2016, for an introduction to this research field). The rules of the political game may be respected by many politicians, but there will most likely always be at least some "black sheep" that break the rules and behave "opportunistically" in the sense of Williamson (1985, p. 47): "By opportunism I mean self-interest seeking with guile. [. . .] More generally, opportunism refers to the incomplete or distorted disclosure of information, especially to calculated efforts to mislead, distort, disguise, obfuscate, or otherwise confuse."

Cross-References

- ▶ Political Competition
- ▶ Political Corruption
- ▶ Public Choice: The Virginia School
- ▶ Rent Seeking

References

Bovens M, Goodin RE, Schillemans T (eds) (2016) The Oxford handbook of public accountability. Oxford University Press, Oxford

Braendle T (2016) Do institutions affect citizens' selection into politics? J Econ Surv 30:205–227

Bramoullé Y, Goyal S (2016) Favoritism. J Dev Econ 122:16–27

Buchanan JM (2008) Same players, different game: how better rules make better politics. Constit Polit Econ 19:171–179

Congleton RD, Hillman AL (eds) (2015) Companion to the political economy of rent seeking. Edward Elgar, Cheltenham
Djankov S, La Porta R, Lopez-de-Silanes F, Shleifer A (2010) Disclosure by politicians. Am Econ J App Econ 2:179–209
Downs A (1957) An economic theory of democracy. Harper & Row, New York
Dreher A, Lang VF (2016) The political economy of international organizations. CESifo Working paper no. 6077
Geys B, Mause K (2016) The limits of electoral control: evidence from last-term politicians. Legis Stud Q 41:873–898
Gherghina S, Volintiru C (2017) A new model of clientelism: political parties, public resources, and private contributors. Eur Polit Sci Rev. 9(1):115–137. https://doi.org/10.1017/S1755773915000326
Hodler R, Raschky PA (2014) Regional favoritism. Q J Econ 129:995–1033
Matter U, Stutzer A (2015) The role of lawyer-legislators in shaping the law: evidence from voting on tort reforms. J Law Econ 58:357–384
Mause K (2014) Self-serving legislators? An analysis of the salary-setting institutions of 27 EU parliaments. Constit Polit Econ 25:154–176
McCubbins MD, Noll RG, Weingast BR (2007) The political economy of law. In: Polinsky AM, Shavel S (eds) Handbook of law and economics, vol 2. North-Holland Publishing, Amsterdam, pp 1651–1738
Mueller DC (2003) Public choice III. Cambridge University Press, Cambridge
Parisi F (ed) (2011) Production of legal rules. Vol. 7 of encyclopedia of law and economics, 2nd edn. Cheltenham, Edward Elgar
Persson T, Tabellini G (2004) Constitutions and economic policy. J Econ Perspect 18:75–98
Posner RA (2014) Economic analysis of law, 9th edn. Wolters Kluwer Law & Business, New York
Rose-Ackerman S, Palifka BJ (2016) Corruption and government: causes, consequences, and reform, 2nd edn. Cambridge University Press, Cambridge
van Aaken A, Voigt S (2011) Do individual disclosure rules for parliamentarians improve government effectiveness? Econ Gov 12:301–324
Williamson OE (1985) The economic institutions of capitalism. Firms, markets, relational contracting. The Free Press, New York
Wohlgemuth M (1999) Entry barriers in politics, or: why politics, like natural monopoly, is not organised as an ongoing market-process. Rev Austrian Econ 12:175–200

Polycentric Law

► Customary Law

Porn

► Pornography

Pornography

Samuel Cameron
Division of Economics, University of Bradford, Bradford, UK

Abstract

This entry discusses the problems of definition of pornography and the problems of defining an optimal regulatory strategy.

Synonyms

Porn

Judge Richard A. Posner has repeatedly applied orthodox free market Chicagoan law and economic thinking to porn in several of his books (*Sex and Reason*, *Frontiers of Legal Theory*, *Economic Analysis of Law*). Heterodox economist David George (2001) briefly discusses porn as an instance of what he calls "preference pollution." In this scenario, free markets impose weakness of will upon us leading to the "wrong choice" of consumption basket. We end up less well-off than we might otherwise have been because we are weak and fall prey to market pressures to consume porn. As we are now over a decade further on in the Internet world, his point would presumably be made even more forcefully now.

Regulation of porn, by laws, requires a definition of the activity so that it can be actioned.

This is a long-running problem. Everyone thinks they know what the word "pornography" means, yet legal definitions have been difficult to agree upon. This problem increases as laws come

Samuel Cameron has retired.

into force relying on gradation into extreme or hard-core porn as illegal and lesser types as not. The original meaning was as a description of the life and activities of prostitutes. This derives from the Greek, pornographos (porne = prostitutes and graphein = to write). It acquired its present meaning in the nineteenth century and seems to have two key elements, one being the obscenity of depiction of sexual acts and the other, the exploitation of workers in the production of the material and/or the consumers of the material. There are thus two areas of legislation. The first is to deal with protection of consumers, in their own best interests, and the second to deal with the workers in the industry. The latter has been particularly important in the American economy and less developed in some other territories. A notable instance of this is to be found in the 1995 Child Protection and Obscenity Enforcement Act, passed in the USA following the 1986 report of the Meese Commission. This required porn filmmakers to keep detailed records proving that no underage performers were used.

Much legislation has tended to be under the rubric of "obscenity." Porn has frequently been treated in legal matters as material "with a tendency to deprave or offend" where the decision on how depraving or offensive the matter is being left to a judge and/or jury. American Supreme Court decisions have illustrated the difficulties of Justice Potter Stewart's claim that he would know pornography "when he saw it." Lurking in the Supreme Court decisions has been the notion of an "average Justice Brennan's so-called 'limp dick' test became entrenched as a dividing line in movie censorship person" standard, that is, the judges are agents representing the median preferences of the population.

In terms of welfare economics, the simplest case for control is that porn may be viewed, as mentioned earlier, as a form of "social pollution." Controls can be aimed at the supply or at the demand. As with general crime, the expectation is that punishment will be proportional to harm. The particular dividing line in this respect is with respect to child pornography. The era of more liberal attitudes toward this in some areas of Europe (particularly in Denmark) has now passed.

Most of Europe makes child porn illegal but many parts of the world do not have explicit laws against it. Part of the more liberal attitude, in the past, related to deterrence and safety valve effects. If punishment does not deter then it will occur costs which bring no benefit other than the public satisfaction with retribution and the possible enhanced safety of taking the offenders out of circulation. The latter effect will be negligible for child (or any porn) if there is no complementarity between porn consumption and other offenses. If the offenses are a substitute good for porn consumption, then society can potentially have net benefit from liberal treatment of offenders. This was the core point of the criminologist Berl Kutchinsky about the Danish experience (see Kutchinsky (1985); Kutchinsky and Snare (1999)).

Various empirical studies have been carried out on the possible relationships between availability of pornography and the frequency of sexual crimes. Ad hoc inferences have also been made about cross-country differences in such offending between repressive and more liberal regimes. Due to problems of data accuracy and consistency and difficulties in causality testing, there is no conclusive evidence from these studies that pornography does have a harmful effect. In any case, legislation and policy proceed largely parallel to evidence. The chief force for change in regulation is the continued evolution of Internet markets.

Supply-side regulation is inherently difficult in Internet markets. Thus demand-side controls in Internet markets have been growing following the style of those for circulating media. For printed magazines, we have seen requirements for shrink-wrapping in an opaque cover and/or location in a position in a shop where it is not easily seen, by accident, by children or others likely to take offense. In the case of broadcast porn, we have seen restriction to channels that are not "free to air" (FTA) and can therefore be blocked to inappropriate consumers. However there has been a proliferation, in many countries, of FTA "Babestation" channels which grew up under the rubric of being nonsexually

offensive chat/social entertainment sites. Frequent complaints against breach of this in the UK have been lodged with relatively few fines being levied by the broadcast regulators. These stations continue to broadcast on FTA television material which regular channels would not be allowed to provide. Such channels operate also on the Internet where they provide more and more graphic content. This is of course material that would be legal as porn per se. It would not be an offense to possess copies for private use.

Prevention of access to Internet porn requires that the would-be consumer is unable to access the site. This can be accomplished by specific software that refuses connection to a site. Such software has been mainly aimed at parents wishing to prevent children seeing porn. ISP (Internet service provider) can also lock certain sites as unsuitable.

A fundamental difficulty with expanded regulation is the potential loss of revenue to ISP if blocking is used especially if "opting in to porn" on user sign-up is invoked. This does not come from sales of porn as such but from the degree of Internet traffic it generates which proportionally raises the advertising-based revenue streams. ISP may not be keen on surveillance but its use to target offenders is of much less revenue threat and is very difficult to resist both technically and also politically due to the threats of terrorism.

From a legal point of view, there is a change in some recent results of police seizure of people's computers and Internet history records. That is, intent is coming to the fore as a ground for prosecution, as opposed to mere possession. In 2014 arrests began from a global initiative involving 19 countries which explored data from around 10,000 Internet users had accessed more than 30 websites carrying pedophilia. For practical reasons, the number of suspects was narrowed down to just over 400 suspects of distributing pedophile images. The 19 countries where warrants were executed were Australia, Belgium, Canada, France, Germany, Israel, Italy, Japan, Korea, the Netherlands, New Zealand, Portugal, Russia, Spain, Sweden, Taiwan, Turkey, the UK and the USA. One of the cases in the 2014 trial of entertainer, Rolf Harris, in the UK for sex offenses was brought on grounds of accessing child porn. His Internet searches were offered as evidence by the prosecution on the grounds that he had clearly used words intended to find sexual images of children. His defense pointed out that the models, on such sites, were simply posing as underage given that compliance with employment directives was observed by these sites. Hence, on a consumption-based charge, he was innocent. This was upheld but we see here serious grounds being given for a charge based on intent.

The above measures are intrinsically based on restricting the quantity of a good which has "bad" effects. They incur costs and thus it is possible that there could be "too much" prevention of porn in the sense that aggregate social benefit, of so doing, exceeds aggregate social cost. The costs may include wrongful arrest and prosecution.

Traditional economic models might suggest that price regulation is likely to be more efficient than quantity regulation. A porn tax might seem attractive (analogously to a Pigouvian pollution tax), but in terms of more modern welfare economics, the option of selling permits, or licenses, to porn merchants would seem a better price-based solution. While the sale value of the permits could be determined at a level that is social welfare maximizing, both approaches face the difficulty of being seen as dangerous due to making porn seem like a legitimate business activity.

References

George D (2001) Preference pollution: how markets create the desires we dislike. The University of Michigan Press, Ann Arbor

Kutchinsky B, Snare A (ed) (1999) Law, pornography and crime – the Danish experience, Scandinavian studies in criminology, vol 16. Pax Forlag, Norway.. www.krim.ku.dk\Jesperkrim.ku.dk\Jesper

Kutchinsky B (1985) Pornography and its effects in Denmark and the United States: a rejoinder and beyond. Comparative Social Research, JAI Press, Greenwich, CT

Positional Goods and Legal Orderings

Ugo Pagano[1,2] and Massimiliano Vatiero[3]
[1]University of Siena, Siena, Italy
[2]Central European University, Budapest, Hungary
[3]"Brenno Galli" Chair of Law and Economics, Institute of Law (IDUSI), Università della Svizzera italiana (USI), Lugano, Switzerland

Upon passing by a small village of barbarians, Julius Caesar asserted
"[f]or my part, I had rather be the first man among these fellows than the second man in Rome".
(Plutarch)

Abstract

People consume because others consume, maintained Veblen in 1899. More recently, theoretical, empirical, and experimental articles have argued that people constantly compare themselves to their environments and care greatly about their relative positions.

Given that competition for positions may produce social costs, we adopt a *Law and Economics* approach (i) to suggest legal remedies for positional competition and (ii) to argue that, because legal relations are characterized in turn by positional characteristics, such legal remedies do not represent "free lunches.

The Issue

A positional good is an economic good that depends largely on comparison of one's own consumption with that of others (Hirsch 1976; McAdams 1992; Pagano 1999; Hopkins and Kornienko 2004; Schneider 2007; Vatiero 2009; Fiorito and Vatiero 2013). Positional concerns refer to the fact that individuals consider their rank, relative standing, or position when they evaluate their situation and act upon this evaluation.

For instance, in the above quotation from Plutarch, Julius Caesar (Roman emperor) admitted his willingness to renounce the larger and better private and public goods that he could consume in Rome (e.g., panem et circenses) to gain the position of sovereign (which is a positional good) in a relatively poorer community. Likewise, in John Milton's *Paradise Lost*, Satan stated that it is "[b]etter to *reign* in Hell, than *serve* in Heaven" (emphasis added). In other words, individuals like Caesar and angels like Satan prefer to be the first or the leader in a reference group, even if it is a second-rate one (e.g., a barbarian village or hell). The recent experimental literature corroborates this "position matters" argument (e.g., Solnick and Hemenway 1998, 2005).

Similarly, Hopkins and Kornienko (2004) state, "it is not just that the car is big [enough for needs] but that it is bigger than those owned by the neighbours that also matters" (Hopkins and Kornienko 2004:1087–1088). Carlsson et al. (2007) test this size-matters hypothesis with an experiment and find that people prefer cars bigger than the average size in their society.

A further example is provided by San Gimignano, a small Italian town close to Florence and Siena. In the past, San Gimignano was characterized by about 80 towers (today there are "only" 20 of those towers remaining). The families of San Gimignano did not build towers to live in them or for military defense (because they were unfit for habitation or for fortification) but rather to display their power, affluence, wealth, and status to the rest of the community. Similarly to the "size-matters" argument in Hopkins and Kornienko, in the case of San Gimignano's towers, tallness matters.

Positional concerns are pervasive in numerous socioeconomic domains (see Solnick and Hemenway 2005) and play a pivotal role in people's happiness (e.g., Frey and Stutzer 2002; Clark et al. 2008). This entry wants to deal with negative effects of positional competition and institutional remedies.

Some consequences of competition for positions are illustrated by the following example concerning a labor relationship (see Frank 2012). Consider two types of employment contract, distinguished between wage and safety at work. The former type of contract is characterized by a wage relatively higher *but* a level of workplace safety

relatively lower than those of the latter type. Denoting with w_L, w^H, and R, respectively, the low(er) wage, the high(er) wage, and disutility for unsafe work (e.g., risks of injuries) assume that

$$w_L > w^H - R \quad (1)$$

That is, the disutility due to an unsafe workplace is relatively higher than the increase in wage. In other words, the choice of the contract with unsafe work may produce social costs (i.e., higher health expenditures).

In this game, the Nash equilibrium (*safe work and low wage* for both parties) is efficient (see Fig. 1).

Now assume that positions matter. For instance, a worker with a higher wage can acquire a relatively larg(er) car: that is, there is a positive utility POS deriving from position (e.g., being the worker with the highest wage in the reference group) as well as a corresponding negative utility − POS (for being the worker with the lowest wage in the reference group). The choice of the worker changes as in Fig. 2.

In particular, if $w_L - \text{POS} < w^H - R < w_L < w^H - R + \text{POS}$, then the game becomes a prisoner's dilemma: the Nash equilibrium (both parties choose *unsafe work and high wage*) is Pareto inefficient.

Indeed, even if the second type of contract (lower wage *but* a higher level of workplace safety) is efficient in terms of social welfare (because health expenditures with lower workplace safety are relatively higher than the increase in wage, cf. Eq. 1), each worker prefers the first type of contract because positional competition (on the wage) is important. This means that agents may substitute a non-positional good (i.e., safety at work) with a positional good (wealth), thus causing social costs (i.e., health expenditures). Moreover, if every employee chooses the employment contract with a high wage, then no agent will enjoy a positional advantage in wealth. In equilibrium, no one will consume a positional good (i.e., the biggest car), but all workers will "consume" a lower level of safety at work.

Hence, the consumption of positional goods may produce the following social costs:

1. Agents could substitute a non-positional good (e.g., a private and/or public good) with a positional good, leading to suboptimal equilibria (see also Frank 1985a, b, 2008).

		Friday	
		Safe work and low wage	Unsafe work and high wage
Robinson	Safe work and low wage	$w_L; w_L$	$w_L; w^H - R$
	Unsafe work and high wage	$w^H - R; w_L$	$w^H - R; w^H - R$

Positional Goods and Legal Orderings, Fig. 1 The trade-off between safety (or health) and wage

		Friday	
		Safe work and low wage	Unsafe work and high wage
Robinson	Safe work and low wage	$w_L; w_L$	$w_L - POS; w^H - R + POS$
	Unsafe work and high wage	$w^H - R + POS; w_L - POS$	$w^H - R; w^H - R$

Positional Goods and Legal Orderings, Fig. 2 The choice when positions do matter

2. Because the "parallel investments" in obtaining positional goods may lead to the situation in which no one consumes positional goods, positional competition may waste the economic resources of agents (see Pagano 1999).

How should institutions regulate the competition for positions? What are institutional remedies? A first group, say of *Public Economics*, involves Pigouvian remedies, i.e., a progressive consumption tax on luxury/positional goods (see Frank 1985a, b, 2008). A second group, say of *Law and Economics* (see definition of Marciano 2016, ▶ "Economic Analysis of Law"), may consider rules as means to reduce social costs due to positional competition. This second group is the focus of the next two sections.

Institutional Remedies

How could norms reduce social costs due to positional competition? We investigate three types of *Law and Economics* remedies: norms which (i) restrain/punish the consumption of positional goods, (ii) make the consumption of non-positional goods compulsory, and (iii) encourage cooperation by agents competing in positions.

(i) *Restraining the consumption of positional goods*

Because of the social costs deriving from the consumption of positional goods, the lawmaker may penalize or prohibit the consumption of positional goods. An example is the so-called sumptuary law – Black's Law Dictionary defines such a law as made for the purpose of restraining luxury or extravagance, particularly against inordinate expenditures in the matter of apparel, food, furniture, etc. Sumptuary laws were enacted in Ancient Greece and Rome, from the Middle Ages onward in France and England, and again, in the seventeenth century, in the American Colonies and in feudal Japan (cf. Dari-Mattiacci and Plisecka 2012).

For instance, in ancient Rome, a series of laws governed the materials of which garments could be made and the number of guests at entertainments. In France, Philip IV issued regulations governing the dress and the entertainments of the various social orders. Under later French kings, the use of gold and silver embroidery, silk fabrics, and fine linen was restricted. In 1433, an act of the Scottish Parliament prescribed the lifestyle in Scotland, even going so far as to limit the consumption of pies and baked meats. In feudal Japan, an imperial edict regulated the size of houses and imposed restrictions on the materials that could be used in their construction. Rules in the Tokugawa period in Japan specified the sorts of toys that parents could give their children. For the Aztecs, *macehualtin* – members of the laboring class who displayed finery and precious objects – could be put to death. An interesting and "romantic" view on sumptuary law is in *Dei Sepolcri* by Ugo Foscolo, where the author condemned the Napoleon edict of Saint-Cloud on *positional* expenditures for funerals.

However, attitudes on sumptuary law changed with the Enlightenment, industrial mass production, and consumer-oriented societies. For this reason, there are today limited examples in which legal norms discourage or punish the consumption of a positional good – an exception could be represented by the case of the case of "power": Power is a positional good (Pagano 1999; Vatiero 2009) whose consumption in liberal countries is legally and formally limited, e.g., by antitrust law, labor law, and public law (see Vatiero 2009), which may represent modern forms of sumptuary norms.

Although in liberal society there are not clear examples of legal norms which punish the consumption of positional competition, social norms may do so. This is the case of the tenth commandment "You shall not covet your neighbor's house; you shall not covet your neighbor's wife or his servant or his ox or his donkey or anything that belongs to your neighbor." Because the commandment implies punishment in the case of *envious* choices and conducts, it should affect the behaviors of agents, at least for Christians and Hebrews, and may mitigate positional competition.

(ii) *Making the consumption of non-positional goods compulsory*

Because of the social costs due to the consumption of positional goods, the lawmaker may render (a minimum level of) the consumption of non-positional goods compulsory. That is, the legal system may provide norms which reduce the substitution between positional goods and non-positional goods by defining a non-renegotiable minimum consumption of non-positional goods.

Indeed, labor law establishes non-renegotiable minimum conditions on safety at work. For instance, the employer must provide and maintain a working environment that is safe and without risk to the health of the workers; and workers must use safety equipment with care and act according to prescribed instructions to safeguard their health and protect themselves against injury.

Accordingly, most labor law in the Western countries forbids the renegotiation of these conditions on safety, even against the worker's will. Indeed, the worker may prefer a higher wage at the expense of safety at work. From the perspective of positional competition, this prohibition on renegotiating minimum safety conditions at work is efficient because it reduces the emergence of social costs related to positional competition.

(iii) *Encouraging cooperation among positional competitors*

Because of the social costs due to the consumption of positional goods, the policy maker may encourage cooperation and collaboration among agents (i.e., positional competitors). That is, the wasteful competition for positional goods represents a problem of coordination among agents. For instance, in the case of employment contracts, workers' choices lead to a Pareto-inefficient equilibrium with a lower level of safety at work and nobody consuming a positive level of positional goods. Workers may coordinate to move to a Pareto-superior equilibrium (where nobody still consumes a positive level of positional goods, but all parties have higher levels of safety at work). This implies that cooperation among agents (e.g., workers) may improve the efficiency of individuals' choices (e.g., in terms of labor safety).

In other words, the labor laws which sustain worker unions and collective bargaining could encourage coordination/cooperation among workers and, because unions take the negative effects of positional competition into account, reduce the substitution of non-positional goods for positional goods.

Legal Positions as Positional Goods

While in the preceding section we argued that legal norms may diminish inefficiencies due to positional competition, in this section we illustrate how legal norms in turn create positional concerns.

According to legal theorists such as Wesley N. Hohfeld and old institutionalists like John Commons, each jural relation is linked to a jural correlative (Pagano 2000; Vatiero 2010; Fiorito and Vatiero 2011). The above figure displays the eight fundamental conceptions with which all legal problems may be stated: claim, duty, liberty, no right, power, liability, immunity, and disability (Fig. 3).

Claim/duty relation. In a simplified situation with two agents, say a Dominus and a Servus, a *claim* means that the Dominus has a state-sanctioned assurance that the Servus will behave in a certain way toward Dominus. However, this occurs if and only if the Servus has the *duty* to engage in such behavior with respect to Dominus. That is, a duty is the legal position of the Servus, who is commanded by society to act for the benefit of Dominus, and who will be penalized by society for disobedience. Hence, the correlative of a claim is a duty.

Liberty/no-right relation. *Liberty* stands for one's freedom from the claim of someone else.

	Jural correlatives			
Dominus	Claim	Liberty	Power	Immunity
Servus	Duty	No-right	Liability	Disability

Positional Goods and Legal Orderings, Fig. 3 Fundamental legal conceptions

Similar to the claim-duty relationship, the Dominus has a liberty to behave in a certain way toward Servus if and only if the Servus has *no-right* toward the Dominus to prevent the Dominus from behaving in a certain way. No-right is therefore the legal correlative of a liberty of another party.

Power/liability relation. *Power* is the legal ability to do certain acts that alter legal relations. The Dominus's power is when the Dominus's own voluntary act will cause new legal relations between the Dominus and the Servus, against the Servus's will. This implies that, whenever a power exists, there is at least one other human being whose legal relation will be altered when the power is exercised. The person whose legal relation will be altered is under a *liability*.

Immunity/disability relation. Finally, *immunity* is any legal situation in which a given relation vested in one person cannot be changed by acts of another person. Correlatively, the one who lacks the legal ability to alter the other individual's legal relations is said to be under a *disability*.

The correlative nature of legal positions means that each legal position is available to an individual if and only if a corresponding legal position is occupied by some other individual. In particular, legal positions are adversarial in nature (see also Vatiero 2013a). For instance, claims of one individual imply, at the same time, duties for some another individual and vice versa. That is, the set of actions that defines the claims of the Dominus imposes duties on some individual(s), e.g., the Servus. This brings about the consumption of legal positions with opposite signs: claim by the Dominus, as his/her desired output, and duty by the Servus, as his/her costly input. In a similar manner, the power-liability relationship consists of Dominus's benefit (i.e., the power) as well as Servus's cost (i.e., liability). Hence, any utility deriving from rights and powers must jointly relate with disutility deriving from duties and liabilities (see Fig. 4).

Unlike traditional economic goods, jural positions inevitably involve consumptions and utilities with opposite signs. *Everyone* cannot consume claims, liberties, powers, and immunities; for some individuals, the exercise of these jural positions must imply the exercise of "unfavorable" correlative jural relations (i.e., duties, no rights, liabilities, and disabilities). Because of their adversarial nature, we can represent jural positions as positional goods (see also Pagano 2000; Vatiero 2013a).

Claim/Power	Duty/Disability
Benefits	*Costs*
Desired output	*Costly input*
Utility	*Disutility*

Positional Goods and Legal Orderings, Fig. 4 The adversarial nature of legal relations

This means, moreover, that the definition of rights, duties, powers, etc. with the purpose of mitigating positional concerns, as illustrated in the previous section, can in turn create positional concerns because judicial positions are positional goods. Following Coasean main contribution, no institution is a "free lunch" (Grillo 1995; Pagano 2012; Vatiero 2013b; Pagano and Vatiero 2015). Thus, in the case of positional goods one should take into account benefits deriving from an institutional arrangement able to mitigate positional competition and reduce related social costs but also the costs involved in that institutional arrangement which is in turn characterized, owing to the adversarial nature of legal relations, by positional concerns.

Conclusions

Positions matter in the choices and happiness of agents. People constantly compare themselves to their environments and care greatly about their relative positions, which impact on their choices.

The consumption of positional goods may produce social costs. On the one hand, agents may substitute a non-positional good with a positional good; on the other hand, as an arms race, positional competition may waste the economic resources of positional competitors. Both consequences lead to suboptimal equilibria.

Law and Economics scholars should investigate legal remedies for social costs due to competition for positions. They should take serious account of the fact that legal remedies are not free lunches. In this regard, future research could investigate the costs and benefits that would derive from the definition of a Coasean market (Cf. Medema 2014, ▶ "Coase Theorem"; Bertrand 2015, ▶ "Coase and Property Rights" in this encyclopedia) for positional goods.

Cross-References

- ▶ Coase and Property Rights
- ▶ Coase Theorem
- ▶ Economic Analysis of Law
- ▶ Externalities
- ▶ Institutional Economics

References

Carlsson F, Johansson-Stenman O, Martinsson P (2007) Do you enjoy having more than others? Survey evidence of positional goods. Economica 74:586–598

Clark AE, Frijters P, Shields MA (2008) Relative income, happiness, and utility: an explanation for the Easterlin paradox and other puzzles. J Econ Lit 74(2):425–467

Dari-Mattiacci G, Plisecka AE (2012) Luxury in ancient Rome. An economic analysis of the scope, timing and enforcement of sumptuary laws. Leg Roots, Int J Roman Law, Leg Hist Comp Law 1:189–216

Fiorito L, Vatiero M (2011) Beyond legal relations: Wesley Newcomb Hohfeld's influence on American institutionalism. J Econ Issues 45(1):199–222

Fiorito L, Vatiero M (2013) A joint reading of positional and relational goods. Econ Politica – J Anal Inst Econ 30(1):87–96

Frank RH (1985a) The demand for unobservable and other nonpositional goods. Am Econ Rev 75(1):101–116

Frank RH (1985b) Choosing the right pond: human behavior and the quest for status. Oxford University Press, New York

Frank RH (2008) Should public policy respond to positional externalities? J Public Econ 92:1777–1786

Frank RH (2012) The Darwin economy. Princeton University Press, Princeton

Frey BS, Stutzer A (2002) What can economists learn from happiness research? J Econ Lit 40:402–435

Grillo M (1995) Introduzione. In: Grillo M (ed) Impresa, Mercato e Diritto. Bologna, Mulino

Hirsch F (1976) The social limits to growth. Harvard University Press, Cambridge

Hopkins E, Kornienko T (2004) Running to keep in the same place: consumer choice as a game of status. Am Econ Rev 94(4):1085–1107

McAdams RH (1992) Relative preferences. Yale Law J 102(1):1–104

Pagano U (1999) Is power an economic good? Notes on social scarcity and the economics of positional goods. In: Bowles S, Franzini M, Pagano U (eds) The politics and the economics of power. Routledge, London, pp 116–145

Pagano U (2000) Public markets, private orderings and corporate governance. Int Rev Law Econ 20(4):453–477

Pagano U (2012) No institution is a free lunch: a reconstruction of Ronald Coase. Int Rev Econ 59(2):189–200

Pagano U, Vatiero M (2015) Costly institutions as substitutes: novelty and limits of the Coasian approach. J Inst Econ 11(2):265–281. Coase memorial issue

Schneider M (2007) The nature, history and significance of the concept of positional goods. Hist Econ Rev 45:60–81

Solnick SJ, Hemenway D (1998) Is more always better? A survey on positional concerns. J Econ Behav Organ 37(3):373–383

Solnick SJ, Hemenway D (2005) Are positional concerns stronger in some domains than in others? Am Econ Rev 95(2):147–151

Vatiero M (2009) Understanding power. A law and economics' approach. VDM-Verlag, Saarbrucken

Vatiero M (2010) From W. N. Hohfeld to J. R. Commons, and beyond? Am J Econ Sociol 69(2):840–866

Vatiero M (2013a) Positional goods and Robert Lee Hale's legal economics. J Inst Econ 9(3):351–362

Vatiero M (2013b) Alla ricerca di regole (e istituzioni) efficienti. Riv Crit Dirit Priv 31(1):123–138

Veblen T (1899) The theory of the leisure class. MacMillan, New York

Posner, Richard

Alain Marciano
MRE and University of Montpellier, Montpellier, France
Faculté d'Economie, Université de Montpellier and LAMETA-UMR CNRS, Montpellier, France

Abstract

Posner is one of the main contributors to what is known as "economic analysis of law." In this entry, we restrict our presentation to a few controversial claims he made (efficiency, wealth-maximization, Hicks-Kaldor, judicial decision making).

Biography

Richard A. Posner was born in New York city January 11, 1939; he graduated from Harvard Law School and taught at Stanford University Law School and at the University of Chicago Law School; he is judge of the US Court of Appeals for the Seventh Circuit where he was appointed in 1981 and for which he was chief judge from 1993 to 2000.

General Presentation

The first point that could be noted is that Posner is one of the most important judges of all times in the USA. Christopher McCurdy and Ryan Thompson wrote in 2011, he "can arguably be called the most influential judge currently on the bench" (p. 50). Indeed, Posner is "a staple in legal casebooks . . . his name continues to show up in public discourse and peer judge interviews" (2011, p. 3). They cite a few figures that impressively witness of Posner's influence: between 1998 and 2000, Judge Posner was cited 1,406 times (figure computed by Choi and Mitu Gulati 2004); between 1989 and 1991, he ranked third on the list of the "Top Twenty-Five Prestige Scores (Klein and Morrisroe 1999) and finally 118 Judge Posner opinions appeared in casebooks used in the 1999–2000 school year – ten times more than 90% of federal circuit judges. Yet, the American Bar Association – on a scale going from "exceptionally well qualified," "well qualified," "qualified," to "not qualified" – gave him the "lowest possible ratings, 'qualified/not qualified'" (Lott 2006). This is also impressive and reveals that, as important as he might be, Posner remains controversial in his judicial decision making.

Second, Posner is and has always been a public intellectual (Fleury and Marciano 2013). Writing for nonacademic audiences has never been secondary for Posner. It became more and more important these last 10 years, after 2004, when he launched the Becker-Posner blog (becker-posner-blog.com/), with economist and 1992 Nobel prize winner Gary Becker. What Posner posts on these blogs might seem a bit far-fetched sometimes. It would have been useful and interesting to look at these writings in detail. For a lack of place, we left it aside.

Third, and obviously a major aspect of Posner's work: his academic and scientific writing. Posner has been so prolific that is impossible to summarize his ideas – one can find his list of publications on the website of the University of Chicago (see http://www.law.uchicago.edu/faculty/posner-r). One can nonetheless emphasize a few important ideas he put forward.

Innovative and Original Aspects

Economic Analysis of Law

In 1973 was published the first edition of *Economic Analysis of Law*. At the time, the book was reviewed as a contribution in *law and economics*: the book could "serve very well for a law and economics course" (Diamond 1974, p. 294) and indeed is a "coursebook in law-and-economics" (Krier 1974, p. 1697). At best, it was noted that "[w]ith the publication of Richard A. Posner's economic analysis of law, that field of learning known as 'Law and Economics' [had] reached a stage of extended explicitness" (Leff 1974, p. 451). What was not viewed was that Posner had launched a new field. Economic analysis of law is not simply another name for from law and economics (see Harnay and Marciano 2009). By contrast with "law and economics" that defines economics by its subject matter, an economic analysis of law assumes that economics is a method, an approach or a set of tools that is used to understand nonmarket phenomena. It emerged in the early 1970s under the influence of Gary Becker (Fleury 2015). In this perspective, the law became an object that economists could analyze. Becker (1968), William Landes (1967, 1968, 1971) or Isaac Ehrlich (1970) proposed the first economic analyses of legal problems. Posner was not only the first one to name the field, with the title of his book, he was also the first to systematically adopt such an approach (see criticisms, in particular, in Backhaus 2017 and Malecka 2017).

Efficiency, as Wealth Maximization

One of the consequences of an economic analysis of law lies in the possibility of using the concept of efficiency to assess legal systems: rules, behaviors or judicial decisions are legitimate when or because they are efficient, which, to Posner, means when or because they maximize society's wealth.

Let us start by saying that Posner defines wealth in monetary terms exclusively, as

> the value in dollars or dollar equivalents (. . .) of everything in society. It is measured by what people are willing to pay for something or, if they already own it, what they demand in money to give it up. The only kind of preference that counts in a system of wealth maximization is thus one that is backed up by money – in other words, that is registered in a market

(1979, p. 119). Or, "the sum of all tangible and intangible goods and services, weighted by prices of two sorts: offer prices (what people are willing to pay for goods they do not already own); and asking prices (what people demand to sell what they own" (Posner 1995, p. 356).

Such definition of wealth is not as narrow as might appear at first sight. Indeed, Posner adopts a rather extended conception of markets. First, he does not restrict to explicit markets. But he considers implicit markets, in which services are sold that can be monetized by reference to substitute services sold in explicit markets. This enables Posner to use the wealth maximization criterion to assess the pecuniary value of a wide set of goods and services. Second, he also considers hypothetical markets, defined as those markets in which high transaction costs prevent individuals to transact with each other voluntarily and, therefore, efficient transactions to occur – for instance, individuals may be forced into involuntary transactions in the case of accidents, when a victim is obliged to transact with the injurer and the victim and the injurer cannot agree on a common price. Then, using a third party (e.g., a court) within a hypothetical market to determine the price of the transaction is a way to have the transaction occur actually and to achieve an efficient allocation of resources. Indeed, agents will consent to the accident-as-a-transaction as long as it is wealth-maximizing, i.e., when it is less costly to let the accident occur than to prevent it. On the contrary, they will let the accident occur when it is wealth-enhancing. Then, transactions within hypothetical markets reconcile the possibility of involuntary transactions with the idea of individual consent that lies at the core of the wealth maximization criterion.

Indeed, Posner equates efficiency with wealth maximization. Thus, from this perspective, "[r]esources are efficiently allocated in a system of wealth maximization when there is no reallocation that would increase the wealth of society" (Posner 1980a, p. 243). That is, Posner remains in the standard maximizing framework of utilitarianism but substitutes wealth for "utility."

Using wealth, instead of utility, as most economists usually do, allows Posner to solve the vexed problems of how to measure and to compare individual utilities. In a system based on "wealth maximization," interpersonal comparison of wealth is possible, making it possible to compare the gains of one individual or a group of individuals to the losses incurred by another individual or group. Therefore, a transaction will be considered as legitimate and desirable when it increases the wealth of the society. Posner gives the following example:

> I offer you $5 for a bag of oranges, you accept, and the exchange is consummated. We can be confident that the wealth of the society has been increased. Before the transaction you had a bag of oranges worth less than $5 to you and I had $5; after the transaction you have $5 and I have a bag of oranges worth more than $5 to me. We are both richer, as measured by the money value we attach to the goods in question.

(1979, p. 120). The transaction is then legitimate and even desirable.

Another example is given by Posner: "[c]onsider an accident that inflicts a cost of $100 with a probability of .01 and that would have cost $3 to avoid. The accident is said to be a wealth-maximizing 'transaction' [. . .] because the expected accident cost ($1) is less than the cost of avoidance" (1995, p. 358). Thus, clearly, what this accident costs to the society is less than what it would have costed to avoid it. An

individual who does not spend the $3 to avoid a $1 cost is not negligent. Conversely, if the cost of avoidance is lower than the accident cost, not only the accident is not wealth-maximizing but also an individual who would have not tried to avoid the accident would have been negligent. This is precisely why Posner also praised judge Learned Hand for his 1942 decision in a liability case: it was a wealth maximizing – and therefore efficient – decision (e.g., Posner 1972, p. 32, or 1999, p. 91).

Wealth and Capacity to Pay

Posner considers that wealth must be determined by using individuals' willingness to pay for a good rather than by its price. This is understandable for two reasons: first, on implicit and hypothetical markets, there are no prices and therefore the only way to value a transaction is to know what individuals would like to pay for a good; and second, the real value of a good for an individual is not represented by the price but by his or her willigness to pay. Wealth relates to consumers' surplus.

But then, one must also note that individuals may not only value a good they desire. They should be able to pay for it. Posner meant capacity to pay as a means to determine wealth:

> a desire not backed by ability to pay has no standing–such a desire is neither an offer price nor an asking price. I may desperately desire a BMW, but if I am unwilling or unable to pay its purchase price, society's wealth would not be increased by transferring the BMW from its present owner to me. (1995, p. 357)

or,

> that wanting something very much, but not being able to pay more for it than its owner or competing demanders, does not establish a claim to a good in a system of wealth maximization, although it might do so in a system of utility maximization. Wealth maximization thus excludes claims based on pure desire-claims not backed up by willingness (implying ability) to pay. (1980a, p. 243)

One would say that for a criterion that Posner also wanted to be "ethical" – he "argued that "wealth maximization" provides an ethically attractive norm for social and political choices, such as those made by courts asked to determine whether negligence or strict liability should be the rule for deciding whether an injurer must compensate his victim" (1985, p. 85) – grounding "wealth" on individual's "capacity to pay" is problematic. Posner replied mainly by saying that one should distinguish efficiency from justice and the allocation from the distribution of resources. If a rule, a legal decision or a policy increases and maximizes wealth, then it directly benefits to certain individuals and not to others. Nothing prevents the use of redistributive policy to compensate the effects of such a wealth-increasing rule, decision, or policy.

This relates to another and quite important aspect of Posner's conception of wealth maximization: the Kaldor-Hicks criterion.

Kaldor-Hicks and Wealth Maximization

In economics, efficiency is measured in terms of Pareto optimality. However, Posner insisted that "Pareto superiority is not a necessary condition to be wealth maximizing" (Posner 1995, p. 357). Posner distinguished wealth from utility maximization à la Pareto. To him, a transaction, decision, or policy may be "wealth maximizing even if the victim is *not* compensated" (Posner 1995, p. 358; emphasis added). Precisely, to know if an allocation of resources is efficient without the Pareto criterion, which implies effective compensation, one may use the Kaldor-Hicks test, which rests on potential compensation. In other words, effective compensation is not required to characterize a move from one situation to another one: "[u]nder the Kaldor-Hicks definition of efficiency, which is also widely used by economists, a reallocation of resources is efficient if it enables the gainers to compensate the losers, whether or not they actually do so. This is equivalent to wealth maximization" (Posner 1980a, p. 244). Or, to put it in the more figurative terms Posner himself uses, an efficient decision in the Kaldor-Hicks sense consists in "making the pie larger without worrying about how the relative size of the slices changes" (2000, p. 1155).

Interestingly, this latter formulation reveals one important feature of Kaldor-Hicks efficiency and, therefore, of Posner's wealth maximization criterion: the exclusion of distributional issues and "normative considerations of distributive justice"

as the Kaldor-Hicks criterion "treats a dollar as worth the same to everyone." (2000, p. 1154). However:

> to the extent that distributive justice can be shown to be the proper business of some other branch of government or policy instrument (for example, redistributive taxation and spending) and that ignoring distributive considerations in the particular domain of decision making that is under consideration will not have systematic and substantive distributive consequences, it is possible to set distributive considerations to one side and use the Kaldor-Hicks approach with a good conscience. This assumes, as I have said, that efficiency in the Kaldor-Hicks sense ... is a social value. (Posner 2000, pp. 1154–1155)

From that perspective, Posner adds, "efficiency in the Kaldor-Hicks sense is accepted as a social value, albeit not the only social value." (Posner 2000, p. 1154). In addition, "even if Kaldor-Hicks efficiency has no social value", it is "a convenient instrument" (Posner 2000, p. 1156) that can be used to decide whether or not one should make a decision. This is where Posner links the "Kaldor-Hicks concept of efficiency" to cost-benefit analysis: decisions can be made, rules modified, policy measures taken by comparing costs and benefits without the winners being obliged to compensate the losers. Then, if the benefits are greater than the costs, wealth is maximized; on the contrary, if costs exceed benefits, wealth is not maximized. Thus, the legitimacy of CBA comes from the use of a Kaldor-Hicks definition of efficiency Reciprocally, Posner's defense of CBA is also a defense of the Kaldor-Hicks principle.

Common Law and Efficiency

A major positive claim in the law and economics literature that was first formulated by Posner is that the common law is efficient. Indeed, the development of common law is mainly driven by judicial decisions issued in response to the needs of the society through cases. When facing new cases for which no precedent is available, judges make decisions relying on the economic efficiency criterion, defined as wealth maximization or Kaldor-Hicks efficiency. Thus, they are assumed to have a preference for efficiency over other moral values (such as income redistribution) and to share that definition of efficiency as a desirable social goal with the society. Further, even if they prefer other values, they issue decisions promoting efficiency because the limitations of the judicial process constrain them in their ability to pursue other objectives. In particular, unlike legislators, they are poorly equipped with tools allowing them to engage in effective wealth redistribution or other unattainable values (see for instance Posner 1973, 1979, 2010). As a consequence, judges having internalized efficiency as a socially desirable goal, even unconsciously, adopt an economic logic when they make their decisions. Indeed,

> [T]he character of common law litigation forces a confrontation with economic issues. The typical common law case involves a dispute between two parties over which one should bear a loss. In searching for a reasonably objective and impartial standard, as the traditions of the bench require him to do, the judge can hardly fail to consider whether the loss was the product of wasteful, uneconomical resource use. In a culture of scarcity, this is an urgent, an inescapable question. And at least an approximation to the answer is in most cases reasonably accessible to intuition and common sense. (Posner 1973, p. 99)

Parties to litigation consent to efficient decisions made by courts because their self-interest is promoted by supporting such decisions. Namely, "by doing so they increase the wealth of the society; they will get a share of that increased wealth; and there is no alternative norm that would yield a larger share" (Posner 1980b, p. 505). Furthermore, independent judges are insulated from interest group pressures and personal factors (Posner 1993) and can therefore promote social efficiency.

Resting on the examination of particular common law fields, much of Posner's works are forceful attempts to determine what legal rules would be efficient and to examine how they correspond to the legal rules that exist. Along this line, he analyzes various fields of common law, including contract law, liability rules, property law, criminal law, family law, and sex law, using the lens of economic efficiency. Then, common law appears to differ from a collection of multiple and conflicting precedents whose coherence may at

first sight be uneasy to grasp. But it may rather be best understood as shaped by an economic – albeit most of the time implicit – economic logic underlying all judicial decisions and intending to maximize social wealth. Posner's theory of the efficiency of common law helps thus rationalize common law within a single – economic – framework. In his own words (Posner 2011, pp. 315–316),

> economics is the deep structure of the common law, and the doctrines of that law are the surface structure. The doctrines, understood in economic terms, form a coherent system for inducing people to behave efficiently, not only in explicit markets but across the whole range of social interactions.

Posner's claim that the common law is efficient has fueled an important literature, either criticizing or agreeing with him. One of the most debated issues deals with the mechanisms likely to explain the evolution of law towards efficiency. A substantial body of literature has raised doubts about the existence of such mechanisms, for various reasons. First, the judicial preference for efficiency over other moral values and preferences (for leisure, reputation) has been a matter of considerable debate, whereas the apparent disinterest of judges in efficiency, at least as expressed in common-law court opinions, has been emphasized. Thus, in response to Posner's seminal analysis, various demand-side models have been developed, calling attention to forces other than judicial preferences in explaining the law and suggesting that selective litigation of rules by parties may actually drive legal evolution and help promote efficient legal common law rules, regardless of judicial preferences, as inefficient rules may be challenged before courts and therefore overturned more often than efficient ones (see for instance Goodman 1978; Landes and Posner 1979; Priest 1977; Rubin 1977). Also in response to Posner's theory of the evolution of common law, other works have cast doubt on the very fact that common law is efficient. Some authors point out that common law adjudication may rather lead to cycles of efficiency and inefficiency (Cooter and Kornhauser 1980). Others argue that the excessive amount of information needed for judges to decide cases may prevent them from making efficient decisions (Aranson 1992; Rizzo 1980). It has also been argued that common law may lead to inefficient lock-in and path-dependency when it preserves obsolescent rules. Furthermore, Tullock (1980) has claimed that the English common law adjudication is less efficient than legislation – public choice critics mainly focusing on the idea that judicial processes are subject to the same sorts of interest-group pressures as are legislatures. More recently, the legal origins literature has revivified the debates on the common law efficiency through its attempts to demonstrate the superior economic performance of common-law legal systems over other systems.

Judges (Judicial Preferences and Behavior)

Posner devoted a lot of work analyzing Judges and judicial behavior. It was one of the first areas that he viewed as important when he started to use economics to analyze the law. In 1971, he thus wrote that judges seem to display a certain propensity towards efficiency – they "are guided by concern with economic efficiency" (1971, p. 223) and "think in economic terms" (1971, p. 224). Or, in 1973, he insisted that one can "*assume* that judges make their decision in accordance with the criterion of efficiency" (1973, p. 325; emphasis added). What was important to Posner was to explain the efficiency of the Common Law. To be more precise, there were two versions of this theory. First, the positive version holds that rational judges having a preference for efficiency do actually issue decisions that maximize social wealth; hence, common law tends towards efficiency. Second, the normative view argues that judicial decisions should help common law to evolve towards efficiency, as rational judges should make their decisions with the wealth maximizing criterion (Posner 1980a).

Eventually, Posner criticized the traditional and idealized view on judicial behavior according to which public interest may be a judicial goal – assuming this is actually the case would amount to acknowledging that judges do not behave as ordinary human beings. To him, there is no reason to argue that judges are different from other individuals. Hence, it can be assumed that they are self-

interested utility maximizers – they maximize "the same thing as everybody else does" (1993). Their utility function may contain both monetary and nonmonetary arguments, such as leisure, popularity, prestige, and reputation. Judicial decisions may also have some pure consumption value for judges who may therefore derive some kind of positive utility from their decisions close to that derived by "artists or craftsmen who value aesthetic excellence in their field" (Posner 2008).

One of the consequences of this view is that judges may not follow the precedent and may be led to innovate. This means that this is consistent with the idea that judges create the law, that they acting as *interstitial legislators* and significant policy-makers.

Over time, Posner himself has also continued documenting the key role played by judges from both a theoretical and empirical standpoint, regularly publishing contributions on adjudication and judicial behavior – the room continuously dedicated to that issue in the successive editions of his textbook *Economic Analysis of Law*, as well as his most recent publications (2008, or 2013 with Epstein and Landes), obviously express his continuing interest for the topic.

Legacy

It is impossible to summarize Posner's legacy. His impact on economic analyses of law is huge. It is probably not exaggerate to claim that all contributors to law and economics/economic analysis of law are in one way or another "Posnerians". Either because they agree or because they disagree with him. Posner has radically transformed law and economics and also probably the legal system.

Cross-References

- Becker, Gary S.
- Coase, Ronald
- Economic Analysis of Law
- Wealth Maximization: Efficiency and Equality Considerations

References

Aranson PH (1992) The common law as central economic planning. Constit Polit Econ 3(3): 289–317

Backhaus JG (2017) Lawyers' economics versus economic analysis of law: a critique of professor Posner's "economic" approach to law by reference to a case concerning damages for loss of earning capacity. Eur J Law Econ 43(3):517–534

Becker GS (1968) Crime and punishment: an economic approach. J Polit Econ 76(2):169–217

Choi SJ, Mitu Gulati G (2004) Choosing the next supreme court justice: an empirical ranking of judge performance. Calif Law Rev 78(1):23–117

Cooter RD, Kornhauser LA (1980) Can litigation improve the law without the help of judges? J Leg Stud, 9:139–163

Diamond PA (1974) Review of economic analysis of law, by Richard A. Posner. Bell J of Econ and Management Science 5(1):294–300

Domnarski W (2016) Richard Posner. Oxford University Press, New York

Ehrlich I (1970) Participation in illegitimate activities: an economic analysis. PhD dissertation, Chicago

Epstein L, Landes WM, Posner RA (2013) The behavior of federal judges: a theoretical and empirical study of rational choice. Harvard University Press, Cambridge

Fleury J-B (2015) Massive influence with scarce contributions: the rationalizing economist Gary S. Becker, 1930–2014. Eur J Law Econ 39:3–9

Goodman JC (1978) An economic theory of the evolution of common law. J Leg Stud 7(2):393–406

Harnay S, Marciano A (2009) Economics and the law: from 'law and economics' to an economic analysis of law. J Hist Econ Thought 31(2):215–232

Fleury J-B, Marciano A (2013) Becker and Posner: freedom of speech and public Intellectualship. Hist Polit Econ 45(Suppl):254–278

Klein D, Morrisroe D (1999). The prestige and influence of individual judges on the U.S. courts of appeals. J Leg Stud 28(2), 371–391

Krier JE (1974) Review of economic analysis of law by Richard A. Posner. Univ Pa Law Rev 122(6): 1664–1705

Landes WM (1967) The effect of state fair employment laws on the economic position of nonwhites. Am Econ Rev 57(2):578–590

Landes WM (1968) The economics of fair employment law. J Polit Econ 76(4):507–552

Landes WM (1971) An economic analysis of the courts. J Law Econ 14(1):61–107

Landes WM, Posner RA (1979) Adjudication as a private good. J Leg Stud 8(2):235–284

Leff AA (1974) Economic analysis of law: some realism about nominalism. Virginia Law Review 60(3): 451–82

Lott J (2006) Pulling rank. New York Times, 25 Jan. http://www.nytimes.com/2006/01/25/opinion/25Lott.html?_r=2&
Malecka M (2017) Posner versus Kelsen: the challenges for scientific analysis of law. Eur J Law Econ 43(3):495–516
McCurdy CC, Thompson RP (2011) The power of Posner: a study of prestige and influence in the federal judiciary. Idaho Law Rev 48(1):49–71
Medema SG (2010) Richard A. Posner. in The Elgar Companion to the Chicago School of Economics, Edward Elgar, Cheltenham, 306–310
Posner RA (1971) Killing or wounding to protect a property interest. J Law Econ 14(1):201–232
Posner RA (1972) A theory of negligence. J Leg Stud 1(1):29–96
Posner RA (1973) Economic analysis of law, 1st edn. Aspen Publishers, Boston
Posner RA (1979) Utilitarianism, economics, and legal theory. J Leg Stud 8(1):103–140
Posner RA (1980a) The value of wealth: a comment on Dworkin and Kronman. J Leg Stud 9(2):243–252
Posner RA (1980b) The ethical and political basis if the efficiency norm in common law adjudication. Hostra Law Rev 8:487–507
Posner RA (1985) Wealth maximization revisited. Notre Dame J Law Ethics Public Policy 2:85–105
Posner RA (1993) What do judges and justices maximize? (the same thing everybody else does). Supreme Court Econ Rev 3:1–41
Posner RA (1995) The problem of jurisprudence. Harvard University Press, Cambridge
Posner RA (1999) The law and economics of the economic expert witness. J Econ Perspect 13(2):91–99
Posner RA (2000) Cost-benefit analysis: definition, justification, and comment on conference papers. J Leg Stud 29(2):1153–1177
Posner RA (2008) How judges think. Harvard University Press, Cambridge
Posner RA (2010) Some realism about judges: a reply to Edwards and Livermore. Duke Law J 59(3): 1177–1186
Posner RA (2011) Economic analysis of law, 8th edn. Aspen Publishers, Boston
Priest GL (1977) The common law process and the selection of efficient rules. J Leg Stud 6(1):65–82
Rizzo MJ (1980) The mirage of efficiency. Hofstra Law Rev 8:641–658
Rubin PH (1977) Why is the common law efficient? J Leg Stud 6(1):51–63
Tullock G (1980) Trials on trial, the pure theory of legal procedure. Columbia University Press, New York

Post-grant Patent Review

► Patent Opposition

Power Indices

Dennis Leech
University of Warwick, Coventry, UK

Abstract

Many voting bodies are constituted on a principle of accountability whereby a member's influence is intended to be a reflection of a measure of size such as financial contribution or population. Examples include the joint stock company, the US Electoral College, the IMF executive and the EU council of ministers.

Power indices are a tool for addressing the (often ignored) problem inherent in this: that the constitution defines the voting rules and not the effective voting powers they imply. Indices due to Penrose-Banzhaf and Shapley-Shubik are often used. That they ignore preferences is often cited as a limitation with regard to positivistic analysis of existing constitutions but an advantage when they are used to address the normative problem of designing voting rules. Power indices can reveal hidden properties of voting rules that are not obvious at a superficial level. The issue of how to construct behavioral power indices that do take account of preferences remains an important research dimension. Power indices can also help us understand multi-tiered governance structures such as federal constitutions or corporate networks, an area where there is need for further research.

Power Indices

Many decision-making bodies are designed with voting rules that assign different voting powers to different participants. Examples include intergovernmental organizations such as the United Nations security council, the Bretton Woods institutions, the Council of Ministers of the European Union, and so on. Many are formally constituted to embody a principle of accountability whereby a member's influence through voting is intended to

be a reflection of financial contribution. An example is the joint stock company whereby shareholders' voting powers are directly proportional to the number of shares they own.

Power indices (more properly perhaps known as voting power indices because power is a much broader concept than is simply defined by formal voting rules) are a tool for addressing the important problem inherent in this: that the constitution defines the voting rules and not the voting powers they imply. For example a shareholder who owns 40% of a company's voting shares could have nearly 100% of the voting power in the sense of the ability to determine a decision taken by voting if the other 60% of shares are widely held. But on the other hand, it could be worth less than 40% of the decision-making power if there are other large bloc shareholders. This problem is rarely acknowledged by policy makers and even more rarely addressed. The best reference on voting power is Felsenthal and Machover (1998).

Power indices are quantitative measures of the influence of each voter in terms of the voter's ability to affect the outcome whenever a hypothetical vote is taken. The basis of the definition of the index is an examination of each possible "yes/no" vote that is theoretically possible. A voter has voting power whenever he (or she or it if the voter is an institution) can change the outcome of a vote from "no" to "yes" by changing the way he casts his vote. The power index for any voter is then the proportion of voting outcomes – taking into account all the possible ways all the voters can cast their votes – for which this is possible.

A real-world example of why this approach is important is provided by the history of the original European Economic Community when it was founded in 1958. The council of ministers adopted a system of qualified majority voting with weighted voting whereby the three largest countries (France, Italy, and West Germany) all had four votes each, the smaller countries (Belgium and Netherlands) two, and tiny Luxembourg one. This was intended to ensure that Luxembourg got due recognition as a sovereign state beyond what its tiny population size might warrant if representation was to be strictly proportional to population. So the intention of the framers of the constitution of the EEC was that voting weights be less than proportional to population so that the larger countries would be underrepresented relative to their populations. However an examination of the voting rule against possible voting outcomes indicates that this assumption is far from the truth. The voting rule was that a qualified majority decision required a threshold of 12 or more weighted votes to pass. Therefore, in order for the vote of Luxembourg to be able to make a difference, that is, for Luxembourg have voting power, there would have to be 11 "yes" votes from the other countries. But the constitution guaranteed that was impossible. So Luxembourg had zero voting power although it had almost 6% of the voting weight. The subject of the design of the system of qualified majority voting for the council of ministers, particularly in the Nice and Lisbon treaties, has given rise to a substantial literature on voting power indices.

All this is straightforward, but turning the idea of measuring voting power into a practical voting power index is not. How voting outcomes should be counted – whether orderings or divisions of voters – is not obvious, and this has given rise to different indices that give different results in general. There are two so-called classical indices, often referred to as the Penrose-Banzhaf index (PBI) and the Shapley-Shubik index (SSI) (though nomenclature varies between authors). The SSI is based on the idea that voting outcomes are essentially orderings of voters; each ordering counts equally as a different possible outcome of the voting rules. Each voter has power to the extent they can be the pivot or swing voter able to change the decision from "no" to "yes." On the other hand, the PBI treats each possible division of voters into two groups voting "for" and "against' a decision. A power index is then defined for each voter as the proportion of outcomes for which that voter is the pivot or swing (Banzhaf 1964; Shapley and Shubik 1954; Penrose 1946).

These definitions lend themselves to interpretation as probabilities, each voting outcome being deemed to occur randomly with equal probability, and are often expressed in such terms. This interpretation has led to criticism by scholars who

object to the idea of random voting as a description of real-world behavior. However this probabilistic interpretation is not essential, and a simple interpretation in terms of relative frequencies of outcomes avoids the need to assume anything about behavior. Nevertheless, because direct computation of power indices from basic definitions requires consideration of all outcomes, this can be difficult when the number of voters is large; then algorithms based formally on probability theory are useful (Leech 2003).

Much of the literature fails to get to grips with the issue of which of the two classical power indices is better as a measure of voting power. There tends to be a preponderance in applied work for the PBI, probably because the idea of an outcome of a vote as a division has more intuitive appeal than an ordering, but many studies report the two indices side by side, leaving it to the reader to decide between them. The case against the SSI was first made by Coleman (1971) who argued strongly against the SSI on the basis that treating all orderings of voters is theoretically implausible and the weighting the formula requires is not justified. See Leech (2002), which also showed that results from applying the SSI to British company shareholdings gave results at variance with empirical evidence about company control.

A second major issue, besides the choice of which coalition model to use, relates to what sort of analysis is being done using the voting power indices. Is a voting power index meant to tell us about how voters actually behave in reality or about nothing more than theoretical possibilities? The failure of many writers to recognize this point has led to a lot of – often strongly expressed – discussion. In fact it is important to distinguish between two types of analysis for which power indices are useful: normative and positive.

Normative analysis is concerned with the design of voting rules. It is essential that this is done "a priori" without regard to the preferences or likely voting behavior of the voters. The voting rule is seen as being designed behind a Rawlsian veil of ignorance that attributes equal importance to all voting outcomes that can theoretically occur. It would clearly be morally wrong, for instance, when considering the enlargement of a voting entity such as the European Union, to allocate votes to a new member on the basis of its likely voting behavior: to give it different weighted votes as to whether it was likely to support or oppose a particular grouping. Nothing other than an objective measure such as population would be reasonable as a basis for choosing its voting weight. Therefore for this normative analysis, it would be wrong to take account of preferences (Braham and Holler 2005).

This point has been ignored by positivist writers who have tended to be more interested in the behavior of voters with respect to a *given* voting rule than the *choice* of voting rule. From a positivist point of view, a priori or "constitutional" power indices, the PBI and SSI, have been criticized for not taking account of preferences which would assign different probabilities to different outcomes. In the EU council, for example, a positivist description of voting behavior would give a higher probability to outcomes where France and Germany vote on the same side than where the UK and France or Germany are in agreement. The PBI and SSI are clearly limited measures of voting power from this point of view.

Voter	Weight	BPI	Weight	BPI	Weight	BPI
1	9	0.0	9	5.0	9	10.0
2	17	20.0	16	15.0	9	10.0
3	17	20.0	17	18.3	17	16.7
4	18	20.0	17	18.3	18	16.7
5	19	20.0	17	18.3	19	16.7
6	20	20.0	24	25.0	27	30.0

Percentages. Threshold 51. Normalized BPI. Computations using algorithm at Leech and Leech

There is a need for much more work on developing "a posteriori" or "behavioral" power indices which do take account of preferences and are able to give an empirical account of the distribution of voting power.

One aspect of the normative use of power indices is in enabling the discovery of hidden properties of constitutional designs. In the EEC example above, it was fairly easy to spot that there was a voter with positive weight but zero power. But more complicated voting rules require voting

power analysis. The table below shows some possibilities that can occur with six voters. Fairly innocuous-looking changes to the voting weights assigned to other voters can make a profound difference to the share of voting power enjoyed by the voter number 1 with 9% of the votes, from 0% to 10%. This illustrates a particularly useful application of voting power analysis in being able to find that a voter with positive weight has in fact zero power. Such a voter is known in the literature as dummy. This feature of voting power analysis is especially useful because it is independent both of the index used and also whether or not preferences are taken into account.

The main open research question is the choice of which of the two so-called "classical" power indices, BPI or SSI, is better. So far there is no consensus on the best index to measure voting power. Some scholars prefer the mathematical rigor of the SSI, while others argue that the assumptions underlying the BPI contain more realism. There is also the challenge of developing behavioral power indices.

Much discussion recently has centered on the properties of voting power indices in *large* voting bodies, that is, when the number of voters is large in some sense. It has been argued that there is a general asymptotic tendency, as the number of voters increases without limit, for relative voting powers to become proportional to relative voting weight. Lindner and Machover (2004) have come up with sufficient conditions for this, which they called "Penrose's limit theorem." Leech (2013) however has shown that this is a special case, and there is no tendency toward proportionality of voting power indices and voting weight in general.

An important extension of the voting power indices method as an analytical tool is to multilevel voting rules such as voting power in federal systems (e.g., the US electoral college, see Miller (2013)), company groups with complex ownership structures Crama and Leruth (2007), and many other applications.

Voting power remains an active research field. Two recently published edited volumes containing applications are Holler and Nurmi (2013) and Fara et al. (2014).

References

Banzhaf JF III (1964) Weighted voting doesn't work: a mathematical analysis. Rutgers Law Rev 19:317

Braham M, Holler MJ (2005) The impossibility of a preference-based power index. J Theor Polit 17(1):137–157

Coleman JS (1971) Control of collectivities and the power of a collectivity to act. Soc Choice:269–300

Crama Y, Leruth L (2007) Control and voting power in corporate networks: concepts and computational aspects. Eur J Oper Res 178(3):879–893

Fara R, Leech D, Salles M (2014) Voting power and procedures, Springer

Felsenthal D, Machover M (1998) The measurement of voting power. Edward Elgar, Cheltenham

Holler MJ, Nurmi H (2013) Power, voting, and voting power: 30 years after. Springer, Berlin

Leech D (2002) An empirical comparison of the performance of classical power indices. Political Stud 50(1):1–22

Leech D (2003) Computing power indices for large voting games. Manag Sci 49:831–838

Leech D (2013) Power indices in large voting bodies. Public Choice 155(1–2):61–79

Leech D, Leech R Algorithms for voting power indices: www.ecaae.ac.uk

Lindner I, Machover M (2004) LS Penrose's limit theorem: proof of some special cases. Math Soc Sci 47(1):37–49

Miller NR (2013) A priori voting power and the US Electoral College. In: Holler MJ, Nurmi H (eds) Power, voting, and voting power: 30 years after. Springer, Berlin, pp 411–442

Penrose LS (1946) The elementary statistics of majority voting. J R Stat Soc 109:53–57

Shapley LS, Shubik M (1954) A method for evaluating the distribution of power in a committee system. Am Polit Sci Rev 48(9):787–792

Preferential Tariffs

Laura Huici Sancho
Department of International Law and Economics, University of Barcelona, Barcelona, Spain

Abstract

Tariff preferences refer to measures that involve tariff reductions which benefit certain goods originating in certain countries. Historically tariff preferences were agreed on a reciprocal basis to recognize special relations between states. The multilateralization of

trade implied the generalization of preferences, which become an instrument for development cooperation. Tariff preferences might be accorded, unilaterally or through regional agreement schemes, to developing countries. However the progressive reduction of tariffs in international trade limits the role of tariff preferences as an instrument to agree a special treatment to certain kind of countries. Besides the increasing diversity of states requires of flexible instruments that allow special and differential treatment to better adapt to development, financial and trade needs.

Synonyms

Commercial benefits; Commercial concessions; Tariff benefits; Tariff concessions

Definition

Preferences are concessions which involve special treatment in trade relations between two or more countries, economic integration organizations, or regional groupings. Generally speaking, a benefit stemming from a treaty that is not extended to all states, whether they are party to it or not, may be regarded as a *preference*. However, in international economic law, the term *preferences* is used more specifically to refer to measures that involve tariff reductions which benefit only goods originating in certain countries. With this meaning, preferences aim to liberalize trade relations, since they involve reductions or even the elimination of tariffs. Yet they upset the principle of no discrimination because, by definition, they do not apply equally to all states. So being contrary to the Most Favoured Nation Clause (MFN), in the GATT/WTO system, preferences can find justification in article XXIV, the Enabling Clause, or requesting a specific waiver (GATT article XXV:5 – WTO article IX:3).

Preferences may be reserved only for certain types of products originating in certain countries; they may be limited qualitatively to certain goods and/or be subjected to quantitative restrictions. This unequal treatment, which is characteristic of preferences, has turned them into a major instrument of development cooperation, and they are thus used for building relations between developed and developing countries. Nevertheless, the increasing use of the MFN in trade agreements, the multilateralization of such with the signing of GATT, the progress made in liberalizing trade as a result of successive rounds of trade negotiations, the creation of the WTO, and the expansion of its Member States are factors that have led to a declining role of tariff *preferences* as tariff reductions have gradually become more widespread.

An Historical Overview of Preferences: An Evolving Institution

The notion of preferences may be defined through their historical evolution with emphasis placed on their main features and most controversial points. With this objective in mind, four stages may be distinguished: first, preferences are incorporated into trade agreements on a reciprocal basis and respond primarily to factors concerning particular links between states; second, as part of the multilateralization of trade relations, these tend to favor the generalization of preferences limited to relations between developed and developing countries, thus eliminating reciprocity; third, preferences tend to place more conditions on compliance with certain requirements, particularly environmental and labor protection; and fourth, a crisis of the unilateral tariff preferences model has arisen, raising doubts concerning their interest and continuity in the near future, while regional preferential trade agreements are growing and new models of preferential treatment have appeared.

Reciprocal and Special Preferences

Traditionally, the establishment of a *preferential* relationship between parties was the subject of trade agreements in which states made mutual concessions. Often, the establishment of these preferences responded to the existence of cultural, historical, or geographical ties among the affected countries which sought to encourage and enhance their trade relations with a reciprocal recognition

of preferences. It should be noted that preferential and reciprocal tariff treatment among members is the basis of the free trade areas and customs unions which constitute, in this sense, specific examples of special preferential arrangements.

Granting preferential tariffs became common in relations between the metropolis and its colonies once these had become independent states. One example is the *imperial preferences* that were first established by the United Kingdom among its colonies and later continued with the Commonwealth countries. Despite initial opposition, "at the Imperial Economic Conference of 1932 in Ottawa the United Kingdom finally accepted the extension of imperial preference throughout the Empire. Eleven bilateral agreements were signed and the whole system of Empire preferences was augmented and co-ordinated" (Clute and Wilson 1958, p. 464). As regards the terms of these agreements, "the imperial preference system is a maze of fiscal arrangements, by which each participant admits into its markets certain exports of the others at lower tariff rates than are imposed upon the same products when originating elsewhere" (Feis 1945–1946, p. 663). This was a unique exclusive system of special preferences for Commonwealth members. If a member of the Commonwealth ever offered the same preferential treatment to a nonmember state, it would have to obtain a *waiver* that would allow it to do so. This unequal treatment raised tensions in relations with nonparticipating countries, in particular the USA, which led the multilateralization process of trade relations on the basis of MFN. At the signing of the GATT in 1947, the compatibility of imperial preferences with the MFN clause was recognized but limiting their future expansion (Clute and Wilson 1958, p. 467).

Another example of preferences granted to former colonies are Association Agreements signed by the European Economic Community and the African, Caribbean and Pacific Countries (ACP). The need to provide a special relationship with states that had formerly been colonies of a Member State is explicitly recognized in the Treaty of Rome signed in 1957, which established the European Economic Community (EEC). Based on this, the EEC granted preferences to several countries that were mainly former French and Belgian colonies. In the Yaoundé Convention, held in 1963, the EEC implemented a progressive reduction of tariffs on products originating in the 18 African and Malagasy States which participated in this Convention (Article 2 of the Convention), and tariffs were completely eliminated for certain tropical products. In return, these countries undertook to eliminate any differential treatment between the six Member States of the EEC and to reduce tariffs on their products, although they did allow some flexibility at this point to protect their domestic industry (Article 3 of the Convention). In 1969, a second Yaoundé Convention was signed extending these reciprocal preferences and three more African countries were included.

The situation evolved with the signing of the Lomé Convention in 1975. The UK's entry into the EEC led to this Convention being expanded to the Association of Caribbean and Pacific States, reaching a total of 46 members. The United Kingdom was forced to leave its Commonwealth model behind and adapt to the EEC Association model. However, countries such as Canada, Australia, New Zealand, and India were not included in the statute of association and were forced to seek other means of cooperation such as trade agreements adopted on the basis of Article 113 of the Treaty of Rome. As regards the system of preferences, the Lomé I Convention signed in 1975 preserved the tariff preferences of the Yaoundé Conventions but reciprocity was reduced. Products from the ACP States benefited from a reduction or elimination of tariffs in the EEC (articles 2 and 3 of the Convention), while European products only received the MFN treatment among the Associated States (article 7 of the Convention). These general features of the system of preferences were continued in the successive Lomé II Convention signed in 1979 and Lomé III Convention signed in 1984. Even after the EEC's Generalized System of Preferences (GSP) was implemented, the Associated States benefited from better preferences than other developing countries (Grilli 1994, p. 148).

Along with the agreements with the ACP countries, the EEC also signed trade and cooperation

agreements with other states. This group included, among others, agreements with the Maghreb, Masreq, and other countries from around the Mediterranean together with several Asian and Latin American countries, which also recognized a system of preferences for them, not as high as the ACP, but still prioritized them (Grilli 1994, p. 151). In these agreements, preferences were not subject to providing direct reciprocity but were significantly limited to protect the EEC's *sensitive* products or sectors.

The nonreciprocity of preferences as well as the abolition of special regimes or reverse preferences and their replacement by a GSP was repeatedly called for by several states at the Second United Nations Conference on Trade and Development (UNCTAD) held in New Delhi in 1968 (UNCTAD doc. TD/97, Vol. I, 1968). The USA demanded that states receiving these preferences should give them up if they wished to receive preferential treatment from them. Opposing this, the EEC, the United Kingdom, and the states benefitting from these special preferences insisted that the generalization of preferences should not harm them. However, all participants agreed to eliminate the need for reciprocity and turn preferences into an instrument for development cooperation which would benefit all developing countries and serve to increase exports and aid access to international trade relations.

The Generalization of Preferences

The signing of the GATT Agreement in 1947 established the application of the MFN clause in trade relations, thus prioritizing the principle of nondiscrimination between the participating states. Although the GATT was not unaware of the existence of special regimes which, as pointed out in the previous section, prioritized relations between some participating states, these were found to be compatible with the system. Along with this, following World War II, the idea spread that aid should be given by developed countries to developing countries. The United Nations Charter linked the need for cooperation among states to reduce economic and social differences in order to ensure international peace and security (Chapter IX of the UN Charter).

It was the developing countries that would eventually introduce into the debate the need to regulate international trade to address their specific needs. As early as 1961, India called for the establishment of general tariff preferences in the GATT. In a joint declaration, titled Appendix to Resolution 1897 (XVIII) of the United Nations General Assembly, 11 November 1963, several developing countries stated that "international trade could become a more powerful instrument and vehicle of economic development not only through the expansion of the traditional exports of the developing countries, but also through the development of markets for their new products." This was the aim of the First UNCTAD Conference held in 1964, which supported the idea of establishing a GSP to favor developing countries in order to increase and diversify their commercial relations (Final act, doc. E/CONF/46/139, 1964).

In the Charter of Algiers, approved at the Ministerial Meeting of the Group of 77 on 24 October 1967, urgent immediate measures were called for to encourage trade with developing countries and a series of principles was also passed for a general system of tariff preferences that "should provide for unrestricted and duty-free access to the markets of all the developed countries for all manufactures and semi-manufactures from all developing countries." This model was debated in the Second UNCTAD Conference which ended with the unanimous adoption of Resolution 21 (II), *on preferential or free entry of exports of manufactures and semimanufactures of developing countries to developed countries,* whereby an acceptable mutual and Generalized System of Preferences was agreed and implemented, with no reciprocity or discrimination which would be advantageous for developing countries. In sum, UNCTAD adopted the idea of a "'nondiscriminatory' system of discrimination in favor of all less developed countries" (Metzger 1967, p. 770). However, the developed countries' position on this issue varied widely.

Not all developed countries were willing to grant a GSP with the same conditions and features. As pointed out in the document addressed to UNCTAD by the OECD in 1969, "each donor country nonetheless reserves the right of taking

action appropriate to its own possibilities and taking account of certain features of its particular circumstances" (doc. OECD C(69)142). Thus developed countries believed that a GSP must be voluntary. Furthermore, donor countries failed to agree on the possibility of establishing quantitative restrictions, identifying beneficiary states, or limiting the types of products that may benefit from preferences. So, far from achieving the establishment of a single GSP for all developing countries, a proliferation of GSP arose.

The GSP approved by Australia in 1966 provided "non-reciprocal preferences on selected manufactures and semi-manufactures from developing countries up to the levels of specified quotas for each commodity." The scheme was aimed at less developed countries which could seek to benefit from certain preferential annual import quotas for certain products, unless they were already competitive by themselves (GATT Doc L/2627, April 4, 1966 on Australia). This first scheme which was adopted for 1 year has since continued to be extended and modified, successively amending fees and/or scope of preferences.

In 1971, the EEC approved its first GSP providing tariff preferences for developing countries. The preferences were limited to a series of regulations that established criteria to identify which products were regarded as originating in developing countries and therefore most likely to benefit from the preferences. It was a complex and seemingly rigid system which aimed to recognize substantial advantages for transformed agricultural products and semimanufactured or manufactured goods from developing countries while preserving the interests of the industries of the Member States and Associated States (Hoffmann, 1971 p. 650). The preferences were highly limited in the case of agricultural products, subject to quantitative restrictions, and mainly aimed at the Group of 77 Member States and Hong Kong. The EEC would also gradually modify this first scheme and has renewed its GSP ever since.

Throughout 1971 and early 1972, other developed countries implemented their respective GSP. However, just like the previous two, all included safeguard mechanisms and largely excluded products that might be of interest for developing countries, which significantly limited, at least in the short term, the effects of GSP on exports (Hoffmann 1971, p. 660). However, differential treatment included in the GSP contradicted the obligations stemming from the MFN under Article I of the GATT, and so it was necessary to obtain a waiver to implement it and provide continuity. From this perspective, the need soon arose for a provision that would generally uphold this special and differential treatment received by developing countries.

The continuity of differential treatment and tariff preferences for developing countries was considered from the outset of the Tokyo Round negotiations. The Enabling Clause was part of the resolutions adopted at the end of the negotiations in 1979 and aimed to accommodate differential and more favorable treatment for developing countries. This clause was based on the compatibility of "preferential tariff treatment accorded by developed contracting parties to products originating in developing countries in accordance with the Generalized System of Preferences" (Article 2 (a), GATT doc. L/4903). The problem was how to interpret the terms of the Enabling Clause in relation to the MFN clause and particularly, whether it allowed the interpretation made by developed countries as regards distinguishing developing countries.

There is no absolute answer to this question. For one, neither the GATT nor the WTO provides a definition or list of developing countries, while this group is becoming increasingly more diverse. Above and beyond the group of least developing countries (LDC) that are identified by the UN, or the orientations that may arise from the World Bank's classifications or the human development index, it is states that grant preferences that unilaterally decide which they regard as developing countries. Additionally, the Enabling Clause states that "any differential and more favorable treatment provided under this clause shall (. . .) be designed and, if necessary, modified, to respond positively to the development, financial and trade needs of developing countries" (article 3 (c) GATT doc. L4903). These terms may be interpreted as providing different treatment according to different needs which may be

consistent with the system. The different interpretations regarding this were highlighted by the WTO's Dispute Settlement Body (DSB) in the case lodged by India against the EU due to its GSP in 2002. As will be seen in the following section, India questioned whether the special arrangement to combat drug production and trafficking should not be open to any developing country which might benefit from the European Union's GSP. The Panel's report concluded that "the term non-discriminatory (. . .) requires that identical tariff preferences under the GSP be provided to all Developing Countries without differentiation" (Doc. WT/DS246/R, above no. 6). Later, the Appellate Body (AB) overruled this opinion stating that "the term 'developing countries' in paragraph 2(a) should not be read to mean 'all' Developing Countries" (WTO doc. WT/DS246/AB/R, above no 25). Differential treatment might be acceptable as it would be defined according to "financial and trade needs of developing countries."

In short, although GSP have indeed led to a generalization of preferences for the beneficiary states, they have not done so in an absolute sense, but rather, this has meant that potentially any state regarded as a developing country by countries granting the preferences can benefit from the same if it proves it meets the requirements of such a scheme. The Decision on Implementation-Related Issues and Concerns adopted in 2001 as part of the Doha Round states that "preferences should be non-discriminatory" but are not specifically required (WTO doc. WT/MIN (01)/17, 41 LL.M. 757 (2002)). This relative notion of a generalized treatment is exacerbated when features of conditions unrelated to trade are added to the granting of preferences.

The Conditionality of Preferences: A New Way to Reverse Preferences?

The Enabling Clause specifies that "the developed countries do not expect reciprocity for commitments made by them in trade negotiations to reduce or remove tariffs and other barriers to the trade of developing countries, i.e., developed countries do not expect the developing countries, in the course of trade negotiations, to make contributions which are inconsistent with their individual development, financial and trade needs" (article 5, GATT doc. L/4903). This section consolidates the nonreciprocity of preferences in relations between developed and developing countries. However, once again the reference made to the "development, financial and trade needs of developing countries" opens up the gate used by developed countries to condition the preferences granted to developing countries.

In its preamble, the Lomé IV Convention signed between the European Community (EC) and the ACP countries in 1990 emphasized the close relationship between promoting and respecting human rights and environmental protection for developing states. The aim of the association is to promote sustainable development and does not depend solely or primarily on an increase in trade relations nor, particularly, exports of these countries. With the review of the Convention in 1995, aid to the Associated States was linked to fulfilling the objectives of respecting human rights and the principles of good governance since noncompliance may justify withdrawal of aid or preferential treatment. The Cotonou Agreement, after 2000, has maintained this link of association with the aim of sustainable development and good governance. The erosion of preferences as a result of the reduction of general tariffs prioritizes measures of positive cooperation by consolidating the requirements of preferential trade treatment on these other obligations which are regarded as intrinsic to the effective development of countries.

In the same vein, in the early 1990s, the EC would also establish additional special tariff preferences under its GSP for products originating in the Andean and Central American states with the aim of aiding the fight against drug production and trafficking. Combating the production and trafficking of drugs was seen to be essential for the development of these countries, and in this sense, the special scheme seemed justified. Special incentive arrangements were introduced to protect labor rights and the environment a few years later, in 1994 and 1996. Again the EC granted additional preferences to developing

countries for adopting and implementing, on a domestic level, international environmental standards for agriculture, the substance of the standards laid down by the ITTO relating to the sustainable management of forests and the International Labor Organization Conventions numbers 87 and 98 concerning the application of the principles of the right to organize and to bargain collectively, and number 138 concerning the minimum age for admission to employment. In these cases, protecting the environment and recognizing labor rights were presented as "special development, financial and trade needs of developing countries."

In 2001, the Council Regulation 2501/2001, which approved the GSP scheme for the period 2002–2004, finally grouped into one single legislative instrument the provisions relative to the general preference scheme and four special regimes for least developed countries – the everything but arms initiative – to combat drug production and trafficking; for the protection of labor rights; and for the protection of the environment. The first two were aimed only at certain developing countries. This included the LDC, identified as such by the United Nations, whose particular situation is explicitly recognized in the Enabling Clause and the waiver agreed in 1999, renewed in 2009 (doc. WT/L/304, 1990 & WT/L/759, 2009). The regime to combat drug production and trafficking was only aimed at states which the EU included in their list of beneficiaries without giving objective criteria. However, the two other special arrangements are open to any developing country that already participates in the GSP and is shown to comply with the additional conditions imposed on them by such. According to the EU, all these special regimes respond to special development, financial and trade needs of developing countries, and are, therefore, based on the Enabling Clause.

Given this interpretation, India argued in its application to the WTO's DSB that the special arrangement to combat drug production and trafficking was discriminatory because the list of beneficiary states was unilaterally set by the EU without providing any criteria which, objectively, has led to some developing countries and not others to be included as beneficiaries of the EU's GSP. As mentioned above, the panel and the AB disagreed on this point yet, although the AB found that differential treatment for developing countries is, in principle, compatible with the Enabling Clause (since it must adapt to the specific needs of each country), it concluded that the EU's special drugs regime is contrary to it, as the criteria that led to defining the list of beneficiary states (WTO doc. WT/DS246/AB/R, above 25) were not justified. However, in this case no questions were raised over the conditionality of the preferences contained in special schemes, but how countries benefitting from the preferences in the special scheme were chosen, which meant these preferences would not be available to all developing countries that might have the same needs. In short, it is not discriminatory to provide differential treatment to a different situation, but it is if the situation is the same or very similar.

Following this decision, the EU revised its GSP and Regulation 980/2005 established the GSP plus, which included three regimes of preferences: a general arrangement, a special incentive arrangement for sustainable development and good governance, and a special arrangement for LDC (article 1 of the Regulation). Apart from the LDC scheme, the other special scheme addresses countries ratifying or undertaking to ratify the agreements specified in Annex III of the Regulation. This lists what the EU considers the major universal international treaties for the protection of human rights, labor rights, environmental protection, and those protecting principles of good governance. Countries that can prove they meet the conditions may request to benefit from additional preferences under the special regime and the European Commission must ensure they comply with the requirements. Regulations 732/2008 and 978/2012 maintain three distinct regimes, without ever questioning their compatibility with WTO rules. Against this, some authors argue that the choice of a closed list of international treaties, such as the sections in Annex III of the EU Regulation, is not flexible and cannot adapt to the needs and reality of each developing country (Bartels 2007, p. 884; Turksen 2009, p.965; Wardhaugh 2013, p 839).

USA's GSP shares the carrot-and-stick approach used by the EU and also establishes differences between developing countries submitted to complying with certain conditions or criteria. Instead of setting out special regimes as part of the GSP, the USA has established specific regional preference programs and has obtained waivers for the Caribbean Basin Economic Recovery Act, the Andean Trade Preference Act, and the African Growth and Opportunity Act. All these programs establish more benefits than the US GSP, but they are restricted to some specific countries and submitted to US foreign policy interests (Kennedy 2012). Besides the USA maintain their GSP. To identify the beneficiary developing countries in the GSP, the conditions are, for instance, that they may not be communist countries; they must respect labor rights and must wipe out child labor, cooperate in the fight against terrorism, respect intellectual property rights, and apply international arbitration decisions. These conditions are established and applied unilaterally and discretionally by the USA. For example, more than 20 years ago, the USA decided to reduce up to 50% the benefits accorded to Argentina through the GSP because they had estimated that national legislation had not granted enough protection to intellectual property rights (Lavopa and Dalle 2012).

Conditionality may seem like a form of reciprocity as it requires a certain action to be taken by developing countries to benefit from a preference (Kishore 2011, p. 895). Subjecting preferences to respecting certain human rights, protecting the environment or the democratic principles of government introduces issues as factors of development that are not directly related to trade. In the Doha Round, developing countries were able to keep these issues out of the negotiations. The unilateral basis which defined the GSP and other preferential trade relations has allowed developed countries to reintroduce them into the debate by conditioning the preferences. Indeed, it is hard to deny that these issues are key to the development of a country, but the way that these are presented seem to respond more to the needs of developed countries than developing countries.

Erosion of Preferences, but Not Their Disappearance

The impact of preferences on developing countries has differed depending on whether they have been channeled through special regional agreements, such as the Association Agreements between the EU and the ACP States or the US Caribbean Basin Initiative, or by establishing a GSP. Developing countries participating in regional preferential agreements enjoy a special and privileged status compared to the rest. The concessions under these agreements are larger, tariffs are reduced or even eliminated, and the product range is wider. On this point the GSP face three problems. First, they are also of a unilaterally voluntary nature which makes them respond more to considerations imposed by the donor states than the most urgent needs of the beneficiary countries. Second, they are open to a larger group of states and are therefore more diverse, and not every country can always benefit from the same levels of preferences. Third, implementing GSP preferences involves a difficult process which leads to an increase in costs, sometimes making it unattractive to the potential beneficiary. It is up to the developing countries to apply for these preferences and demonstrate that their products comply with the regulations of origin set out in GSP, which may involve additional costs and reduce the final benefit. From this perspective, many authors criticize the results achieved by the GSP.

The general reduction of tariffs in international trade has also progressively limited the role of tariff preferences as an instrument for development cooperation. The issue of preference erosion has been widely debated in by authors for years. In regional preferential agreements, this loss has been offset by other means of cooperation, such as financial or technical assistance. But the GSP are the result of a different logic and have also been stretched by the existence and proliferation of other preferential trade agreements which provide greater preferences than those general schemes and due to the loss of margin for preferential treatment as a result of successive negotiation rounds.

Furthermore, the temporary nature that characterizes the special and differential treatment for

developing countries in the WTO creates some uncertainty. It seeks to promote development so that once the goal is achieved *exceptions* to the principle of nondiscrimination and the MFN governing multilateral trade relations will become meaningless. This feeling of uncertainty is greater in the case of a strictly unilateral GSP. A GSP depends on the willingness of the states to establish such, and they may eventually decide not to continue it. However, it is worth noting that the main GSP still persist today. The EU, the USA, and Russia have renewed their GSP in 2013; Australia did so in 2012; and Japan in 2010. But, for example, the EU has severely reduced the list of beneficiary states excluding from 2015 those states that have been classified by the World Bank as upper-middle-income countries since 2011 (European Commission delegated Regulation 1421/2013). Ways remain to be explored that can empower the role of tariff preferences as tools for development. Agreements that facilitate the management of trade in goods subject to tariff preferences would represent a qualitative edge in response to an open, general, debate in the WTO (Bartels and Häberli 2010; Persson 2012).

Before concluding it should be recalled that a more general definition of the term "preference" means any treatment that is more beneficial for a state. Other agreements also provide preferences with this broader meaning that have been of great interest to the beneficiaries. Of special interest is the decision referring to preferential treatment to services and service suppliers of LDC, 2011, which allows members to grant "preferential treatment to services and service suppliers of Least Developed Countries without according the same treatment to like services and service suppliers of all other Members" (doc. WT/L/847). The terms of this document are quite similar to those of the Enabling Clause, because preferential treatment is justified in order to increase the participation of LDCs in trade in services, which are expected to become generalized preferences for all LDC, but of a unilateral nature from the perspective of the donor state, and set temporarily.

Nevertheless, there is an increasing diversity of states in different levels of social and economic development. Today it is difficult to maintain the division between developed and developing countries as clearly separate and stable groups in different areas of international cooperation (Pauwelyn 2013). This is reflected by developing countries acting less in common, which may weaken their ability to negotiate, but also that of developed countries, some of which face serious domestic problems with a much lower growth rate than in other so-called developing countries. A system based on a generalization of preferences does not adapt well to this reality than more than ever requires ample flexibility to adapt to the "development, financial and trade needs" of the beneficiary states.

References

Monographs

Grilli ER (1994) The European community and the developing countries. Cambridge University Press, Cambridge

Journal Articles

Bartels L (2007) The WTO legality of the EU's GSP+ arrangement. J Int Econ Law 10(4):869–886

Bartels L, Häberli C (2010) Binding tariff preferences for developing countries under Article II GATT. J Int Econ Law 13(4):969–995

Clute RE, Wilson RR (1958) The commonwealth and the favored-nation usage. Am J Int Law 52: 455–468

Feis H (1945–1946) The future of British imperial preferences. Foreign Aff 661: 661–674

Kennedy KC (2012) The generalized system of preferences after four decades: conditionality and the shrinking margin or preference. Mich State Int Law Rev 20:521–668

Kishore P (2011) Conditionalities in the generalized system of preferences as instruments of global economic governance. Int Lawyer 45:895–902

Lavopa F, Dalle D (2012) "¿Hay vida después del SGP? Implicancias de la posible exclusión de Argentina de los sistemas generalizados de preferencias de Estados Unidos y la Unión Europea", Cátedra OMC FLACSO Argentina, August 2012. Available at: http://catedraomc.flacso.org.ar/wp-content/uploads/2012/08/Anexo-1_SGP.pdf

Persson M (2012) From trade preferences to trade facilitation: taking stock of the issues. Econ Open Access-Open Assess E-J 6:1–33

Turksen U (2009) The WTO law and the EC's GSP+ arrangement. J World Trade 43(5):927–968
Wardhaugh B (2013) GSP+ and human rights: is the EU's approach the right one? J Int Econ Law 16:827–846

Further Reading

Monographs

Hoekman B, Özden Ç (2006) Trade preferences and differential treatment of developing countries. Edward Elgar Publishing Limited, Northampton

Journal Articles

Bartels L (2003) The WTO enabling clause and positive conditionality in the European community's GSP program. J Int Econ Law 6(2):507–532
Breda Dos Santos N, Farias R, Cunha R (2005) Generalized system of preferences in General Agreement on Tariffs and Trade/World Trade Organization: history and current issues. J World Trade 39(4): 637–670
Charnovitz S, Bartels L, Howse R, Bradley J, Pauwelyn J, Regan D (2004) The Appellate Body's GSP decision. World Trade Rev 3(2):239–265
Fernandez-Pons X, Lavopa F (2014) ¿Ojo por ojo, diente por diente? Trazando los límites de las contramedidas comerciales en el marco del Derecho de la OMC. In: Linares D, Luciano M. (Coord.): Controversias en la Organización Mundial del Comercio. Protagonismo y estrategias de los países en desarrollo. Editorial FLACSO Argentina – Teseo, Buenos Aires, pp 233–299
Grossman GM, Sykes AO (2005) A preference for development: the law and economics of GSP. World Trade Rev 4(1):41–67
Hofmann G (1971) Les préférences tarifaires généralisées. Cah de Dr Eur 6:641–661
Howse R (2003) India's WTO challenge to drug enforcement conditions in the European community generalized system of preferences: a little known case with major repercussions for 'Political' conditionality in US trade policy. Chic J Int Law 4:385
Huici-Sancho L (2006) What kind of 'Generalized' systems of preferences? Eur J Law Econ 21: 267–283
Inama S (2011) The reform of the EC GSP rules of origin: per aspera ad astra? J World Trade 45(3):577–603
Inama S (2003) Trade preferences and the WTO negotiations on market access – battling for compensation of erosion of GSP, ACP and other trade preferences or assessing and improving their utilization and value by addressing rules of origin and graduation? J World Trade 37(5):959–976
Irish M (2007) GSP tariffs and conditionality: a comment on EC–Preferences. J World Trade 41(4):683–698
Khorana S, Yeung MT, Kerr WA, Perdikis N (2012) The battle over the EU's proposed humanitarian trade preferences for Pakistan: a case study in multifaceted protectionism. J World Trade 46(1):33–59
López-Jurado C (2011) La oferta comercial preferencial de la Union europea a los países en vías de desarrollo: modalidades e interacciones. Revista de Derecho Comunitario Europeo 39:443–483
Manero-Salvador A (2007) El SGP+ como elemento de la cooperación entre Europa y América Latina. Revista Electrónica Iberoamericana 1(1). Available at: http://www.urjc.es/ceib/investigacion/publicaciones/REIB_01_A_Manero_Salvador.pdf
Metzger SD (1967) UNCTAD, developments in the law and institutions of international economic relations. Am J Int Law 61:756–775
Ochieng C, Milton O (2007) The EU–ACP economic partnership agreements and the 'Development Question': constraints and opportunities posed by article XXIV and special and differential treatment provisions of the WTO. J Int Econ Law 10(2):363–395
Pauwelyn J (2013) The end of differential treatment for developing countries? Lessons from the trade and climate change regimes. Rev Eur Int Environ Law 22:29–41
Shaffer G, Apea Y (2005) Institutional choice in the generalized system of preferences case: who decides the conditions for trade preferences? The law and politics of rights. J World Trade 39(6):977–1008
Yusuf AA (1980) Differential and more favourable treatment: the GATT enabling clause. J World Trade Law 14(6):488–507

Prisoner's Dilemma

Wolfgang Buchholz and Michael Eichenseer
University of Regensburg, Regensburg, Germany

Abstract

The Prisoner's Dilemma (PD) is probably the most famous two-person game in which a fundamental divergence between individual and collective rationalities arises: If the agents play noncooperatively, an equilibrium is achieved which, however, does not constitute the best available solution. Such a PD situation characterizes many situations of voluntary cooperation, e.g., the provision of the global public good climate protection. But in reality agents are – despite the predictions of the PD game – often willing to cooperate voluntarily to some degree which has been confirmed by experimental economics. Furthermore there are a lot of institutional devices which help overcome the cooperation dilemma in a PD situation.

The Standard Prisoner Story

The Prisoner's Dilemma (PD) is probably one of the most famous games analyzed in game theory. Its name refers to a rather artificial story conceived by the American mathematician Albert Tucker in order to convey the structure of the PD game to a broader public: Two prisoners – A(ndrea) and B(ritney) – who have jointly committed crimes are interrogated by the police separately without any chance of communicating with each other. Actually the police cannot take enough evidence to convict them of a serious crime (what the police clearly would like to do) but only of a smaller crime which would send both A and B to prison for 1 year. Without further provisions neither A nor B will have an incentive to tell the truth about the serious crime. Anticipating this discretion, the police offer a principal witness regulation: If one of the two prisoners betrays the other, she will be set free immediately while the thus convicted perpetrator will have to go to jail for 3 years. This regulation, however, only applies if just one of the two prisoners confesses and sends her accomplice to the doom. Otherwise, i.e., when both prisoners give evidence, both A and B will have to serve 2 years.

Even though this story has some drawbacks (e.g., as it is not obvious why in the case of a unilateral confession the betrayed prisoner should get a harder punishment), it highlights the cooperation dilemma that occurs in a PD situation: When acting in complete isolation from one another, the best each prisoner can do is to admit to a more serious crime – independent of what she expects her co-perpetrator to do. Such a behavior follows from a rational calculation which runs as follows: If my counterpart confesses, I will be imprisoned for 3 years if I remain silent but only for 2 years if I confess, too. Likewise, if the other one remains silent, my best strategy is again to confess because in this case I can benefit from the principal witness regulation and avoid staying in jail any longer.

In technical terms this means that in a PD, confessing is the dominant strategy for player A as well as for player B. Thus the game's outcome will be that each prisoner confesses and goes to jail for 2 years. This, however, does not seem to be a satisfactory solution for either of them: If both remained silent instead, they could improve their situation since this would save each of them 1 year in prison. However, the problem is that the personal incentives which are guiding the agents' behavior when they act independently and in complete isolation will prevent the attainment of the more favorable outcome. Thus for players A and B, a dilemma between individual and collective rationalities arises which constitutes the distinctive feature of the PD game.

In game theory the incentive structure of the PD game and the determination of the equilibrium solution are usually described by the normal form representation as depicted in Figure 1. Here the numbers 1, 2, 3, and 4 indicate the ranking of each player's personal utilities (or "payoffs") depending on the four possible strategy pairs chosen by A and B: Player A's rank number appears in the upper and B's rank number in the lower part of each cell of the matrix. For example, number 4 in the lower left cell indicates that A reaches her most favorable position if she confesses while B does not. The stars behind the numbers indicate the optimal (Nash) reactions of each player which are obtained for player A by comparing her rank numbers in each row and, analogously, for player B by comparing her rank numbers in each column. The unique Nash equilibrium (confess, confess) of the PD game is situated in the upper left cell of the normal form where the stars of A and B coincide, i.e., where the optimal reactions of both players are consistent. A comparison between the upper left and the lower right cell (don't confess, don't confess) shows the inferiority of the noncooperative outcome relative to the hypothetical cooperative solution (Fig. 1).

Public Good Provision

The incentive structure underlying the PD game characterizes many situations of mutual cooperation between two or more agents. A very prominent example for such a cooperation problem is the voluntary provision of a public good which – particularly due to the perception of

Prisoner's Dilemma, Fig. 1 Normal form payoff matrix for the PD

A / B	Confess	Don't confess
Confess	2* / 2*	1 / 4*
Don't confess	4* / 1	3 / 3

climate protection as a global public good – has received a lot of attention over the past few years. How a PD situation arises in the case of public good provision is even easier to explain than with Tucker's standard story about the two prisoners.

For the public good scenario, we assume that the players A and B are two countries that share the same environmental medium (water or air) and thus suffer from the same environmental damage. Each country has two options, either to take an abatement measure to curb its pollution which entails costs (c) for this country or not to abate and to carry on as usual. If none of the countries abate, no environmental damage is avoided and the environmental benefit for each country is $B(0) = 0$. If instead one single country abates, both countries will enjoy the environmental benefit $B(1) > 0$, as there are spillovers of pollution between the two countries. In other words, there is neither rivalry in nor excludability from the benefits of an abatement measure, i.e., environmental quality has the well-known properties of a public good. If not only one but both countries make an abatement effort, then each country gets an environmental benefit equal to $B(2) > B(1)$.

Concerning the relation between environmental benefits and abatement costs, it is now supposed that:

- $B(2) - c > 0$ or equivalently $B(2) > c$ which means that collective abatement efforts are beneficial for both countries.
- $B(2) - B(1) < c$ and $B(1) < c$ which says that making an own abatement effort is never worthwhile for a country irrespective of whether the other country does abate or not.

Given these conditions we obtain the same payoff structure as depicted in Figure 1 by identifying the strategy "abatement" (or "cooperation") with "not confess" and "nonabatement" (or "noncooperation") with "confess" in the prisoners' story. Voluntary public good contributions made by payoff-maximizing countries noncooperatively lead to an undersupply of the public good, i.e., to insufficient abatement activity. At the same time, there are lower net benefits for both countries compared to the cooperative outcome with collective action on public good provision. The best solution for each country is to be a free rider on the abatement activities of the other country.

In the two-country case, it is well plausible that game types which are different from PD may arise (see Osborne 2009). So if $B(1) > c$ is assumed while $B(2) - B(1) < c$ still holds, a chicken game with two asymmetric Nash equilibria results which are characterized by unilateral abatement measures of just one country. In this case unilateral abatement pays for each country (e.g., to avoid an environmental catastrophe) but not additional abatement when abatement has already been done by the other country.

Notwithstanding that, the two-agent PD serves as the prototype for a more general public goods game with many agents, where each agent affects the total public good supply not much through his own costly contributions but benefits considerably from the contributions of the many other agents. This extension to many agents further emphasizes the empirical relevance of the PD game.

Ways Out of the Dilemma

Given the unsatisfactory outcome of a noncooperatively played PD game for its participants, naturally the question arises how to implement the cooperative solution and thus to overcome the cooperation dilemma. In order to achieve such a Pareto improvement over the noncooperative Nash equilibrium, several approaches appear to be viable.

Agents can take up negotiations and enter into an agreement in which they commit to performing the cooperative action, i.e., remaining silent or abating as in the two examples above. The difficulty of this approach is, however, that the cooperation problem may only be shifted from the level of participation to the level of internal stability of a cooperative arrangement. This is due to the fact that in a PD each agent will not want to keep her promise and implement the cooperative action after the agreement has been concluded. Defection incentives may be avoided and stability may be ensured by introducing sanctioning mechanisms. At the national level the government punishes tax evaders who do not pay their share in the cost of public goods. In line with that, PD-like scenarios already have played a major role in the traditional contract theories of the state as developed by the philosophers Th. Hobbes and J. Locke centuries ago.

However, if no central authority with sanctioning power exists (as it is typically the case for global public goods), other forms of punishment may help to solve the dilemma and to bring about mutual cooperation. If the PD is not a one-shot game but rather iterated over many periods, it becomes possible for an agent to punish her partner's noncooperation in an earlier period by refusing to cooperate in subsequent periods. The largest part of the many punishment strategies that have been conceived by game theorists are variations of Rapoport and Chammah's (1965) simple tit-for-tat TFT strategy where each player directly reciprocates by repeating the other player's preceding action. Refinements of TFT mainly differ with respect to the other player's history of actions which trigger punishment, the length of the punishment period, and the integration of random elements in the punishment strategy. Beginning with Axelrod (1984), the success of these strategies has been tested by the use of computer tournaments. In the meantime, new strategies based on recognition and learning ("Pavlov") or gradual punishment mechanisms have proved to be even more successful in maximizing individual payoffs (Kendall et al. 2007). Note, however, that given strictly rational agents, punishment strategies only work as desired if the PD game is (at least potentially) repeated over infinitely many periods and future payoffs are not too heavily discounted.

In real-life situations punishment may also occur outside the PD by some kind of "issue linkage" which, e.g., could mean that cooperation on the provision of international public goods is fostered by the threat of trade sanctions. Moreover, if formal governmental institutions are absent, people may organize themselves and establish rules and mechanisms for an improved public good provision. In the context of local and regional commons (as fisheries or grazing lands) as specific types of public goods, the Nobel laureate E. Ostrom (1990) has demonstrated the efficiency of such self-organized governance systems as well as their potential limitations in a lot of field studies.

The basic assumptions underlying most theoretical treatments of PD are complete individual rationality and material self-interest as the players' sole motivation. These requirements are clearly not satisfied for real people in many situations (see Fehr and Schmidt 2006 for a review). Thus it does not come as a surprise that a lot of experimental studies have shown that noncooperation is not the standard outcome: In a one-shot public goods game subjects usually contribute between 40% and 60% of their endowment (Ostrom 2000), and cooperation rates in one-shot as well as in repeated PD games turn out to be significantly higher than zero ranging between 5% and 96.9% with an average of about 45% (Sally 1995; Jones 2008). The experimental setting, however, has a large impact on cooperation rates: Framing a PD as a "Stock Market Game",

for example, reduces cooperation rates while cooperation is encouraged by calling it a "Community Game" (Ellingsen et al. 2012). Moreover, individual behavior in PD games seems to be influenced by gender (Croson and Gneezy 2009) and the cultural and ethnical background of the players (Cox et al. 1991).

Recently, Khadjavi and Lange (2013) have provided evidence on behavior in a PD game played by a subject group of female inmates who after all did cooperate in 55% of all cases in the one-shot PD scenario. Ultimately the PD has thus returned to the place where it virtually started: The prison.

References

Axelrod R (1984) The evolution of cooperation. Basic Books, New York

Cox TH, Lobel S, McLeod PL (1991) Effects of ethic group cultural differences on cooperative and competitive behavior on a group task. Acad Manage J 34: 827–847

Croson R, Gneezy U (2009) Gender differences in preferences. J Econ Lit 47:1–27

Ellingsen T, Johannesson M, Mollerstrom J, Munkhammar S (2012) Social framing effects: preferences or beliefs? Games Econ Behav 76:117–130

Fehr E, Schmidt KM (2006) The economics of fairness, reciprocity and altruism – Experimental evidence and new theories. In: Kolm SC (ed) Handbook of the economics of giving, altruism and reciprocity. Elsevier, Amsterdam, pp 615–691

Jones G (2008) Are smarter groups more cooperative? Evidence from prisoner's dilemma experiments, 1959–2003. J Econ Behav Organ 68:489–497

Kendall G, Yao X, Chong SY (2007) The iterated prisoner's dilemma: 20 years on. Advances in natural computation. World Scientific Pub Co, Singapore

Khadjavi M, Lange A (2013) Prisoners and their dilemma. J Econ Behav Organ 92:163–175

Osborne MJ (2009) An introduction to game theory. Oxford University Press, New York

Ostrom E (1990) Governing the commons: the evolution of institutions for collective action. Cambridge University Press, Cambridge

Ostrom E (2000) Collective action and the evolution of social norms. Econ Perspect 14:137–158

Rapoport A, Chammah AM (1965) Prisoner's dilemma: a study of conflict and cooperation. Michigan University Press, Ann Arbor

Sally D (1995) Conversation and cooperation in social dilemmas: a meta-analysis of experiments from 1958 to 1992. Ration Soc 7:58–92

Prisons

Paolo Buonanno[1] and Juan F. Vargas[2]
[1]Department of Management, Economics and Quantitative Methods, University of Bergamo, Bergamo, Italy
[2]Department of Economics, Universidad del Rosario, Bogota, Colombia

Synonyms

Detention; Imprisonment; Incarceration

Definition

Prison (or incarceration) is one of the most important forms of sanction in many modern criminal system. Incarceration may impact the overall level of crime and affect criminal behavior through several channels: incapacitation, deterrence (general deterrence and specific deterrence), rehabilitation. Understanding whether individuals respond to any of the effects outlined above is a crucial aspect for the design of effective crime reduction policies.

Introduction

The most important form of criminal sanction in many modern criminal justice systems is incarceration. The utilitarian theory of punishment developed by Cesare Beccaria (1764) and Jeremy Bentham (1789) defines the extent to which incarceration or imprisonment may be considered beneficial for society as a whole. The utilitarian rationale justifies any form of punishment, under the condition that the associated benefits are higher than the implied costs. In particular, if punishment deters or incapacitates criminals, or facilitates their rehabilitation – thus bringing benefits to most individuals of a society – then punishment is socially "good," and therefore justifiable.

From a theoretical perspective, prisons have at least five social functions. First, incarceration mechanically incapacitates criminally active individuals from committing other crimes during the length of conviction. This is the **incapacitation effect**. Second, other things equal the threat of imprisonment could deter potential criminals from offending. Indeed, in the economic model of crime, proposed by Becker (1968), potential offenders consider the expected costs of committing a crime, which are related to the probability of being captured and condemned, and to the severity of the punishment. This is the **general deterrence effect**. In turn, the **specific deterrence effect** appears when formerly convicted individuals behave in such way to avoid returning to prison. A fourth social function of prisons arises insofar as these types of facilities could potentially help rehabilitating criminals, thus favoring their social reintegration and preventing them from recommitting crimes. **This is the rehabilitation effect**. On the other hand, however, serving time in a prison may increase the criminal propensity of inmates, for instance, by enhancing their criminal network. This is **criminogenic effect**.

Understanding whether individuals respond to any of the effects outlined above is a crucial aspect for the design of effective crime reduction policies. Over the last decade or so, the economics literature has made significant progress in studying the empirical relevance of each of the mechanisms through which prisons can help reducing crime. With a few exceptions, most of this literature has focused on the US case, thus limiting somehow the external validity of its findings.

Incapacitation

The incapacitation effect of imprisonment is extremely challenging to estimate. The earlier criminology literature based its estimates on inmate interviews. Prisoners are asked about their criminal activity prior to their last arrest, and researchers use their responses to extrapolate the counterfactual criminal behavior in the absence of imprisonment (Marvell and Moody 1994, review this literature). The resulting estimates, that show a large variance and thus provide little useful information, have been criticized (see Miles and Ludwig 2007) for their reliance on self-reported offenses and their sensitivity to outliers (prisoners who report an unrealistically large criminal record).

A more credible approach is the one followed by Owens (2009), who exploits a natural experiment that took place in the state of Maryland in 2001. Following a law change, judges stopped considering juvenile criminal records when sentencing adult offenders, so actual sentences went down. The author then examines the crimes committed by released offenders during the time they would have otherwise been incarcerated had the law not changed. The estimated incapacitation effects are much smaller (about one fifth) than the average effect from the research based on inmates' self-reported past offences. A critique of this strategy is, however, that it hardly distinguished between incapacitation and rehabilitation or specific deterrence.

Natural experiments of this sort are, however, less than common. A different empirical approach correlates aggregate crime and imprisonment data, usually exploiting state-level variation for the USA (see Marvell and Moody 1994). This approach has, however, two main limitations. On the one hand, estimates aggregate the incapacitation and the general deterrence effects. On the other, it is challenging to find causal effects at high levels of aggregation. Again, natural experiments can offer solutions. Levitt (1996), for instance, takes advantage of the fact that state-level prison overcrowding lawsuits exogenously reduce incarceration rates. Using an instrumental variables strategy, the author finds that a 1% increase in prison time reduces state-level violent crimes by 0.4%. While this approach solves the latter critique, it cannot disentangle the imprisonment and the general deterrence effects. Other instrumental variables estimates of the joint incapacitation/deterrence effects are Johnson and Raphael (2012) for the case of USA and Barbarino and Mastrobuoni (2014) for the case of Italy.

An interesting theoretical insight of the incapacitation effect is that, as the incarceration rate increases, the average incapacitation decreases.

This is because everything else equal higher incarceration rates imply that the marginal criminal is less dangerous. Liedka et al. (2006) for the case of the USA, Vollaard (2013) for the Dutch case, and Buonanno and Raphael (2013) for the case of Italy find evidence in favor of this hypothesis.

General Deterrence

General deterrence is a function of both prison length and prison conditions (Drago and Galbiati 2012). Regarding sentence length, the available evidence suggests that the severity of punishment does deter potential criminals from engaging in illegal behavior, as predicted by the standard economic model of crime. However, as mentioned above, the standard empirical approach – that compares incarceration rates to aggregate (state-level) crime – tends to confound the effect of incapacitation.

Again, natural experiments can come in handy. Kessler and Levitt (1999), for instance, exploit the exogenous variation in prison length provided by the so-called three strikes laws in the USA. The authors use California's Proposition 8 to show that, consistent with an aggregate deterrence effect, sentence enhancements of persistent offenders cause a 4% aggregate crime reduction immediately after the passage of the law. Likewise, exploiting the exogenous variation that the 2006 Italian Collective Clemency Bill created on the expected future punishment of released offenders if recommitting a crime, Drago et al. (2009) find that larger expected punishments dissuade recidivism.

Prison conditions also factor in the expected punishment and thus have a deterrent role. Using inmates' death rates as a proxy of (bad) conditions, Katz et al. (2003) corroborate that worse facilities are associated with less crime in the USA. The policy implication of this evidence is, however, not clear. On the one hand, prison conditions are more readily manipulatable than sentences length. On the other, it is hard to argue that facilities should be made harsher for crime reduction. Moreover, the size effects obtained by Katz et al. (2003) are rather small, so the cost-benefit balance of manipulating prison conditions is unclear. It is perhaps more promising to study the specific deterrent effect of prisons for individual who have spent jail time.

Specific Deterrence and Rehabilitation

Prisons vary substantially in terms of the quality of facilities, the services and activities offered, and the average physical space on inmates. All of these aspects potentially affect a released individual's propensity to recidivate. However, the myriad of factors that condition inmates' future behavior, and the difficulty to measure them in a systematic and comparable fashion, make it difficult to empirically investigate specific deterrence and rehabilitation mechanisms of (future) crime prevention. The available evidence is, therefore, scarce and it has limited external validity. Moreover, in the absence of individual-level data, it is hard to empirically separate the specific deterrence effect of prisons from the general deterrence effect.

Recent evidence from Italy (Drago et al. 2011) and Colombia (Tobón 2017) suggest that worse prison conditions (respectively higher mortality and higher overcrowding) increase the likelihood that former prisoners reengage in criminal behavior. In turn, evidence from the USA (Kuziemko 2007) and France (Maurin and Ouss 2009) suggests that experiencing longer sentences reduce recidivism.

Criminogenic Effect

While the evidence above points toward a positive (crime-reducing) social function of prisons, prisons can also breed crime. Indeed, inmates not only interact with a prison's facilities and rehabilitation opportunities but also – and primarily – with other inmates. From their peers, inmates can access information (about criminal markets and illegal lucrative opportunities), be exposed to a specific set of values and social norms, and get acquainted with people with similar preferences and abilities while finding

potential complementarities that reduce the costs of committing crimes.

Understanding the scope of prison-specific peer effects is crucial for the design of policies aiming at improving the effectiveness of the prison system. For instance, there is evidence that prison networks persist after inmates are released. Specifically, using data from the 2006 Italian Collective Clemency Bill, Drago and Galbiati (2012) find that sharing a prison facility with inmates of one's same nationality predicts a correlated former prisoners' post release criminal behavior.

Also, peer effects are stronger for younger individuals, the brain of whom is more malleable, and thus more vulnerable to stimuli from the environment (Giedd 2004). Consistent with this observation, Bayer et al. (2009) estimate large peer effects on subsequent criminal behavior for juvenile offenders of the same correctional facility in Florida.

To conclude, there are several mechanisms through which prisons can influence society's crime rates. While challenging, empirically distinguishing the contribution of each one of them to reduce (or exacerbate) crime, is of foremost importance for public policy.

References

Barbarino A, Mastrobuoni G (2014) The incapacitation effect of incarceration: evidence from several Italian collective pardons. Am Econ J Econ Pol 6(1):1–37

Bayer P, Hjalmarsson R, Pozen D (2009) Building criminal capital behind bars: Peer effects in juvenile corrections. Q J Econ 124(1):105–147

Beccaria C (1764) On crime and punishments

Becker GS (1968) Crime and punishment: an economic approach. J Polit Econ 76(2):169–217

Bentham J (1789) An introduction to the principles of morals and legislation

Buonanno P, Raphael S (2013) Incarceration and incapacitation: evidence from the 2006 Italian collective pardon. Am Econ Rev 103(6):2437–2465

Drago F, Galbiati R (2012) Indirect effects of a policy altering criminal behavior: evidence from the Italian prison experiment. Am Econ J Appl Econ 4(2): 199–218

Drago F, Galbiati R, Vertova P (2009) The deterrent effects of prison: evidence from a natural experiment. J Polit Econ 117(2):257–280

Drago F, Galbiati R, Vertova P (2011) Prison conditions and recidivism. Am Law Econ Rev 13(1):103–130

Giedd JN (2004) Structural magnetic resonance imaging of the adolescent brain. Ann N Y Acad Sci 1021(1): 77–85

Johnson R, Raphael S (2012) How much crime reduction does the marginal prisoner buy? J Law Econ 55(2):275–310

Katz L, Levitt SD, Shustorovich E (2003) Prison conditions, capital punishment, and deterrence. Am Law Econ Rev 5:318–343

Kessler D, Levitt SD (1999) Using sentence enhancements to distinguish between deterrence and incapacitation. J Law Econ 42(1):343–364

Kuziemko I (2007) Going Off Parole: how the elimination of discretionary prison release affects the social cost of crime. Working paper no 13 380 (September), NBER, Cambridge, MA

Levitt, SD. (1996). The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation. Q J Econ, 111(2), pp. 319–51

Liedka RV, Piehl AM, Useem B (2006) The crime-control effect of incarceration: does scale matter? Criminol Public Policy 5:245–276

Marvell TB, Moody CE Jr (1994) Prison population growth and crime reduction. J Quant Criminol 10(2):109–140

Maurin E, Ouss A (2009) Sentence reductions and recidivism: lessons from the Bastille Day quasi experiment. IZA discussion paper no 3 990 (February), Inst. Study Labor, Bonn

Miles TJ, Ludwig J (2007) The silence of the lambdas: deterring incapacitation research. J Quant Criminol 23(4):287–301

Owens EG (2009) More time, less crime? Estimating the incapacitative effect of sentence enhancements. J Law Econ 52(3):551–579

Tobón S (2017) Prison conditions, human capital and recidivism: evidence from an expansion in prison capacity, documento CEDE Universidad de Los Andes, no 2017–7

Vollaard B (2013) Preventing crime through selective incapacitation. Econ J 123(567):262–284

Privacy

Ignacio N. Cofone
NYU Information Law Institute,
New York, NY, USA

Synonyms

Data privacy; Information privacy

Definition

There is no universal agreement on what is privacy. Within law and economics, privacy has been modeled as concealment of personal information (Posner 1978, 1981) and as the standard deviation in the probability distribution forming people's perception over our personal information, with privacy loss being a tighter posterior in that distribution (Cofone and Robertson 2018a).

Introduction

The law and economics literature on information privacy revolves around two dialogues, one normative and one empirical. To a large extent, these dialogues have remained separate. The normative dialogue asks whether privacy is worth protecting, to what extent, and how. The empirical dialogue focuses on better understanding consumer behavior through experimental methods.

Normative Privacy

The puzzling aspect of the normative debate is that while much of the normative economic literature argues categorically against privacy protection, most of the legal literature does not address these arguments and moves directly to second order questions of degree and manner: to what extent privacy should be protected and how.

For Posner (1978, 1981), privacy creates an information asymmetry that disadvantages the buyers in the market (e.g., employers in job interviews), thereby re-distributing wealth and creating inefficiencies. There are individuals with good and bad traits, and the individuals with bad traits want to hide them (privacy), while the individuals with good traits want to show them. Privacy allows individual to hide their traits to deceive others and thus reduces the information with which the market allocates resources. Stigler (1980) adds that people who want information about someone protected by privacy will use other less precise (and usually less intrusive and costlier) substitutes as a proxy for the data they want to acquire, so privacy law is at best ineffective. This is similar to an application of the Coase theorem: as long as transaction costs remain low, whether there is a privacy rule (which allocates the right over information to the person to whom that information refers) or there is no privacy rule (which allocates information to whoever finds it) is irrelevant for the information's final allocation. For that reason, under a no-transaction-cost condition, introducing disturbances in the market such as information asymmetries, which would be welfare decreasing, is unjustified. Later research takes a similar approach for information in the financial market, arguing that allowing firms to collect and sell consumer data leads to increased overall welfare (Kahn et al. 2005; Kim and Wagman 2015).

Some resistance to the idea that privacy is necessarily inefficient is later introduced. Introducing a model where individuals suffer reputational harm from the loss of privacy, Daughety and Reinganum (2010) show that privacy protection allows people to engage in an optimal level of desired activities. Cofone (2017) shows that the conclusion that privacy is always inefficient holds only in a static world, but not in a dynamic world where information production is costly. Gradwohl and Smorodinsky (2017) show how privacy concerns affect people's choice of actions in strategic settings.

In the legal literature, Solove (2006) introduces a taxonomy of six different conceptions of privacy: (i) as the right to be left alone, (ii) as autonomy, (iii) as secrecy or concealment of discreditable information, (iv) as control over one's personal information, (v) as personhood and preservation of one's dignity, and (vi) as intimacy. The two conceptions that use law and economics arguments to approach privacy normatively have been secrecy (the approach taken by Posner and Stigler) and control. The latter mostly involves a discussion on whether a property right protection over personal data is warranted. While some suggest a protection regime similar to that of property (Murphy 1995; Schwartz 1999, 2004), others are skeptical of the idea. They argue that the reasons to protect property or intellectual property and the reasons to protect privacy are different (Litman

2000; Samuelson 2000) and that the debate around privacy should focus not on property but on autonomy (Cohen 2000).

More recent literature addresses privacy harms. Calo (2011) distinguishes two different dimensions of privacy harms: intrinsic privacy harms, which relate to privacy in itself, and extrinsic harms, which relate to material damage that is independent of privacy, such as financial damage resulting from identity theft or annoyances from spam – where the latter have been the focus of the prior privacy literature. Calo (2014) explores extrinsic privacy harms further and shows how people's personal data can be used to influence their decisions. Along the same lines, Conitzer et al. (2012) show that, counterintuitively, both a low and a very high level of privacy lead to price discrimination and produce extrinsic privacy harms by allowing sellers to capture all surplus, while an intermediate level of privacy is best for consumers. Regarding intrinsic privacy harms, Cofone (2016) introduces a method to quantify intrinsic privacy harms in the context of medicine, which presents a unique context in which privacy preferences can be measured with a dollar value. Cofone and Robertson (2018a) introduce a model of privacy harms that takes into account privacy preferences and therefore captures intrinsic privacy harms, separating it from reputational harms, and apply it to tort law and criminal procedure. Cofone and Robertson (2018b) apply a modification of such model to the context of consumer privacy and show that the nonbelief in the law of large numbers can justify consumer privacy regulations.

Empirical Privacy

The empirical law and economics dialogue takes place mainly in the economics of privacy literature, and it has been called "the new economics of privacy" (Acquisti et al. 2015). It consists of insights on how consumers behave regarding their personal information, and contains sometimes unexplored policy consequences. One of its central findings is that privacy decisions are affected by cognitive and behavioral biases, such as immediate gratification or status quo bias (John et al. 2011). Acquisti et al. (2013), for example, show that consumers exhibit a much higher discrepancy than normal between their willingness to pay and their willingness to accept than in other choices: the value that consumers place in protecting their personal information differs to up to five times depending on the framing of the question. It has also been demonstrated that privacy concerns vary both with context and with personal traits (Acquisti et al. 2015).

Another central finding, and perhaps the central puzzle of this second dialogue, is the "privacy paradox." The paradox describes the phenomenon in which people declare a high value for their private information in surveys but, in incentivized experiments, they disclose such information for low compensation. Studies on how much people value their privacy show an inconsistency between their declared concern for privacy and their actual behavior online (e.g., Norberg et al. 2007; Beresford et al. 2012). Privacy concerns are a weak predictor of the amount of personal information disclosed. However, consumers are not completely unreactive to privacy. Advertisements that are targeted and obtrusive are more likely to trigger privacy concerns among users than advertisements that do not match the content of a website (obtrusive but not targeted or do not impede visibility (targeted) but not obtrusive) (Goldfarb and Tucker 2011). People's ability to deal with privacy choices changes with their complexity (John et al. 2011). Consumers are also willing to pay to purchase from merchants that protect their privacy when privacy policies are available and their content is salient (Tsai et al. 2011). And, as shown through a series of surveys asking about personal financial information, privacy concerns have increased over time (Goldfarb and Tucker 2012).

Another major finding is the extent to which consumers are unaware of what policies they agree to and what happens with their data – uncertainty that is usually referred to as "consumer unawareness." Milne and Culnan (2004) show that people rarely understand privacy policies and license agreements. McDonald and Cranor (2008) show that a disproportionate amount of

time and energy would be needed to read all privacy policies one is involved with. Hoofnagle et al. (2012) show that the advertising industry finds new ways of tracking that consumers are often unaware of. Hoofnagle and Urban (2014) show that consumers are unaware of the extent in which they are subject to behavioral targeting and believe that there are implied duties of confidentiality also when these do not exist. There is discussion on whether privacy notices are effective at counteracting this. Some show empirical evidence of their ineffectiveness to increase consumer awareness (Martin 2016). Others find that simplifying such disclosures along standard best-practices also has no effect (Ben-Shahar and Chilton 2016) and that consumers are unreactive of different types of language in privacy policies (Strahilevitz and Kugler 2016).

Cross-References

- ▶ Privacy Regulation
- ▶ Signalling

References

Acquisti A, John LK, Loewenstein G (2013) What is privacy worth? Journal of Legal Studies 42(2):249

Acquisti A, Brandimarte L, Loewenstein G (2015) Privacy and human behavior in the age of information. Science 347(6221):509

Ben-Shahar O, Chilton A (2016) Simplification of privacy disclosures: an experimental test. Journal of Legal Studies 45(S2):41

Beresford A, Kübler D, Preibusch S (2012) Unwillingness to pay for privacy: a field experiment. Economic Letters 117:25

Calo R (2011) The boundaries of privacy harm. Indiana Law Journal 86(3):1131

Calo R (2014) Digital market manipulation. George Washington Law Review 82(4):995

Cofone IN (2016) A healthy amount of privacy: quantifying privacy concerns in medicine. Cleveland State Law Review 65(1):1

Cofone IN (2017) The dynamic effect of information privacy law. Minnesota Journal of Law, Science & Technology 18(2):517

Cofone IN, Robertson AZ (2018a) Privacy harms. Hastings Law Journal 69(4):101

Cofone IN, Robertson AZ (2018b) Consumer privacy in a behavioral world. Hastings Law Journal 69(6):1193

Cohen J (2000) Examined lives: informational privacy and the subject as object. Stanford Law Review 52(5):1373

Conitzer V, Taylor CR, Wagman L (2012) Hide and seek: costly consumer privacy in a market with repeat purchases. Marketing Science 31(2):277

Daughety AF, Reinganum JF (2010) Public goods, social pressure, and the choice between privacy and publicity. American Economic Journal: Microeconomics 2(2):191

Goldfarb A, Tucker C (2011) Online display advertising: targeting and obtrusiveness. Marketing Science 30(3): 389

Goldfarb A, Tucker C (2012) Shifts in privacy concerns. American Economic Review Papers & Proceedings 102(3):349

Gradwohl R, Smorodinsky R (2017) Perception games and privacy. Games & Economic Behavior 104:293

Hoofnagle CJ, Urban JM (2014) Alan Westin's privacy homo economicus. Wake Forest Law Review 49(2):261

Hoofnagle CJA, Soltani A, Good SN,Wambach DJ, Ayenson MD (2012) Behavioral advertising: the offer you can't refuse. Harvard Law & Policy Review 6 (2):273

John LK, Acquisti A, Loewenstein G (2011) Strangers on the plane: context-dependent willingness to divulge sensitive information. Journal of Consumer Research 37(5):858

Kahn CM, McAndrews J, Roberds W (2005) Money is privacy. International Economic Review 46(2):377

Kim JH, Wagman L (2015) Screening incentives and privacy protection in financial markets: a theoretical and empirical analysis. RAND Journal of Economics 46(1):1

Litman J (2000) Information privacy/information property. Stanford Law Review 52(5):1283

Martin K (2016) Do privacy notices matter? Comparing the impact of violating formal privacy notices and informal privacy norms on consumer trust online. Journal of Legal Studies 45(S2):191

McDonald AM, Cranor LF (2008) The cost of reading privacy policies. I/S: Journal of Law and Policy for the Information Society 4: 543

Milne GR, Culnan MJ (2004) Strategies for reducing online privacy risks: why consumers read (or don't read) online privacy notices. Journal of Interactive Marketing 18(3):15

Murphy R (1995) Property rights in personal information: an economic defense of privacy. Georgetown Law Journal 84(7):2381

Norberg PA, Horne DR, Horne DA (2007) The privacy paradox: personal information disclosure intentions versus behaviors. Journal of Consumer Affairs 41(1):100

Posner R (1978) The right of privacy. Georgia Law Review 12(3):393

Posner R (1981) The economics of privacy. American Economic Review Papers & Proceedings 71(2):405

Samuelson P (2000) Privacy as intellectual property? Stanford Law Review 52(5):1125

Schwartz P (1999) Privacy and democracy in cyberspace. Vanderbilt Law Review 52(6):1607

Schwartz P (2004) Property, privacy, and personal data. Harvard Law Review 117(7):2056
Solove D (2006) A taxonomy of privacy. University of Pennsylvania Law Review 154(3):477
Stigler G (1980) An introduction to privacy in economics and politics. Journal of Legal Studies 9(2):623
Strahilevitz L, Kugler MB (2016) Is privacy policy language irrelevant to consumers? Journal of Legal Studies 45(S2):69
Tsai JY, Egelman S, Cranor LF, Acquisti A (2011) The effect of online privacy information on purchasing behavior: an experimental study. Information Systems Research 22(2):254

Privacy Regulation

Fabrice Rochelandet[1] and Silvio H. T. Tai[2,3]
[1]IRCAV, Université Sorbonne Nouvelle Paris 3, Paris Cedex 05, France
[2]PPGE, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil
[3]RITM, Université Paris-Sud, Paris, France

Abstract

Privacy regulation is a set of rules and enforcement tools designed to control the collection and use of personal information. Not only these rules aim at protecting privacy, but also reducing the scope of misuses of information including identity theft, higher prices, spam, and effort spent by individuals to protect their data. The spectrum of privacy instruments varies from purely (self-regulating) market-based solutions to regulatory-orientated rules and from ex ante to ex post tools. The protection of personal data involves costs for firms, such as the restriction of available information and detrimental effects on innovation. From the consumer point of view, costs are related to consent and information issues, such as reading or writing privacy charts, complying with privacy standards and adopting privacy-enhancing technologies. Since there are privacy trade-offs arising from the interaction between this regulation and other economic and social issues, the economic impacts of privacy regulation depend to the adequacy of the privacy protection arrangement to the context.

Definition

Privacy regulation is a set of rules and enforcement tools designed to control the collection and use of personal information.

Privacy Regulation

Privacy regulation is a set of rules and enforcement tools designed to control the collection and use of personal information. Not only these rules aim at protecting privacy but also reducing the scope of misuses of information including identity theft, higher prices, spam, and effort spent by individuals to protect their data. It articulates legal rules (US *Privacy Act*, French law "informatique et libertés"), the role of medias and civil society, the use of technologies (encryption, anonymous web browsing), and the actual behavior of organizations and individuals, that are more or less privacy sensitive. The weighting of each of those factors varies among countries: for instance, the role of medias in the USA proves to be stronger than in European countries. As for the public intervention to protect privacy, the USA and the EU have adopted two different approaches. While the USA is limited to ad hoc regulation, the EU puts the spotlight on general regulation based on principles.

Various instruments permit to regulate privacy. There is a spectrum of solutions between purely market-based ones (reductionist approach supporting self-regulation of firms) to regulatory-orientated rules (strict application of notice and consent mechanisms). Many authors focus on the cost-benefits analysis and comparisons of these solutions (Swire 1997; Bennett and Raab 2006; Koops et al. 2006). Privacy instruments can also range from ex ante tools that apply before the damages (e.g. legal prohibition to collect data on minors, privacy by design rules, the fair information practices) to ex post solutions (e.g. a liability rule, the right to be forgotten).

Literature on privacy regulation has emerged since the 1970s with the enactment of the US Privacy Act. In particular, Judge Richard Posner (1981) considers it as a source of market

inefficiency. By hindering the capacity of firms to gather personal data, privacy is supposed to reduce the information available in the markets about consumers, workers, borrowers, and so on. However, the advent of the Internet and digital platforms renew these issues. Moral hazard issues derive from the costs to observe the uses that firms make from the data they gather over individuals. Their well-being is undermined from such a position of imperfect or asymmetric information regarding who collect their data, when, for what purposes, and even with what consequences. At the same time, regulating firms can impose them important legal costs that could have detrimental effects on innovation and business.

The design of an efficient regulation raises many difficulties. One is to assess the optimal rules of the regulation because this requires a large amount of information that regulators rarely know. For instance, what are exactly the actual harms due to privacy infringement and how to evaluate them or who are the "polluters" in a digital environment? It proves very uneasy to know who gather what kind of information and to do what.

Another problem is that privacy protection can be a source of failures (Rubin and Lenard 2001) either because of their implementation costs (e.g., detection of unlawful behavior) or because of their consequences (e.g., restrictions on digital innovations). The costs of implementing privacy regulation might be higher than those rising from privacy abuse. For instance, internet users benefit from free online services, social networking service, and targeted advertising thanks to the exploitation of their online data. Quite often, privacy regulations require a combination of various instruments (institutional design), and thereby they can interact with other regulations and laws. Some authors support self-regulation by considering firms (advertisers, website, and platform operators) as sensitive to the consumer response to their strategies whenever these ones prove intrusive (invasive or too much targeted ads, spam, and unacceptable price discrimination) and irritate consumers (Anderson and Simester 2010). Hence, firms can engage themselves in pro-privacy competition, for instance, by choosing to actually comply with their privacy commitment (Jamal et al. 2003). However, the efficiency of self-regulation supposes a certain ability of individuals to be informed and the capacity of firms to provide protection of privacy without excessive costs (Swire 1997). Here, the advantage of more strict regulatory solution is to permit individuals to save on costs associated with the need for complex information inherent with various privacy policies and interaction with firms that exploit their data (Milberg et al. 2000).

From an economic point of view, the efficiency of any privacy regulation is to minimize the social cost associated with the exploitation personal data while maximizing its social welfare. Therefore, there are at least three criteria to evaluate the efficiency of a privacy instrument (Rallet and Rochelandet 2011). First, a privacy instrument can be considered as efficient whenever it minimizes the transaction costs associated with its setup and enforcement. Privacy involves costs for the consumers, including the costs of consent when they want to use an online service. More generally, private agents can incur information costs, i.e., when reading or writing privacy charts, complying with privacy standards, adopting privacy-enhancing technologies, and so on. The same prevails for public institutions when they set and enforce privacy rules with the surveillance and detection of unlawful behavior. All these costs are unequally allocated among institutions, organizations, and individuals, and hence the resulting allocation could affect the social benefits of such or such privacy instrument. For instance, a property rule in favor of individuals can be inefficient because these ones are not always in the best position to assess the value of their data, leading to overestimation from their part and then less innovation and socially useful exploitation of their data. Similarly, charging too stringent privacy norms to firms can be costly for them to comply with, leading to underinvestment in innovation and quality of service.

Secondly, in order for privacy regulation to be efficient, rules assigned to every regulated agent must be followed by them. However, incentives as

well as financial or cognitive ability might be insufficient. For instance, psychological biases can lead to overexposure of individuals over digital platforms, even though they have strong preferences for privacy (Acquisti et al. 2015). In addition, firms can benefit from asymmetric information when it appears to be costly to observe their behavior, resulting in an exploitation of resources in the market for personal data (Schwartz 2004). Moreover, privacy rules can lead to firms to adopt inefficient decisions. For instance, they can decide to adopt business models (Cecere and Rochelandet 2013) or to locate their business according to the actual level of privacy (Rochelandet and Tai 2016).

Thirdly, There are privacy trade-offs arising from the interaction between this regulation and other social issues such as innovation, health, antitrust, and public order. How to best protect privacy without harming the positive effects of information sharing and exploitation? To some extent, privacy rules can complement or, conversely, be opposed to other policies through its effects on the functioning of markets. In a market perspective, ex post regulation is held to be more in favor of innovation by letting the market and innovation process work, even if it means subsequent privacy harms. However, such damages might be irremediable because they result from the use of information goods that are nonrival and then hard to seal off. However, too vigorous privacy regulation can be inconsistent with other social issues such as health, public order, freedom of speech, and innovation. For instance, the fact that European Union Data Protection Directive is more restrictive than the US Privacy Act could lead to a reduction in welfare (for instance, in health services, Miller and Tucker 2011) or a reduction in the general level of privacy protection (Rochelandet and Tai 2016 show that stringent privacy laws may lead firms to locate in less stringent countries, what weakens the actual privacy protection).

Possible approaches to privacy protection range from market-based solution to formal regulation. Thus, different frameworks are in-between self-regulation and strict regulatory protection. The economic impacts of privacy regulation depend to the adequacy of the privacy protection arrangement to the economy context.

References

Acquisti A, Brandimarte L, Loewenstein G (2015) Privacy and human behavior in the age of information. Science 347(6221)

Anderson E, Simester D (2010) Price stickiness and customer antagonism. Q J Econ 125(2):729–765

Bennett CJ, Raab CD (2006) The governance of privacy: policy instruments in global perspective. MIT Press, Cambridge, MA

Cecere G, Rochelandet F (2013) Privacy intrusiveness and web audience: empirical evidences. Telecommun Policy 37(10):1004–1014

Jamal K, Maier M, Sunder S (2003) Privacy in e-commerce: development of reporting standards, disclosure, and assurance services in an unregulated market. J Account Res 41(2):285–309

Koops BJ, Prins C, Schellekens M, Lips M (eds) (2006) Starting points for ICT regulation: deconstructing prevalent policy one-liners, Information technology & law series, vol 9. T.M.C. Asser Press, The Hague

Milberg SJ, Smith HJ, Burke SJ (2000) Information privacy: corporate management and national regulation. Organ Sci 11(1):35–57

Miller AR, Tucker CE (2011) Can health care information technology save babies? J Polit Econ 119 (2):289–324

Posner RA (1981) The economics of privacy. Am Econ Rev 71(2):405–409

Rallet A, Rochelandet F (2011) La régulation des données personnelles face au web relationnel : une voie sans issue ? Réseaux 29(167):19–47

Rochelandet R, Tai SHT (2016) Do privacy laws affect the location decisions of internet firms? Evidence for privacy havens. Eur J Law Econ 42(2): 339–368

Rubin PH, Lenard TM (2001) Privacy and the commercial use of personal information. Kluwer Academic Publishers, Norwell, MA, USA

Schwartz P (2004) Property, privacy, and personal data. Harv Law Rev 117(7):2056–2128. https://doi.org/10.2307/4093335

Swire PP (1997) Markets, self-regulation, and government enforcement in the protection of personal information. In: Privacy and self-regulation in the information age. U.S. Department of Commerce, Washington, DC

Private Law

► Customary Law

Private Property: Origins

Enrico Colombatto[1,2] and Valerio Tavormina[3]
[1]Università di Torino, Turin, Italy
[2]IREF (Institut de Recherches Économiques et Fiscales), Lyon, France
[3]Università Cattolica del Sacro Cuore, Milan, Italy

Abstract

This entry focuses on the legitimacy of private property and analyzes the process of first appropriation. In particular, we examine and comment the different views on the origin of private property rights that have emerged through the history of economic and legal thinking, from Democritus to de Jasay. These views have been grouped in two broad categories: consequentialism and fundamental principles. Although consequentialism is now dominant among economists and inchoate in the legal profession, we observe that it is in fact an alibi for discretionary policymaking by the authority. By definition, fundamentalist approaches generate rules that limit discretion. However, we show that some fundamentalist views rest on questionable a priori statements. De Jasay's argument based on the presumption of liberty is perhaps the only perspective that escapes this criticism.

Introduction

Property rights play a crucial role in economics: They define the very essence of this discipline, which studies how individuals exchange in order to enhance their welfare, subject to scarcity constraints. If there were no property rights, grabbing and looting would replace exchange, and the time horizon of any economic activity would depend on how effectively and at which cost each individual can protect the goods under his/her control.

One can identify three categories of property rights regimes: common property, centralized property, and private property (see also Waldron 2016). Common property defines a context in which the notion of property is abolished or severely restricted. It is in fact equivalent to the absence of property rights: Every individual in the relevant community/group is free to claim and appropriate everything he/she finds. The underlying idea is that individuals produce for their own immediate consumption or for the community at large (altruistic behavior with an expectation that the others will behave likewise) and that each individual has a right to take and consume whatever he likes. Theft is thus de facto legalized, unless it involves physical violence. Actually, one could even argue that theft no longer exists, because there are no legal owners and no member of the community has the right to prevent others from taking.

Not surprisingly, this social arrangement is of limited practical interest. It would quickly lead to the demise of the community (hardly anybody would produce goods and services, except for situations in which they can be manufactured secretly and consumed immediately) or to slavery (the most effective looters would force the rest of the community to produce and surrender their output). Of course, in the latter case, the slave masters would be the actual owners.

By contrast, centralized property corresponds to a system in which property rights are clearly assigned and belong to a central authority. This authority can be an individual, such as a dictator or an absolute sovereign. For example, since the end of the XVI century, absolute monarchies were based on the idea that God gave all the existing resources to one individual, who would then manage them in the interest of the community and in accord with God's design. The central authority can also be a set of individuals chosen through a shared procedure (elections). This set of individuals often operates by majority voting (e.g., parliaments) or assigns to other parties the power to decide on its behalf (e.g., an agency). Centrally planned economies and modern social democracies follow these patterns. In these circumstances, the expression "private property" is not absent from the debate, but it is in fact misleading. For example, in modern democracies, the central authority enjoys full power to encroach upon

somebody's property. This means that in a social democracy, the central authority allows individuals to make use of the goods in their possession within the limits defined by the central authority itself. As a result, the ultimate owner is actually the central authority, rather than the individual. In these cases, therefore, individuals engage in economic activities with other individuals, subject to their expectations about encroachment by the authority on their preferences and (temporary) property. When individuals interact with the government widely understood (policymakers, bureaucrats, agencies), they know that their counterparts are the ultimate judges of their own behavior and therefore enjoy discretionary power to which little opposition is possible. As mentioned above, property is merely temporary and far from absolute.

The recognition and enforcement of private property rights are the founding pillars of a free-market economy. Private property means that individuals have absolute, exclusive, and permanent rights on what they legally own: they can do whatever they like with their property, nobody can interfere with their decisions, and there is nobody to whom these rights must be returned after a given time period. Thus, under this regime, each individual engages in unfettered voluntary exchange, subject to his/her compliance with the freedom-from-coercion principle (no violence and no cheating are allowed) and insofar as he/she respects the private property of the other individuals. Put differently, the legitimacy of private property and the freedom-from-coercion principle specify the moral foundations of a free-market economy. By contrast, the illegitimacy of private property and the limits to private property define the features of the centralized economies, regardless of the political format – dictatorship or social democracy.

The debate on the origin and legitimacy of private property is thus of crucial importance, since it defines the very features of an economic system (institutions), the role of government, and the content of economic policymaking (see also Alchian 1965). Briefly put, the debate on the legitimacy of private property focuses on two areas. One regards property rights on natural resources, while the other regards property rights on manmade goods, services, and intangibles. Research on the origins of private property analyzes and explains the process through which an individual can legitimately appropriate a resource never previously appropriated by somebody else. Moreover, the literature examines whether the goods, services, and ideas produced by the individual become private property of the producer; or whether they actually belong to somebody else, e.g., the community of which the producer is a member; or whether they are part of the common pool and thus belong to nobody.

The following sections provide a critical analysis of these two agendas by considering the different views developed through the history of economic and legal thinking. In particular, the next section is devoted to the consequentialist, natural right, and religious beliefs before John Locke. Section "Aquinas and Locke on the Process of Appropriation" explores the notion of property as the result of rightful appropriation. Section "A Different Natural Right Approach to Private Property: Natural Dominion" focuses on natural dominion, while section "Recent Free-Market Agendas: From Demsetz to De Jasay" discusses the more recent approaches. The final sections conclude and offer a brief outline of the debate about intellectual property rights.

Private Property Before John Locke: Democritus, Aristotle, and the Etruscan Legacy

The first attempt to justify individual appropriation – as opposed to common property – harks back to Democritus (460–370 BC). In his view, private property is justified because it leads to superior economic results (efficiency). In a sentence, individuals make better use of the resources when they own them (Diels 1903: 55 B, frag. 279). In today's wording, one would say that when private property rights are well defined and enforced, the positive and negative externalities generated by the use of a resource are minimized, and efficiency is enhanced.

According to this approach, therefore, private property is just because it leads to desirable outcomes, efficiency being the criterion that defines desirability. Surely, this view is squarely in the consequentialist camp: morality defines justice, and, in turn, the desirability of the outcomes associated with one action or one institutional arrangement defines morality.

However, considering desirability (indirectly) equivalent to justice presents one major weakness. It fails to specify who should define desirability. This is not a trivial aspect, since it is apparent that different individuals are likely to have different preferences and thus attach different meanings to the notion of "desirable." For example, one could follow Plato and argue that since each individual should pursue virtue, and since private property encourages greed and social tensions to the detriment of virtue and peaceful coexistence, private property should be outlawed. Likewise, the very notion of efficiency is ambiguous. Efficient is whatever enhances value. Yet, value is subjective, and, therefore, the notion of efficiency can vary across individuals and groups of individuals, depending on preferences, religion, traditions, and culture. Certainly, technical efficiency is by no means enough to measure social happiness, whatever this means.

The upshot is that if one follows this consequentialist viewpoint, the presence, the boundary, and the stability of private property rights are conditional on a notion of desirability that is necessarily arbitrary and subject to change over time. Put differently, by resorting to the notion of desirability in order to legitimize private property, one actually avoids analyzing the foundations of private property and moves instead to comparing different concepts of social desirability. Yet, this comparison is hardly conducive to a useful answer. In order to avoid arbitrary rule, therefore, the notion of social desirability requires that the members of the community unanimously agree on a hypothetical desirability function (social preferences). Moreover, this notion also requires that the community members unanimously agree on the institutional tools that allow a society to obtain the desired set of goals (policymaking): when raising revenues to finance desirable public expenditure, a tariff on car imports is different than a tax on inheritance. Failure to meet these two social choice constraints – shared goals and shared instruments – ensures that we can say little or nothing about the consequentialist justification of property in a society.

Aristotle (384–322 BC) enriched Democritus' arguments in favor of private property by mentioning the role of human nature. According to his line of reasoning, since history and factual observation show that individuals like to own resources and goods, one must necessarily conclude that the principle of ownership is indeed part of human nature (see Politica: I,8 and II,5). Hence, the institution of private property is part of natural law, and denying or constraining private property would amount to negating the essence of each human being and of the natural order. The Romans also referred to natural law: the use of reason, the observation of reality, and consistency with tradition are the instruments through which cultivated men discover the natural law – "si in unum sententiae concurrunt, id, quod ita sentiunt, legis vicem optinet" (Gaii Inst., I § 7). And private property is part of the natural law (Gaii Inst., II §§ 66, 69, 73, 79).

However, and despite its simplicity and consistency, the Aristotelian version is also vulnerable to doubts. By claiming that the foundation of private property lies in natural law, one makes a statement that raises additional questions, rather than answers the original query. Put differently, by claiming that private property is founded on natural law, one wonders whether private property is indeed a component of natural law. One can also ask why natural law is superior to positive law (manmade legislation) and – last but not least – who has the final word on all these questions.

A third and final vision on private property was typical of the Etruscans and very popular with the Romans. As mentioned in Liggio and Chafuen (2004), religion is a family matter: each family's ancestors are sacred and present a sacred link with the land upon which the family insists. Violating that land and – more generally – the family's belongings was sacrilegious and usually deserved capital punishment, as the Romulus and Remus legend witnesses. Thus, there could be no religion

(and no family) without private property. Hence, private property is not a matter of economic efficiency, nor is it related to human nature. Rather, it is the necessary connection between the household and the gods. It is part of the essence of classical civilization and an element that will also play a crucial role in the defense of private property put forward by the early Fathers, together with the obligation of sharing with "the poor [...] the fruits of [the owner's] labour" (Lactantius, Divine Institutes, V, 5).

Aquinas and Locke on the Process of Appropriation

St. Thomas Aquinas (1225–1274) produced two original insights, from which a theory of private property legitimized by rightful appropriation emerged. Aquinas agreed with the dominant classical worldview on the desirability/efficiency of private property. Yet, he made a second point, which was new and of great importance. He argued that unused resources do belong to humankind. However, when an individual combines them with his own labor, his claim to the resources trumps all others' and transforms possession into private property (this is one way of reading Summa Theologiae, II-II, 66, 1). A few years later, this argument was forcefully repeated, clarified, and expanded by John of Paris (1250–1306), a student of Aquinas': "lay property is not granted to the community as a whole. . ., but is acquired by individual people through their own skill, labour and diligence, and individuals, as individuals, have right and power over it and valid lordship; each person may order his own and dispose, administer, hold or alienate it as he wishes, so long as he causes no injury to anyone else since he is lord" (quoted in O'Donovan and O'Donovan 1999: 403; see also Rothbard 1995: 57; and Kilcullen and Robinson 2017).

In modern times, John Locke's Second Treatise (1689, chapter V) popularized and expanded the line of theorizing initially proposed by St. Thomas and John of Paris and later enriched by the Dominican Francisco de Vitoria (1483–1546), the Levellers, and Earl Shaftesbury (see Lepage 1985: 63; Rothbard 1995: 315–317; see also Vaughn 1980). Locke's line of thinking can be summarized in six points:

1. God gave natural resources to humankind, and are part of the so-called state of nature.
2. Every individual is the owner of himself.
3. By mixing natural resources with his own manual labour, ideas and creative efforts, the individual removes the resources from state of nature and makes them his own.
4. Property includes the resources initially appropriated as well as their fruit.
5. Since God would disapprove of wastage or abuse, appropriation is no longer valid when private property involves wastages, or when appropriation prevents other people from satisfying their needs.

   From a normative viewpoint, Locke argues that the preservation of property is the only function of the social contract, which is the origin of government, an institution formed by individuals who agree on finding a peaceful solution to disputes. In fact,
6. Private property pre-dates government and justifies its existence.

Clearly, the notion of the "state of nature" plays a key role in the Lockean context. This concept was actually introduced by Juán de Mariana at the end of the sixteenth century. It is a synonym for the common pool and defines a situation in which resources do not belong to anybody, not even to the community as a whole. Put differently, one could argue that Locke did not theorize private property as an institution inherent in nature or legitimate per se, but rather as a desirable manmade institution that originates from common property (in the state of nature) and ownership of one's own self. It materializes through an act of appropriation (removal of a resource from the state of nature), is consistent with God's will, and is subject to constraints dictated by God's will.

All in all, Locke's theory remains incomplete and not entirely satisfactory. As observed earlier, his key elements are the emphasis on the role of

self-ownership, which allegedly justifies ownership on what the individual removes from the state of nature (the finders-are-keepers principle already formulated by Gaius, Inst. § 66), and the fact that God is not opposed to private property. However, this is Locke's weakness; it is not evident that self-ownership justifies grabbing from the common pool. Locke's view about the legitimacy of private property is more a description of how private property emerges, rather than an explanation of legitimacy. In other words, the essence of Locke's argument boils down to claiming that God created common property and that He does not object to grabbing for a good purpose (wealth creation). In this light, one may indeed argue that Locke's line of reasoning rests on consequentialism (wealth creation), possibly mixed with an act of faith (God wants people to improve their welfare well beyond what is needed to survive and private property serves that purpose). Hence, the Lockean vision not only presents the limitations typical of all consequentialist approaches but also relies on one's vision about religion and is burdened by the famous proviso which, if taken literally, de facto prevents first appropriation in the presence of scarcity or makes first appropriation conditional on everybody else's consensus.

A Different Natural Right Approach to Private Property: Natural Dominion

Consequentialism and appropriation are surely the most popular arguments in favor of private property. As mentioned earlier, consequentialism harks back to classical Greece and maintains that private property is justified by the desirable results (efficiency) it produces. By contrast, the case for appropriation was put forward in the Middle Ages; it was further developed by Locke and implies acceptance of the dominion thesis: consistent with God's design, men are free to exploit natural resources (including animals), and private property is the outcome of appropriation by an individual or a group of individuals, as long as no other human being is harmed.

However, the Middle Ages not only claimed that private property is desirable and consistent with God's design. Not long before the end of its pontificate, Pope John XXII (1249–1334) argued that private property is just also from a deontological perspective. In his view, and in contrast with other property regimes, private property stems from "pure" natural rights. As argued in the encyclical *Quia vir reprobu*s (1329), which the Pope released to condemn Franciscan pauperism, God owns whatever exists, and since man has been created in the image of God, the principle of private property is necessarily embedded in the very nature of each human being. Hence, limiting private property would amount to disputing God's design. Of course, in this case, the adjective "pure" is important, since the pure version of natural rights emphasizes that these rights are embedded in each individual since his/her birth.

This natural-dominion approach differs from the Aristotelian version, according to which natural rights are those revealed by spontaneous behavior and tradition. And it also differs from the rationalist version, which owes a great deal to Grotius (1583–1645), according to whom natural law coincides with what is needed to ensure survival, while natural rights are manmade absolute rights dictated by reason and independent of the circumstances. Within the framework suggested by Grotius, therefore, sociability is the moral benchmark with reference to which all institutions are evaluated, and private property is the operational tool to meet sociability (man's inclination to live in harmony with other human beings).

Of course, the normative consequences of the natural-dominion thesis are important. According to this approach, private property is all but sacred and can never be encroached upon; whereas according to the rationalist perspective, private property ends up being the result of human constructivism and has nothing to do with religion. This also affects the very notion of right. In the former case, a right corresponds to one's freedom to use a resource as he pleases; whereas in the latter (rationalist) case, it corresponds to one's claim to enjoy something, as dictated by the (hopefully) enlightened authority.

Recent Free-Market Agendas: From Demsetz to De Jasay

In recent times, the debate on the origins of private property has subsided, with few exceptions. Indeed, mainstream theorizing has taken an evolutionary turn. Bordering with Benthamite utilitarianism, Demsetz (1967) and Pejovich (1972) pioneered an approach that neglects the moral justifications of private property and tends to treat this institution as the spontaneous result that emerges when groups of individuals face the problem of scarcity. In this context, the term "spontaneous" underscores the fact that private property is the result of human action, but not the result of constructivist design. Rather, it originates from trial and error, so that bad solutions lead to unsatisfactory performance and tend to be discarded in favor of superior institutional answers (see also Barzel 1989: Chap. 6).

In brief, it is claimed that private property emerges when unrestrained access to scarce resources leads to inefficiencies (overexploitation) and when a community finds it appropriate to generate incentives that drive individual behavior in a desirable way. Put differently, private property is a way of granting access only to selected individuals (the legitimate owners) and of managing valuable goods by internalizing potential externalities. Two consequences follow. First, private property is a manmade institution, the features of which evolve according to the current environmental conditions (transaction costs, preferences, technology). Thus, it can be neither absolute nor perpetual. Second, this evolutionary approach raises a normative issue: Who defines the features of a private property right system? This literature draws attention to the institutional entrepreneurs, who devise and experiment alternative property arrangements, and to the judges, who carry out the necessary cost-benefit analyses following which property rights are assigned and reassigned. As a result, good arrangements survive and bad arrangements are corrected, ignored, or discarded. However, when judges, experts, and lawmakers are responsible for the assignment, the recognition, and the enforcement of property rights, the outcome is de facto determined by government, and spontaneity necessarily falls victim to state coercion. Not unlike Jeremy Bentham and John Stuart Mill, therefore, one hopes that coercion pursues the common interest, whatever this means; and one neglects to notice that the interests of the majority end up trumping those of the minority.

More generally, the recent evolutionary approach mentioned above follows the Lockean tradition in that it has little to say about the origins of private property. Similar to Locke, it provides a consequentialist description of how private property emerges. However, Locke based his consequentialist approach on God's will. By contrast, the law-and-economics tradition to which Demsetz and Pejovich belong bases its consequentialist claim on the utilitarian stamp of an enlightened government. This explains why, and in contrast with the Lockean tradition, the evolutionary approach takes for granted that government precedes private property: legislators create the law in accord with utilitarian principles, and the law defines private property.

Not surprisingly, in order to recover the essence of the debate on the origin of private property, one has to focus on those libertarian scholars who deny the legitimacy of government as a coercive authority, a role that necessarily transforms the state into the source of property. In particular, by drawing on John Locke, these unorthodox authors maintain that private property preexists government and that, therefore, government cannot encroach upon private property.

In order to develop their argument, the libertarians follow two lines of reasoning, which do without God and religion and ignore Locke's proviso (see Bouillon 2011). The first one starts from the notion of self-ownership and generalizes this principle into an ethics of private property. A second perspective focuses on a different notion of legitimacy, which turns the burden of proving the legitimacy of private property upside down by relying on the notion of Paretian optimality.

Rothbard (1974) is probably the most prominent proponent of the first view, based on an e contrario argument. Rothbard justifies appropriation by observing that we live in a world of scarcity and that a resource is scarce if at least two individuals want to exploit it. Now, individual

A does not need to ask permission to exploit/appropriate the resource if it is a first appropriation. Thus, A becomes the lawful first owner. By contrast, if A must ask permission, it means that B (or somebody else before him) had already become the owner (or the co-owner) through first appropriation. Put differently, the Rothbardian perspective transforms the debate on private property into an analysis of the legitimacy of first appropriation. If one denies the right of first appropriation, all potential first owners should behave as if all resources were in common, including their own selves and the air they breathe (see also Boaz 1997). This would be absurd, since if each individual had to ask permission to act and breathe from billions of other people, the human race would quickly perish.

Rothbard also rejects the possibility that the individuals are at least partially owned by others – e.g., policymakers – who take decisions in the interest of the rest of the community. In his view, this possibility would be immoral and contradictory. It would be immoral, because all moral rules require generality (all individuals must enjoy the same rights). It would be contradictory, because accepting asymmetric ownership would be equivalent to saying that the human race is actually made of humans and non-/subhumans. According to Rothbard, therefore, the first appropriation is a law of nature because it is necessary for survival and establishes a general principle. Hence, the natural origin of private property. This principle is moral and also applies to the first appropriation of what an individual creates with his own labor.

De Jasay (1991, 1998) is an advocate of the second approach, based on the so-called presumption of liberty, i.e., on the idea that an individual can act as he pleases, unless a challenger falsifies the presumption by raising justified objections. Justified objections stem from three situations: (1) when it is apparent that an action conflicts with (spontaneous) conventions, (2) when the actor violates obligations that he had previously and voluntarily assumed, and (3) when somebody else's liberty to act is impaired (harm). The third condition is actually reminiscent of the weaker Lockean proviso, which is how Robert Nozick (1974: 178–182) identified a situation in which individual A's action prevents others from exploiting opportunities to enhance their well-being. Although situation (3) above remains ambiguous – one could eliminate all ambiguities by understanding the term "harm" as a synonym for physical violence at the expense of the challenger – the core thesis outlined by Nozick and De Jasay is clear: the legitimacy of first appropriation is guaranteed by the lack of opposition by individuals who could claim a previous right over the good/resource. If such opposition had substance, then the finder would not be the first finder and could not keep the good. Another individual is in fact the first finder or the legitimate owner of the goods he obtained from the true first finder.

Put differently, and absent justified opposition, the finders-are-keepers rule is consistent with the presumption of liberty and necessarily represents a Pareto improvement: it makes the finder better off and makes nobody worse off. The second comer would be in a different position, since his claim to the goods appropriated by the finder would violate the presumption of liberty and would not be a Pareto improvement. Of course, the presumption of liberty also applies to private property obtained through exchange or by means of inheritance and to self-ownership.

To summarize, by articulating a presumption of liberty, De Jasay introduces a negative notion of legitimacy: all non-illegitimate actions are legitimate, and the burden of proof lies with the challenger. Since the presumption of liberty is met when nobody has justified cause to object, and since the absence of objection is also the essence of a Pareto improvement, all actions that imply a Pareto improvement are necessarily legitimate. In regard to the legitimacy of private property, therefore, the debate boils down to assessing whether an individual's property is based on a previous act of appropriation that violated the presumption of liberty. Clearly, this would be the case with theft, nationalization, or expropriation.

Preliminary Conclusions

By and large, the arguments in favor of private property have focused on three points:

characterizing property rights in the initial status quo, justifying the legitimacy of private property through consequentialism, and explaining the legitimacy of private property by resorting to fundamental principles.

In regard to the first point, the literature analyzes various possibilities. According to one version, the initial position consists in the state of nature, where resources belong to nobody and can be appropriated by the homesteader (first-user principle), possibly with some qualifications (the Lockean clauses). A second version consists in claiming that the initial position involves common ownership of the resources and that regulating the exploitation of the resources owned by a community is a political issue. In practice, the government decides who can use what and to what extent. Hence, government is in fact the origin of private property. This means that an individual manages what the government assigns to him, subject to the conditions and during the time period established by the authority. Following a third perspective, it has been claimed that the origin of private property is the individual property of one's own self, a concept that defines the very idea of individual and in the absence of which life would be impossible. Finally, according to a fourth (and radical) possibility, it is argued that the initial position is irrelevant and that, given a presumption of liberty based on self-ownership, what matters is the extent to which the various steps that have led to the current status are objectionable. Of course, lack of objections amounts to confirming the legitimacy of homesteading and exchange.

With the partial exception of the view proposed by De Jasay, private property of manufactured goods is considered an extension of the principle of self-ownership. If an individual combines the resources he owns – land, raw materials, and labor – denying his property right over the output would involve an act of aggression and amount to a violation of the freedom-from-coercion principle. On the one hand, since each unit of output is a mix of inputs owned by the producer, each unit of output is necessarily property of the producer. On the other hand, it is manifest that the manufacturer is the first finder of the result of his activity and that nobody can raise justified objections to the manufacturer's exercise of his liberty (acting as a producer and producing and appropriating the result).

As mentioned earlier, the second set of explanations regarding the origins and foundations of private property focuses on private property as an institutional arrangement justified by efficiency: it ensures better economic outcomes and wards off social tensions. Although this approach is dominant among economists and inchoate in the legal profession, it remains problematic. As a matter of fact, explaining why private property exists is not the same as analyzing why it is legitimate (and thus immune to manmade rule making). In particular, it raises the problem of specifying who decides about efficiency: the homesteader or the ruler? As result, different perspectives produce different answers.

Finally, the fundamentalist approaches try to justify private property by resorting to religion, or to natural rights, or to expedience (which differs from consequentialism). As we observed, religion is problematic, in that it has no universal value. If one justifies private property by appealing to faith, different religious views can lead to radically different conclusions. Natural rights are different, since the various approaches underscore one or more natural traits of all human creatures and give birth to different theories of property in accord with human nature(s). Hence, if one assumes that all individuals share a set of natural rights, then private property is inviolable insofar as it is recognized as a natural right or strictly derived from a natural right. In a similar vein, if freedom from aggression is recognized as a natural right, private property is legitimate as long as it does not involve aggression (homesteading doesn't, by definition) and as long as appropriation by the comers does not imply violence (voluntary exchange doesn't). However, if one has doubts about the "natural" nature of private property, or about the general validity of the freedom-from-coercion principle, then the legitimacy of this institution becomes questionable. By contrast, expedience rests on identifying the presumption of liberty as the simplest way of organizing a community (De Jasay 2005). This presumption

leads to the finders-are-keepers rule and requires the second comer to challenge the incumbent owner. The reader might notice three interesting aspects. According to this vision, (1) the origin of private property rests on a negative (ex post) notion of legitimacy; (2) the presumption of liberty is not a moral argument in favor of liberty, but rather an operating principle that minimizes costs; and (3) the negative notion of legitimacy ensures that the burden of proof lies with the challenger, a choice that has no moral content, but enhances social cooperation.

An Extension to Intellectual Property

Theories about the foundations of private property have usually focused on natural resources and the fruits of men's activities. In recent times, however, considerable attention has been devoted to the property rights on intangibles, with particular reference to two areas: information and patents. In all these cases, the research agendas share one common feature and are relatively simple.

In contrast with what happens in the realms of natural resources and material goods, intangibles are frequently characterized by the presence of free riding, i.e., by the possibility that individual A benefits from the activity of individual B, regardless of whether A and B are bound by a contract. Knowledge, information, and ideas are typical examples. Hence, when free riding occurs, A's welfare increases thanks to B's labor and talents, regardless of whether B agrees or feels slighted. This raises a question. Is free riding a problem or perhaps a crime against private property? Or is it just a fact of life, which some people consider harmless and others undesirable?

The research agendas follow the lines of thinking suggested by the questions above, depending on whether one believes that free riding should be restrained or, rather, freely allowed. One approach alludes to fairness/envy. These notions play a significant role when some agents have privileged access to relevant information and exploit this privilege in dealing with allegedly uninformed counterparts. Those who lament that uneven information amounts to unfairness require that the informed disclose everything they know, except for situations in which they actually bought or researched the information from which they benefit or in which the information is easily accessible to anybody (in which case there would be no privilege). Legislation against insider trading follows this view: it is all right if individual A reads the business press or carries out extensive research about company Z and decides to operate accordingly on the stock exchange. However, A cannot exploit the information he obtains if he enjoys an allegedly privileged position. In the case of insider trading, the privilege consists in being an employee or an administrator of Z.

By contrast, the legislation regarding patents exemplifies other views on property rights. One is based on the claim to self-ownership and one on consequentialism. The argument stemming from the assumption of self-ownership rests on the fact that the mind is part of one's own self and that, therefore, what is produced by one's mind is necessarily an extension of the individual, who homesteads it. Put differently, the second inventor/author/discoverer cannot claim property rights on something that others have already removed from the state of nature, a state of nature that includes knowledge, skills, and artistic concepts that humankind ignores. Although this thesis has merit, it runs into a number of problems, which we shall only mention. Most innovations are based on earlier insights and discoveries. This approach would require that all inventors track – and ask permission to – all the legitimate owners of the knowledge that the current inventor is using. Moreover, the difference between a new invention, a marginal improvement on an idea already existing, and the bare use of an existing idea is frequently far from clear. Finally, the very act of exercising one's intellectual abilities hardly justifies preventing other individuals from using their own intellectual abilities, which include thinking, observing, and possibly reproducing. Of course, this line of reasoning resonates with the presumption of liberty mentioned in the previous section, a presumption often used as an argument against the existence of property rights on intangibles and thus against the legitimacy of patents. In other words, one may argue that the theory based on

the extension of one's self regards the appropriation of the result one obtains by using his talent, his labor, his knowledge, and his imagination. Yet, a result is not a process. Put differently, discovering a process and making use of it do not remove knowledge from the state of nature. It merely makes knowledge accessible to the inventor and to the rest of the community. If this is true, one can thus conclude that enriching the state of nature does not involve homesteading and that involuntary altruism is not a source of rights.

The previous comments help understand why the current legislation on intellectual property rights is founded on consequentialism rather than on philosophical theorizing. As it often happens in law-and-economics debates, the debate presents different views. We list two of them. On the one hand, a large portion of the literature neglects to discuss the nature and foundations of property rights on intellectual property. Rather, it maintains that governments driven by utilitarian principles are justified in enforcing property rights on intangibles in order to compensate their authors for the damage suffered at the hands of free riders. In other words, the origin of intellectual property rights is the government, which assigns them in the common interest. The consequentialist counterargument is that too much protection awarded to a set of inventors may prevent potential competitors from improving on the existing technology and developing new insights. The upshot is that the presence of free riding justifies patents and other barriers to entry. However, these barriers should expire after some time and allow new competitors to enter the scene at a relatively low cost.

Yet, there is also a second and more recent consequentialist perspective. As mentioned at the beginning of section "Recent Free-Market Agendas: From Demsetz to De Jasay," some eminent scholars argue that private property originates as a response to scarcity. In other words, private property is legitimate because it is an efficient way of exploiting scarce resources: it enhances exchange, and it allows individuals to distribute consumption over time, possibly taking into consideration also the potential benefits enjoyed by future generations. Intangibles, however, do not present a problem of scarcity. The use of knowledge by one individual does not prevent other individuals from exploiting that very knowledge. Hence, absent scarcity, private property has no reason to exist, and the debate on the origin of private property in the realm of intangibles is moot. From a normative viewpoint, therefore, patents have no legitimacy.

Cross-References

- Consequentialism
- Copyright
- Economic Efficiency
- Intellectual Property: Economic Justification

References

Alchian A (1965) Some economics of property rights. Il Politico 30(4):816–829

Barzel (1989) Economic analysis of property rights. Cambridge University Press, Cambridge

Boaz D (1997) Libertarianism, chapter 3. The Free Press, New York

Bouillon H (2011) Business ethics and the Austrian tradition in economics, chapter 2. Routledge, London

de Jasay A (1991) Choice, contract, consent: a restatement of liberalism. Institute of Economic Affairs, London

de Jasay A (1998) Justice. In: Newman P (ed) The new Palgrave dictionary of economics and the law. MacMillan Reference Limited, London, pp 400–409

de Jasay A (2005) Freedom from a mainly logical perspective. Philosophy 80(4):565–584

Demsetz H (1967) Toward a theory of property rights. Am Econ Rev 57(2):346–359

Diels H (1903) Die Fragmente der Vorsokratiker. Weidmannsche Buchhandlung, Berlin

Kilcullen J, Robinson J (2017) Medieval political philosophy. In: Zalta EN (ed) The Stanford encyclopedia of philosophy (Summer 2017 Edition). forthcoming URL = https://plato.stanford.edu/archives/sum2017/entries/medieval-political/

Lepage H (1985) Pourquoi la propriété. Hachette, Paris

Liggio L, Chafuen A (2004) Cultural and religious foundations of private property. In: Colombatto E (ed) The Elgar companion to the economics of property rights. Edward Elgar, Cheltenham, pp 3–47

Locke J (1689) In: Hollis T (ed) Two treatises of government, London. A Millar et al. 1764 – available at http://oll.libertyfund.org/titles/222

Nozick R (1974) Anarchy, state and utopia. Basic Books, New York
O'Donovan O, O'Donovan JL (eds) (1999) From Irenaeus to Grotius: a sourcebook in Christian political thought. Wm. B. Eerdmans Publishing, Grand Rapids
Pejovich S (1972) Towards an economic theory of the creation and specification of property rights. Rev Soc Econ 30(3):309–325
Rothbard MN (1974) Justice and property rights. In: Blumenfeld S (ed) Property in a humane economy. Open Court, La Salle, pp 101–122
Rothbard MN (1995) Economic thought before Adam Smith, vol I. Edward Elgar, Cheltenham
Vaugh K (1980) John Locke's theory of property: problems of interpretation. Lit Lib: Rev Contemp Lib Thought III (1):5–37
Waldron J (2016) Property and ownership. In: Zalta EN (ed) The stanford encyclopedia of philosophy (Winter 2016 Edition). URL = https://plato.stanford.edu/archives/win2016/entries/property/

Privatization

Brady P. Gordon
Trinity College, University of Dublin, Dublin, Ireland

Abstract

Privatization as a tool for economic management emerged in the 1970s and 1980s in response to the decline of Keynesian economics and the collapse of communism and has expanded to become a pillar of public policy in all of the world's major economies. The economic benefits of privatization emerged out of new economic theories which applied long-established principles of market failure to weaknesses in public sector economic governance. It is seen as a useful way of increasing efficiency by introducing private competition to otherwise inefficient, monopolistic, and politicized operations in the public sector. Since the late twentieth century, privatization outcomes have been increasingly recognized as context-dependent and socially contested, but privatization remains a pervasive and useful instrument of government policy in pursuit of a wide range of economic and social aims.

Definition

Privatization describes the transfer of government assets or functions from public to private ownership. Privatization is the opposite of nationalization, in which governments acquire privately owned assets or operations.

Types of Privatization

Privatization encompasses four broad types of transactions:

Denationalization consists of the divestiture of state-owned assets to private ownership, typically through sale or lease. The first large-scale privatization campaigns at state level consisted largely of denationalization. These were led by the UK, the USA, and Chilean governments in the 1980s and saw the denationalization of many public utilities (i.e., industries in the telecommunications, energy, transportation, sewage, and water sectors), traditionally associated with the public sector. Denationalization allows governments to expose what might be inefficient, subsidized and politicized government-owned monopolies to private competition, and increase short-term revenues.

Concessions consist of the granting of a legal right to fund or construct an infrastructure project and/or operate a public service. The most common example of a concession is the right of a private company to construct and operate a toll road or bridge. Concessions allow governments to engage in public works projects which might otherwise be unaffordable or politically infeasible for the government.

Outsourcing or **tendering** consists of contracting with a private entity to finance or deliver a public service that had hitherto been carried out in-house by a branch of the public service. Tendering processes are often considered to allow governments greater control over privatization outcomes than simple divestment, since it allows the government to set minimum service standards and multiple criteria as the basis for competition. Outsourcing may occur for

entire services or it can be introduced at specific levels of the supply chain in order to open financing, production, or service provision stages to private competition. This may be useful where a government wishes to retain control over the function as a whole. For example, public policy may require government control over the provision of health care, but the operation of specific ambulance services or the construction of hospitals may be outsourced to the private sector without sacrificing that commitment. In the USA, federal and state governments frequently outsource even core state functions such as military and security services.

Deregulation consists of the opening of state-operated or state-controlled activities to private sector competition, allowing the government to retain its ownership but exposing the public entity to increased competition.

In addition to the main categories of privatization, there are also innumerable intermediate forms of privatization. For example, these might consist of public-private partnerships (PPPs) or private finance initiatives (PFIs), wherein an asset is financed entirely or partially by the private sector and the investment is recouped either through reimbursement from the government or through the accrual of profits from the operation of the service.

Historical Development of Privatization

For most of the twentieth century, until the 1970s, public ownership of key industries was a basic tenet of economic policy in each of the world's major economic blocs. In particular, public utilities such as energy, telecommunications, water, sewage, and transport were nearly universally state-owned industries. This began to change in the 1970's and 1980's, when new economic theories began to apply principles of market failure to weaknesses in public management, and the socialist economies of Central and Eastern Europe fell into rapid decline. By the end of the twentieth century, privatization had become a prevalent feature of public service provision in all of the world's major economies. Privatization is now recognised as a useful economic tool with different costs and benefits depending on the economic and social context.

In the liberal market economies of the USA and Western Europe, Keynesian macroeconomic theory prevailed from the late 1930s until the late 1970s. Keynesian economics emphasized the influence of aggregate demand on economic output and therefore the facilitating potential of government intervention in the economy. Keynesianism reached its zenith at the 1944 Bretton Woods Conference with the establishment of the global monetary system, the International Monetary Fund (IMF), and the World Bank. Following the "total war" state mobilizations of World War II, most major manufacturing and utilities industries in western economies were state-owned, and it was widely accepted that key industries should remain so. There were some important instances of privatization during this period, including the privatization of the British steel industry in the 1950s, and coordinated divestitures by the West German government in the 1960s (e.g., the sale of a majority state in Volkswagen in 1961). For the most part, however, public utilities and other important industries remained state-owned. This consensus was entrenched further as increasing prosperity after World War II led to the expansion of welfare states and social-market economies around the world. Keynesian economics informed the management of western market economies for much of the postwar economic expansion.

In Eastern and Central Europe, the rise of socialism sometimes led to near-total appropriation of enterprise and centralised planning of economies by the state during the postwar period. The constitutions of socialist countries based on the 1936 USSR Constitution, for example, stipulated that all productive assets (including firms) were property of "all the people" (i.e., the state). Private markets were severely constricted in eastern-bloc countries.

In South America, Asia, and Africa, encouraged by the new global Keynesian institutions and governments of both socialist and capitalist

countries, developing nations equally saw state control and investment as the best way to achieve fast "catch up" in modernization and industrialization (Parker 2000).

Until the 1970s, economic theorists in all the world's major economic blocs were therefore primarily concerned with identifying imperfections in private markets and making adjustments through public intervention. The reign of Keynesianism in the west and socialism in the east meant that economic thought to the 1980s was dominated by the literature on market failure.

Privatization only began to take root in developed countries during the ascendency of neoliberal economic theory in the late 1970s and early 1980s. Applying lessons of the market-failure literature to the economic performance of states, economists began to identify failures in the government's ability to perceive public wants, and further weaknesses in incentives to satisfy those wants efficiently. Economists such as Ronald Coase and Harold Demsetz questioned the prevailing logic that the mere existence of market failures warranted government intervention. Coase argued against government intervention without first analyzing whether such intervention would improve economic outcomes (Coase 1960), and Demsetz identified problematic inefficiencies and information asymmetries in government-managed organizations (Demsetz 1969, 1968). The essential defense of privatization which emerged from the literature was that the private sector was capable of providing public goods and services more efficiently than governments due to the disciplining effect of competition and profit motives. Transferring public functions to private, for-profit enterprises instilled incentives to cut costs and produce goods in response to demand. Exposure to competition and supply and demand principles were predicted to improve efficiencies in both production and allocation.

Neoliberal economic doctrines began to gain political traction in developed economies in response to the failure of Keynesian policy to deal with the economic malaise and stagflation of the 1970s. It was first enacted at the level of national policy in the UK under the administration of Prime Minister Margaret Thatcher (1979–1990), in the USA under the administration of President Ronald Reagan (1980–1988), and in Chile under the administration of President Augusto Pinochet (1974–1990). This was followed by other Western European economies within a decade. In the European Community in the 1990s, the abolition of customs barriers, the liberalization of national markets, common merger controls, and prohibitions on state aid facilitated the entry of private operators to areas that had previously been dominated by governments (Parker 1998). New European rules implied limits on public interventions. Portugal and France, for example, were required to amend their privatization legislation to allow foreign private investment.

In the transitional economies of Central and Eastern Europe, privatization took root as the result of an altogether different process: The collapse of Communism. Following the collapse of the USSR in 1989, privatization became a pillar of state policy in transitional economies. In Russia itself, two major waves of privatization occurred: The first wave, from 1992 to 1994, saw major state enterprises converted to joint-stock companies and put up for public auction, with Russian citizens given vouchers redeemable for stock in the new companies. A second wave of privatization saw the divestment of much of the energy and telecoms sector to private investors in the late 1990s and early 2000s (Leonard and Pitt-Watson 2013). By 1999, only 4% of registered companies in Russia were owned by the state (Munkholt 2000). Similarly, in post-Warsaw pact and other socialist countries, privatization occurred as both an economic and political reaction to the fall of communism. After the collapse, 90% of industrial capital in Eastern Europe remained in public hands (Lipton and Sachs 1990). Privatization thereafter took a variety of forms, often appearing as employee share purchases, public vouchers, or sales to cooperatives and private shareholders. Economic aims included increased efficiency generally, but often also aimed at increasing economic trade with Western Europe. Political aims included returning property to private owners who might have been deprived of it by the nationalization process under socialist governments.

In developing and emerging countries in Asia, Latin American, and Africa, impetus for privatization derived from two significant sources: As a precondition of foreign investment from developed market economies and as a condition for international aid and structural adjustment programs. Conditionality for financial assistance from the IMF, World Bank, and developed-country governments frequently requires states to meet certain objectives of economic liberalization or to pursue debt sustainability by denationalizing assets. Privatization policies have also often been used to counteract a perceived culture of political instability and to signal a commitment to free enterprise. At the dawn of privatization, between 1988 and 1993, the value of privatizations in developing countries reached US $96bn (World Bank 1995).

China has embraced limited privatization as part of its "Socialist Market System", while keeping political control and ownership of large, key industries in the economy. Yet it has not been immune to the growth of privatization. From 1978 to 1991, the state's share in industrial output fell from 77.6% to 52.9% (Wang 1994). In 1995, the State Council endorsed a policy of maintaining state control on large industries and allowing competition in small industries. The role of the state in industry has gradually declined as part of a measured and evolving process of opening up certain sectors of the economy to international investment without compromising political control of key economic sectors.

By the end of the twentieth century, the use of private actors in facilitating the operations of government had become ubiquitous in government policy around the globe. Privatization is no longer the exclusive domain of certain political systems or poles of the political spectrum. The discussion of governmental functions in liberal market economies is now suffuse with market-style language. Public service has been increasingly submitted to what have been called "New Public Management" regimes based on market-based theories (Hughes 2003). Private actors operate prison and immigration systems, determine regulatory standards for key industries, provide health care, and even conduct policing and military operations on behalf of governments.

Within this new global paradigm, the outcomes and theory of privatization have also begun to come under criticism. The results of privatization have often been mixed and context-conditional. Privatization has generated political opposition in many countries, and the theoretical tenets have come under challenge. In theory, policy, and practice, the extent to which privatization should be enacted in each county has become a much more nuanced and particulate debate.

The Economics of Privatization

The rationales for privatization are invariably couched in terms of economic efficiency. A number of rationales emerge from the literature (Parker 2000):

- Increased economic efficiency by introducing private competition and profit motives to what might otherwise be inefficient, monopolistic, and politicized operations under the public sector
- The reduction of government debt by selling financially burdensome assets or operations
- Increased welfare through increased responsiveness to supply and demand and cheaper goods from competition
- The avoidance of capture by politicians overseeing the provision of services by special interest groups such as trade unions or professional bodies
- Cost savings in accomplishing public works by allowing private investment to supplement or displace government financing
- Increased private ownership of capital and the fostering of liberal markets
- The attraction of mobile capital and foreign direct investment in an increasingly globalized world

Broadly, each of the efficiency arguments for privatization is based on two tenets of economic reasoning: (1) Efficiency gains from competition between firms and (2) efficiency gains from information and incentives within private firms (Dunsire et al. 1988; Vickers and Yarrow 1988). Both sources of reasoning stem from the premise that governments do not always pursue maximum

welfare and that imperfections in markets do not automatically justify government intervention.

Economic Efficiencies from Product Market Competition

Economic efficiency is comprised of allocative efficiency and production efficiency. Allocative efficiency occurs when producers produce only those types of goods that are useful and in demand in society, and where the price is equal to marginal cost. In essence, it is concerned that resources are not expended on less-optimal products, and that the most optimal distribution of resources is achieved. Production efficiency occurs when products are created using the least resources possible for the most efficient level of production. Production efficiency in turn has two constituent elements: technical efficiency and price efficiency. Technical efficiency is concerned with the effectiveness of inputs in creating an output. So, for example, a firm which employs too many workers or has underutilized equipment is technically inefficient. Price efficiency is the degree to which a price reflects all available supply and demand information. So, for example, charging a high price for a product that is in low demand would be price inefficient.

Under conditions of competition, firms will seek to optimize allocative and production efficiency, and price will reflect the most efficient marginal costs available. In essence, firms will strive to create the most useful products for the least cost, and charge prices that reflect supply and demand, maximizing economic well-being (Kirzner 1997). Without competition, a firm may restrict output and raise prices. It will not need to pursue technical or price efficiency. Where the monopoly is state-owned, this applies to governments. Governments which own firms have incentives to eliminate competition by regulating the market, subsidizing inefficient industry, or erecting legal barriers to entry (Littlechild 1986).

Efficiencies from Private Firm Ownership

The second source of efficiency from privatization derives from principal-agent theory. Principal-agent theory concerns the information asymmetries between principals (owners) and agents (managers) of firms. Under this theory, the private sector is considered to have more effective mechanisms of control in place between those who own the firm and those who drive its performance. The first and most obvious reason for this is that incentives for management performance in the private market are often unrestrained by politics and public law. Management may be generously incentivized or quickly demoted more easily than in the public sector, where set pay scales, public unions, and political costs may restrict action. Secondly, a private capital market is considered to have more effective means of control over managers. Managers and boards of directors can be pressured to resign by shareholders. A firm which is losing the value of its shares may see investors sell their shares, further precipitating a drop in the firm's share price. If the share price drops low enough, the firm becomes susceptible to takeovers by more efficient firms, who may then purchase the firm's asset and turn them to more productive uses. In essence, it is argued that inefficient management cannot survive in the private market.

In the public sector, agents are responsible to politicians. The literature on public choice indicates two problems with governments as principal: The first is that the objectives of government are some combination of social welfare, the politician or government's own personal goals, and the aims of special interests groups (Niskanen 1971). Even if weighted only slightly towards the latter two goals, this represents a suboptimal welfare result. Politicians may be subject to political capture and require firms to pursue secondary objectives such as industry employment, high wages, and other aims which may run counter to productive efficiency. These wider social objectives may genuinely contribute to the public welfare; however the literature on public choice suggests that governments are no less susceptible to rent-seeking than private monopolies (Buchanan 1972). For example, a politician elected due to the support of labor interest groups may be insulated from citizen pressure and instead be incentivized to promote high wages in the industry instead of efficient outputs that benefit public welfare as a whole. The second problem is

that, even if the government solely pursues the maximization of efficiency, governments lack information on demand and supply costs. Only under conditions of competition is demand information readily available.

Legal Controls and Methods of Privatization

The divestiture or delegation of basic public services to private actors can raise legal questions both as to legal methods of control on the actions of government and what legal forms privatization may take.

Legal Controls on Privatization

Privatizations are typically subject to three types of legal control: Constitutional law, administrative law, and international or human rights law.

Constitutional law places limits on the existence and exercise of public power. This can include limits on the government's power to privatize. Article 34 of the French Constitution of 1958, for example, states that rules governing the transfer of public assets to the private sector shall be set by law. Privatizations are therefore required to be approved by the legislature. The constitutional principle of equality among citizens and Article 17 of the Declaration of Human Rights (which requires just compensation for expropriation) has been held by the French Constitutional Council to prohibit the sale of public assets at less than the real value. Privatization transactions are therefore justiciable both on the basis of the separation of powers and administrative value. In other countries, such as the UK, the doctrine of parliamentary sovereignty and the discretionary powers of government mean that divestitures by either the legislature (by statute) or the executive (by contract) will not be subject to judicial review without an explicit statutory basis. In the EU, the European Court of Justice has long accepted the existence of an inherent power to delegate where necessary to exercise a power conferred by the Member States, subject to the limit that privatizations on that basis cannot exceed the limits of the power conferred (*Meroni v. High Authority* [1957–58] ECR 133). In the USA, the Fifth and Fourteenth amendments to the US Constitution contain "due process" clauses which stipulate basic elements of fundamental fairness such as the opportunity to be heard and impartiality in decision-making. The courts have derived a constitutional doctrine of non-delegation from these clauses that requires governments to comply with due process and the separation of powers when delegating.

Administrative law seeks to ensure fairness and compliance with the rule of law in the exercise of state power. In both civil law and common law countries, the legislature may set out detailed codes on administrative procedures and public decision-making. In the USA, federal and state legislation encompass a range of procedural acts. The Federal Procedure Act 1946, for example, offers equal or greater procedural protection than even the due process clause of the constitution. In the EU, conditions and limits on the delegation of state power are regulated in detail by Financial Regulation 1081/2010 [2010] OJ L311/9, which sets out the tasks that can be delegated to private entities or agencies. In the USA and UK, the common law principles of judicial review grant the courts a jurisdiction to review the actions of government decision-makers against standards of procedural unfairness, transparency, bias, and rationality (Taggart 1997).

International and human rights law seeks to protect individuals against the inappropriate use of state power. This becomes particularly relevant to privatization when human rights principles contain procedural protections. For example, Article 6 of the European Convention on Human Rights (ECHR) enshrines the right to a public hearing before an independent and impartial tribunal where civil rights are being determined. It is not difficult to imagine how privatization of a social function such as welfare or public housing could have the potential to deprive the individual of the right to an impartial tribunal against determinations of

their civil rights by the private body. The existence of profit motives of a private operator may provide an incentive to disqualify recipients of welfare or housing benefits, for example. Indeed, in *Feldbrugge v. Netherlands* (1986) 8 EHRR 425, the European Court of Human Rights held that such benefits may give rise to civil rights.

Legal Methods of Privatization

In legal terms, the three main techniques for privatization are legislation, contracts, and legal grants (Donnelly 2007). Each method is subject to a different mixture of legal controls in each state.

Privatization by legislation is, in legal terms, perhaps the most conventional method for privatization. Legislative acts are typically justiciable under general constitutional law principles and are therefore likely to be subject to judicial review principles. The same will generally be true of a privatization engaged in by the executive, subject to rules set out in legislation. In many jurisdictions, however, there are also countervailing limits to the legal control of privatization by legislation. In the UK, the doctrine of parliamentary sovereignty precludes judicial review of legislative acts unless provided explicitly by statute. In practice, statutes granting executive powers are typically drafted broadly and are unlikely to preclude privatization where it may be necessary for the function given.

Privatization by contract transfers the source of the private entity's legal responsibility from an act of the legislature to a private contract. Whereas executive and legislative action is bound by constitutional, administrative, and international law, a private party governed by private contract is typically governed by the terms of the contract. In the UK, decisions by private actors operating under contract are not typically subject to judicial review and administrative law principles, even when exercising a public power (Freedland 1994). In the USA, the definition of "state actor" to which constitutional and administrative procedural rules apply has evolved to exclude most private actors acting under a contract. Private actors are exempt from disclosure, oversight, bias, and ethical obligations which apply to public bodies under the US Code of Laws, for instance. Similarly, international human rights obligations such as the ECHR bind state parties only, despite the fact that private entities exercising state power may exercise the same capacity to interfere with individual rights. In the UK, for example, despite the fact that the Human Rights Act 1998 applies to "functions of a public nature", the public or private nature of the entity has tended to predominate, rather than its function. Since the legal controls on the private actor are often limited to the scope of the contract, its drafting will have a significant impact on the government's ability to control outcomes. On one hand, government contracts typically fall within the scope of public procurement regulations, and therefore may allow or require governments to include public-service objectives in pecuniary contracts. On the other hand, many government services are highly variable and it may be inappropriate to limit the criteria for performance evaluation. A contract which pays a prison service on a per head basis, for example, may instate a perverse incentive to extend sentences. Yet the more complex the contract, the more difficult the privatization is to monitor, and the less benefit is gained from privatization.

Legal grants consist of the granting of a legal right or funding to build or operate an asset. In contrast with contracts, grants are typically governed by authorizing legislation and often fall outside the scope of legislated procurement regimes. There is often a broader latitude given to the recipient as to how to achieve the aims of the privatization. This can result in an arbitrary difference in legal treatment between contracts and grants (Donnelly 2007).

Criticisms of Privatization

Economic Criticisms

Privatization theory has come under pressure from the emergence of new theories such as New

Institutional Economics theory, which question the predictive power of privatization models and attempts to demonstrate that the optimal allocation of functions cannot be determined a priori in all cases. Privatization is not universally beneficial, and the optimum allocation of ownership often depends on the economic and social attributes of a specific country. Some contemporary economists, such as Joseph Stiglitz, argue that government can generally outperform private enterprise in economic outcomes (Stiglitz 1995). Critics argue that privatization should only take place where the market could perform as well as the benevolent government (Sappington and Stiglitz 1987). This criticism centers around two main heads:

First, markets are often imperfect, and privatization only generates benefits under specific market conditions. The choice is not always between government monopoly and fully competitive capital markets with effective competition. Many state-owned industries maintain large economies of scale (such as rail networks or telecommunication towers) which they then take with them to the private market to create a private monopoly. There is also evidence that private monopolists may be more harmful than public monopolists (Pint 1991). The OECD concludes that private sector monopolists are much more efficient under conditions of competition but also much better at extracting rents under monopolistic conditions (OECD 2000). They may also be less susceptible to regulation: attempts to regulate them may simply result in the private firm passing on the cost of regulation to consumers without bearing any cost itself (De Fraja 1993). Some economists argue that competition is the crucial factor in efficiency, rather than the public or private nature of the owner (Milward and Parker 1983; Vickers and Yarrow 1988).

Second, capital markets may not make for more efficient principal-agent control. Private sector management may still be able to pursue their own objectives to a large degree without shareholder reaction, and a loss of shareholder value may not necessarily lead to takeovers or share sell-offs.

The empirical outcomes of privatization have not always shown efficiency improvements, particularly where privatization has merely resulted in the transfer of a public-sector monopoly to a private-sector monopoly. While there is strong evidence that privatization lowers prices and improves services where markets work effectively, the outcomes can often be the opposite where the privatization did not result in competition (OECD 2000). In other cases, critics have argued that state-owned assets and rights have been sold at an under-value, with no residual government ability to regulate the industry and ensure pareto outcomes for the citizens. The optimum allocation of private and public control is now a matter of theoretical, empirical, and political debate in individual countries and economies. Economists in favor of privatization have continued to argue that the private sector is a more efficient provider of most government functions. For example, Megginson observes that "private ownership must be considered superior to state ownership in all but the most narrowly defined fields or under very special circumstances" (Megginson 2005). Yet other authors argue that privatization increases costs just as often as it reduces them (Brudney et al. 2004). In a survey of the literature, Parker concludes "Empirical evidence, like the economic theory, does not allow us to be confident of predicting accurately the outcome of any privatization on corporate performance" (Parker 2000).

Legal Criticisms

While the scope of public law varies between jurisdictions, government action is often subject to legal constraints under constitutional law, administrative law, and human rights law, which private actors are not. This may allow public actors to divest themselves of their legal obligations. For example, under EU law, the actions of public institutions may be challenged on the basis of the principles of proportionality and non-discrimination, where private institutions may not. In the USA and the UK, the common law principles of judicial review grant the courts a jurisdiction to review the actions of government decision-makers against standards of unfairness,

transparency, bias, and rationality (Taggart 1997), but common law judicial review does not apply to private actors. The principle of bias, for example, forms part of the natural law principles of procedural fairness at common law. It states that where a decision-maker has a pecuniary interest in a decision, the decision may be void for bias. Yet for-profit private actors (for example, particularly those responsible for social welfare or prison systems) will nearly always have a pecuniary interest in the outcome of decisions which it has been delegated.

In civil law countries as well as common law countries, public law sets out detailed codes on administrative procedure and decision-making, but private law remains a matter of contract. This makes courts and legislators slow to recognize and enforce public law duties on private parties in most jurisdictions, even when those private parties are wielding governmental power (Donnelly 2007). Even under the most effective international human rights regimes, such as the European Convention on Human Rights, individual human rights are enforceable only against the state save in exceptional or indirect cases. Such regimes do not typically apply to the actions of private parties, despite the fact that private parties exercising state power may enjoy the same capacity to interfere with individual rights.

Political Criticisms

Political opposition derives primarily from five outcomes: (1) Private restructuring of labor markets and labor retrenchments may result in worker vulnerability and increased unemployment; (2) privatization may result in the discontinuation of less-profitable services; (3) consumer cost increases may result from the withdrawal of government support or private market imperfections; (4) governments may lose control of outcomes in which there is a strong public interest; and (5) privatization may sacrifice the long-term value of assets for short-term gain, because while there are inevitably a finite amount of public assets, there is an infinite capacity for debt growth.

Political opposition therefore derives from both empirical and ideological observation. Underlying each of these criticisms is the implication that the treatment of citizens as consumers is often inappropriate. First, at an ideological level, citizenship may imply something more than consumerism – it implies a stake in the state itself. For example, as citizen-consumers, the electorate may accept the privatization of postal services provided that efficiency gains are demonstrated, but the electorate as citizens may not accept the privatization of health or military services, regardless of efficiency. The National Health Service in the UK is an example of a public asset that is unlikely to be privatized due to strong public support. Second, at an empirical level, consumers exercise their influence on the state through the exercise of choice. Yet when it comes to public services, citizens will not often have a choice. Welfare recipients or those needing immediate health care, for example, may not have a choice as to the most efficient provider.

References

Brudney JL, Fernandez S, Ryu JE, Deil W (2004) Exploring and explaining contracting out: patterns among the American States. J Public Adm Res Theory 15:393

Buchanan JM (1972) Theory of public choice. University of Michigan Press, Ann Arbor

Coase R (1960) The problem of social cost. J Law Econ 2(1):1–44

Demsetz H (1968) Why regulate utilities? J Law Econ 11(1):55

Demsetz H (1969) Information and efficiency: another viewpoint. J Law Econ 12(1):1

Donnelly C (2007) Delegation of governmental power to private parties. Oxford University Press, New York

Dunsire A, Hartley K, Parker D, Dimitriou B (1988) Organisational status and performance: a conceptual framework for testing public choice theories. Public Adm 66(4):363

Fraja D (1993) Productive efficiency in public and private firms. J Law Econ 50:15

Freedland M (1994) Government by contract and public law. Public Law 86:101

Hughes OW (2003) Public management and administration: an introduction. Palgrave Macmillan, Basingstoke

Kirzner IM (1997) How markets work: disequilibrium, entrepreneurship and discovery. Hobart Paper 113, Institute of Economic Affairs, London

Leonard CS, Pitt-Watson D (2013) Privatization and transition in Russia in the early 1990s. Routledge, New York

Lipton D, Sachs J (1990) Privatization in Eastern Europe: The case of Poland. Brook Pap Econ Act 2:293

Littlechild SC (1986) The fallacy of the mixed economy: an 'Austrian' critique of recent Economic thinking and policy. Hobart Paper 90, Institute of Economic Affairs, London
Megginson W (2005) The financial economics of privatization. Oxford University Press, Oxford, pp 52–96
Milward R, Parker D (1983) Pubic and private enterprise: comparative behaviour and relative efficiency. In: Milward D, Parker L, Rosenthal MT, Topham N (eds) Public sector economics. Longman, London
Munkholt P (2000) Russia seeking a soft landing. Privat Int 137:7
Niskanen WA (1971) Bureaucracy and representative government. Aldine, Chicago
OECD (2000) Privatisation, competition and regulation. OECD, Paris, pp 10–11
Parker D (1998) Privatization in the European Union: theory and policy perspectives. Routledge, London
Parker D (2000) Privatization and corporate performance. Edward Elgar Publishing, Cheltenham, pp xii–xv
Pint E (1991) Nationalization vs regulation of monopolies: the effects of ownership on efficiency. J Public Eco 44(2):131
Sappington D, Stiglitz J (1987) Privatization, information and incentives. J Pol Anal Manag 6(4):110–125
Stiglitz J (1995) Whither socialism? MIT Press, Cambridge
Taggart M (1997) The province of administrative law determined? In: Taggart M (ed) The province of administrative law. Hart, Oxford, pp 1–3
Vickers J, Yarrow G (1988) Privatisation: an economic analysis. MIT Press, Cambridge, MA
Wang S (1994) The compatibility of public ownership and the market economy: a great debate in China. World Affairs 157(1):38
World Bank (1995) Bureaucrats in business: the economics and politics of government ownership. Oxford University Press, Oxford

Product Promotion Strategy

► Promotional Effort

Productivity and Growth

Jaime Vallés-Giménez and Anabel Zárate-Marco
University of Zaragoza, Zaragoza, Spain

Abstract

Economic growth is a long-run process that occurs when an economy's potential output increases, and it can be measured by the product method, the income approach or the expenditure method. Actual growth is the percentage annual increase in national output. Potential growth is the speed at which the economy could grow, i.e. the percentage annual increase in the economy's capacity to produce and it can be shown by an outward shift in the economy's production possibility frontier. The theory and empirical studies suggest that potential economic growth is associated with the increase in the use of factors of production (capital, labor, energy, etc), but primarily to increases in productivity or efficiency with which these factors are used, through advances in labor skills and organization of production or improves in technology. Productivity and economic growth are then closely linked because economic growth occurs when productivity increases to allow for such growth. Productivity is therefore the cornerstone of economic growth.

Productivity is an indicator of the efficiency of production and can be defined as the ratio of output to inputs in production. Higher productivity means that the economy can produce more goods and services at a lower cost per unit. This will help to reduce prices and increase consumer welfare and living standards, because more real income improves people's ability to purchase goods and services, enjoy leisure, improve housing and education and contribute to social and environmental programs. Higher productivity increase total output from the scarce factor resources, causing an outward shift of the production possibility frontier. Productivity also affects our competitive position: the more productive we are the better we are able to compete on world markets. Productivity growth also helps businesses to be more profitable. There are broadly two ways of measuring productivity. On one side are the partial productivity measurements that relate to an input (labor, capital, etc), and on the other side it is a measure of total factor productivity (TFP) or multifactor productivity (MFP), which measures the effects in total output not caused by measured inputs of labor, capital and intermediate outputs.

Definitions

Economic growth is a long-run process that occurs when an economy's potential output increases, i.e., as economy's ability to produce goods and services rises.

Productivity is an indicator of the efficiency of production.

Concepts and Connections Between Growth and Productivity

The fruit of economic activity is the amount of goods and services produced by labor, capital, and other inputs. So arguably, **economic growth** is the increase in the market value of the goods and services produced by an economy over time. It is conventionally measured as the percent rate of increase in real gross domestic product, or real GDP, i.e., (real GDP_t – real GDP_{t-1})/realGDP_{t-1}.

Since GDP is a macroeconomic variable, i.e., it is the result of multiplying the quantities of goods and services produced in one country by their prices, we will only have a proper idea of the growth of an economy's production if we eliminate the distorting effect of inflation on the price of goods produced, and the evolution of real output is analyzed.

Another element to consider in economic growth is the increase in population. Only if the population increase is known can it be determined whether the per capita product increases or not. For this reason, of more importance is the growth of the ratio of real GDP to population (real GDP per capita).

On the other hand, as the GDP figures of each country are measured in its local currency, they have to be converted into a common currency (e.g., dollars or euros) at the current exchange rate, to be able to be compared. But the exchange rate may be a poor indicator of the purchasing power of the currency at home. To compensate for this, GDP can be converted into a common currency at a *purchasing-power parity rate*. This is a rate of exchange that would allow a given amount of money in one country to buy the same amount of goods in another country after exchanging it into the currency of the other country.

Besides this method of measuring GDP, known as the *product method*, GDP can be calculated in other two different ways. The production of goods and services generates incomes for households in the form of wages and salaries, profits, rent, and interest. Therefore, GDP can also be calculated by adding up all of the income received by labor and other inputs in the economy. This is known as the *income approach*. The third method, expenditure method, focuses on the expenditures necessary to purchase the nation's production by different groups in the economy. The four main components are consumption expenditures by households, gross private investment spending principally by firms, government purchases of goods and services, and net exports (exports minus imports). Because of the way the calculations are made, the three methods of calculating GDP must yield the same result.

It is essential to distinguish between actual and potential economic growth. Actual growth is the percentage annual increase in national output: the rate of growth in actual output (GDP). Potential growth is the speed at which the economy could grow. It is the percentage annual increase in the economy's capacity to produce: the rate of growth in potential output. An increase in an economy's productive potential can be shown by an outward shift in the economy's production possibility frontier.

If the potential growth rate exceeds the actual growth rate, there will be an increase in spare capacity and probably an increase in unemployment: there will be a growing gap between potential and actual output. To close this gap, the actual growth rate would temporarily have to exceed the potential growth rate. In the long run, however, the actual growth rate will be limited to the potential growth rate. Although growth in potential output varies to some extent over the years – depending on the rate of advance of technology, the level of investment, and the discovery of new raw materials – it nevertheless tends to be much more steady than the growth in actual output. Actual growth tends to fluctuate. In some years, countries will experience high rates of

economic growth. In other years, economic growth is low or even negative. This cycle of booms and recessions is known as the business cycle or trade cycle. The business cycle moves up and down, creating fluctuations around the long-run trend in economic growth.

Anyway, these measures of growth do not determine economic development as this is a more complex concept as it has social connotations also related to the improvements of living standards in the country. They ignore the distribution of income. Although per capita real income may be increasing in a country, this does not necessarily mean that all the inhabitants of that country are benefiting from this improvement. Maybe along with the growth of real income it takes place a change of income that impoverish certain people while others enjoy her growth above average. While economic growth is necessary, it is not sufficient for progress on reducing poverty. For this reason, the economic growth may be a poor indicator of society's well-being.

Moreover, it does not compute the human costs of production. If production increases, this may be due to technological advance. If, however, it increases as a result of people having to work harder or longer hours, its net benefit will be less. Leisure is a desirable good and so too are pleasant working conditions, but these items are not included in the economic growth figures.

Another problem in the computation of economic growth is some external benefits or costs that are not included in GDP statistics. For example, growth has the disadvantage that can both create negative externalities, e.g., higher levels of noise pollution and lower air quality arising from air pollution and road congestion and causes depletion of resources. And finally, GDP only measures the market economy, thereby excluding "do-it-yourself" and other home-based activities as well as the underground economy, so the GDP statistics understate the true level of production in the economy.

Since the pioneering work of Solow (1956), the theory and empirical studies suggest that potential economic growth is associated with two factors: the increase in the use of factors of production (capital, labor, energy, etc.), but primarily the increase in productivity or efficiency with which these factors are used, through advances in labor skills and organization of production or improvements in technology. Productivity and economic growth are then closely linked because economic growth occurs when productivity increases to allow for such growth. Productivity is therefore the cornerstone of economic growth.

Productivity can be defined as the ratio of output to inputs in production. It is an average measure of how efficiently goods and services are produced. Higher productivity means that the economy can produce more goods and services at a lower cost per unit. This will help to reduce prices and increase consumer welfare and living standards, because more real income improves people's ability to purchase goods and services, enjoy leisure, improve housing and education, and contribute to social and environmental programs. Higher productivity increase total output from the scarce factor resources, causing an outward shift of the production possibility frontier. Productivity also affects our competitive position: the more productive we are the better we are able to compete on world markets. Productivity growth also helps businesses to be more profitable.

There are broadly two ways of measuring productivity. On one side are the partial productivity measurements that relate to an input (labor, capital, etc.), so we can say that there are so many measurements of productivity as resources used in production. This partial productivity has usually been measured in terms of labor, by the availability of data. Labor productivity is then the value of goods and services produced in a period of time, divided by the hours of labor used to produce them. In other words, the labor productivity measure the output produced per unit of labor, usually reported as output per hour worked or output per employed person.

However, partial productivities do not show overall efficiency of the use of all the resources, so it is important to have a simultaneous measurement of the efficiency of the totality of resources, i.e., a measure of total factor productivity (TFP) or multifactor productivity (MFP). TFP measures

the effects in total output not caused by measured inputs of labor, capital, and intermediate outputs. If all inputs are accounted for, total factor productivity (TFP) represents improvements in ways of doing things, that is to say, it can be taken as a measure of an economy's long-term technological change or technological dynamism, which is the primary source of real economic growth. In the short term, however, also reflect unexplained factors such as cyclical variations in labor and capital utilization, economies of scale, and measurement error. Total factor productivity is the most commonly known and widely used method of productivity measurement. However, TFP cannot be measured directly but is a residual. It accounts the residual growth that cannot be explained by the rate of change in the inputs.

Sources of Productivity and Economic Growth

Despite the lack of a unifying theory, there are several partial theories that discuss the role of various factors in determining long-term economic growth. Two main strands can be distinguished. The neoclassical model based on the growth model of Solow (1956) has emphasized the importance of investment. And the more recent theory of endogenous growth developed by Romer (1986, 1990) and Lucas (1988) has drawn attention to human capital and innovation capacity. Furthermore, important contributions on economic growth have been provided by the cumulative causation theory of Myrdal (1957) and by the New Economic Geography school of Krugman (1991). These two schools assert that economic growth tends to be an unbalance process favoring the initially advantaged economies. In addition, other explanations have highlighted the significant role that non-economic (in the conventional sense) factors play on economic performance.

The theoretical and empirical studies about the growth theories have been plentiful and have used differing conceptual and methodological viewpoints (especially interesting is the review of Helpman 2004). These studies have placed emphasis on different explanatory parameters and have offered various insights to the sources of economic growth. The review of this literature suggests that no single policy or factor leads to productivity growth. Rather, it is a matter of getting a lot of interconnecting things right, and by ensuring incentives are aligned, creating an environment where firms can create and take advantage of opportunities. The main factors that may influence the long-term economic growth are as follows:

1. *Physical capital*. The rate of accumulation of physical capital is one of the main factors determining the level of real output per capita although its effects could be more or less permanent depending on the extent to which technological innovation is embodied in new capital. Investment in physical capital is a key determinant of economic growth identified by both neoclassical and endogenous growth models. However, in the neoclassical model, physical capital has impact on the transitional period, while the endogenous growth models argue for more permanent effects. Whatever the transition mechanism from capital accumulation to growth, the significant differences in the investment rate across countries, and over time, point to it as a possible source of cross-country differences in output per capita.

   Physical capital includes factories, tools, computers, machinery, production equipment, and structures such as infrastructure or fixed social capital that are often the result of investments made by the state. These infrastructures range from transport infrastructure (roads, airports, ports, railways), energy, telecommunications, to universities, hospitals, water projects, and other public health measures, e.g., diseases control, etc.

   The more capital workers have at their disposal, generally the better they are able to do their jobs, producing more and better quality output. The role of infrastructures is to expand production, to increase resources, and to enhance the productivity of private capital. A good transport system enhances

cohesion and improves access to outlying regions through the reduction of transport costs for both goods and people traveling for leisure or work. Telecommunications are the modern substitute for the connections made through transport and a prerequisite for the development of industries and modern services that rely on the phone, fax, and data transmission systems. An undersized or inadequate infrastructure, such as an electrical network with frequent failures, cuts electricity to homes and businesses is a major obstacle to economic growth in some countries. Without necessary infrastructure, it can be difficult for firms to be competitive in the international markets. The lack of infrastructure is often a factor holding back some developing economies

2. It is not enough that a worker has good equipment; he must also know what to do with it and how to use it in the safest, most effective, and efficient manner. For this reason, *human capital* is the main source of growth in several endogenous growth models as well as one of the key extensions of the neoclassical growth model, because of its role as a facilitator of both technology adoption from abroad (absorption capacity) and the creation of appropriate domestic technology (innovation). It would be impossible to operate the current economy with a population with the literacy levels and formation of a century ago.

   The term human capital refers both to the improvement and training of manpower produced by the education and knowledge that is incorporated into the work force and by the learning by doing, as well as to the improvements in their health. A population that is well educated and well trained helps a society to increase its ability and acquire as well as use relevant knowledge (absorption capacity). As knowledge is created by a small number of leader countries in technological terms and most countries do not produce state-of the-art technology themselves, these latter countries must acquire the technology from elsewhere via trade or foreign direct investment. Focusing on education, basic education would be important for learning-capacity and utilizing information, while the higher education would be necessary for technological innovation.

   In turn, we must bear in mind that in a population with good health, workers can be more productive and learning ability of children to be greater. A longer life expectancy makes it more attractive to invest in human capital and even foreign direct investment, and savings incentive and productivity can be increased. The governments can also play an important role in the accumulation of human capital from the time they can invest in education.

3. Probably the most important factor for productivity growth and therefore economic growth is the *innovation and technological progress*. In practice, most technological improvements are due to deliberate actions, such as research and development (R&D) carried out in research institute or firms.

   R&D evolves new ideas and designs and is used by firms in search for blueprints of new varieties of products or higher-quality products. New ideas can also take the form of new technologies, new products, or new corporate structures and ways of working. Expenditure on R&D can be considered as an investment in knowledge that translates into new technologies as well as more efficient ways of using existing resources. Such innovations contribute to the expansion of the so-called frontiers of knowledge, and the accumulation of knowledge will generate growth. Hence, technological change emerges from technical innovations generated by research and development, patenting and software, and productivity enhancing developments in the fields of education management and marketing. And workers today are capable of producing more than in the past, even with the same amount of physical and human capital, because the technology has advanced over time.

   The amount of resources that are devoted to R&D can be influenced by government

intervention. In particular, the potential benefits from new ideas may not be fully appropriated by the innovators themselves due to spillover effects, which imply that without policy intervention the private sector would likely engage in less R&D than what could be socially optimal. This can justify some government involvement in R&D, both through direct provision and funding, but also through indirect measures such as tax incentives and protection of intellectual property rights to encourage private-sector R&D.

4. The degree of *openness* of the economy also affects productivity and economic growth. It does through several channels such as exploitation of comparative advantage, technology transfer and diffusion of knowledge, increasing scale economies, and exposure to competition. Trade liberalization promotes competitiveness, efficiency of input, creates incentives to innovate, and ensures that resources are allocated to the most efficient firms. It also forces existing firms to organize work more effectively through imitations of organizational structures and allow to introduce foreign (relatively advanced) technology into domestic production, which in turn has a positive effect on productivity and economic growth. To have a good absorption capacity is a key factor for a good use of the technology transferred. In particular, certain kinds of imports, namely, machinery and equipment relating to foreign R&D, are expected to generate a lot of technology transfer because the technology is often embodied in goods. Moreover, trade liberalization increases investment opportunities and international contacts.
5. *Foreign Direct Investment* can play a crucial role of internationalizing economic activity and can stimulate economic growth by improving technology and productivity. Host economies are expected to benefit from the positive externalities driven by foreign direct investment. Those include knowledge spillovers generated by technology transfers, introduction of new processes, and managerial skills and know-how diffusion to the domestic market.
6. *Macroeconomic stability-oriented policies* can also have a significant impact on economic growth. A stable macroeconomic environment characterized by low and predictable inflation, sustainable budget deficits, and limited departure of the real exchange rate from its equilibrium level, sends important signals to the private sector about the commitment and credibility of a country's authorities to efficiently manage their economy and increase the opportunity set of profitable investments.

   The usual arguments for lower and more stable inflation rates include reduced uncertainty in the economy and enhanced efficiency of the price mechanism. Uncertainty related to higher volatility in inflation could discourage firms from investing in projects that have high returns but also a higher inherent degree of risk. Moreover, large fiscal deficits and high net international debt position make a country vulnerable to global financial shocks and terms of trade shocks (e.g., oil price spikes)
7. *Institutions* also matter for economic growth because they establish the rules of the game in a society or, more formally, they set the humanly devised constraints that shape human interaction (North 1990). Institutions decide how to organize the societies and so determine whether or not the economy prospers. Certain forms societies encourage people to innovate, to take risks, to save for the future, to find better ways of doing things, to learn and educate themselves, solve problems of collective action, and provide public goods, while others do not. Institutions are therefore a key factor for potential economic growth.

   Institutions encompass both informal constraints as customs, traditions, codes, or taboos and formal constraints as property rights, legal rules, constitutions, contract enforcement, the political system, and electoral rules. Also encompass public sector imperfections, the degree to which laws and

regulations is fairly applied, the extent of corruption, etc.

Economic institutions are the set of norms relating to production, allocation, and distribution process of goods and services. They can guarantee the rights of intellectual and industrial property so necessary for there to be greater efficiency in the use of resources and also greater technological progress and innovation leading to economic growth. They must also guarantee some degree of equality of opportunity in society, including such things as equality before the law, so that those with good investment opportunities can take advantage of them. Economic institutions also help to stabilize the economy and ensure the proper functioning of the financial system, so it is necessary to have high rates of savings and investment, a good allocation of resources, specialization of the economy, and creating incentives. A well-developed financial system provides funding for capital accumulation, helps the diffusion of new technologies, mobilizes savings by channeling small savings of individuals into profitable large-scale investments, while offering savers a high degree of liquidity.

Economic institutions can also remove or reduce the tariff barriers and the impediments to foreign investment so they may favor a greater exposure to international competition. They may enforce contracts, discourage unfair or abusive business practices, limit the power of rulers, protect individuals both from one another and from the state, increase safety at work, contribute to public health and safety and the development of a more productive and fairer society, set limits on pollution, help protect consumers from potentially hazardous products, ensure that they are able to make informed choices, and influence investments in physical and human capital and technology and the organization of production.

Economic institutions are linked to political institutions because the latter are necessary to the former work. Political institutions are those that determine the structure of the state and the procedures of the political decision-making process. Political institutions shape the political process that produces legislation and regulation. They also determine the legal system and coordinate the processes that create and enforce the law. Political institutions therefore produce economic institutions and determine their quality. Institutions like democracy and social protection legitimize market outcomes and ensure their endurance. Political institutions can support a market economy by shaping and safeguarding property rights and making the market compatible with social stability and social cohesion. A stable and corruption-free government, a strong independent judiciary, efficient bureaucracy, and political constraint on executive are keys to generate certainty and thus economic growth, since political instability would increase uncertainty, discouraging investment, and eventually hindering economic growth.

8. Certain structural characteristics such as *geographical conditions* are also a powerful driver of economic growth. The country's location, its topography, and access to the sea affect the transport costs and the efficient allocation of resources because where geography is not propitious, the diffusion of technology is more complicated and more costly trade. Thus, it also affects the productivity and competitiveness.

   Moreover, the existence of natural resources in abundance is essential for economic growth. The natural resources of a country such as minerals and oil-resources, soil quality, forest wealth, and good climate river system affect the returns to agriculture and its economic structure. A country deficient in natural resources may not be in a position to develop rapidly, although natural resources are a condition for economic growth necessary but not sufficient to one.
9. *Demographic factors* like population growth, population density, migration, and age distribution can play a major role in economic growth.

The population growth can undermine per capita economic growth, and moreover, composition of the population has important implications for growth. A large working-age population is deemed to be conductive to growth, whereas population with many young and elderly dependents is seen as impediment because influences the dependency ratio, investment and saving behavior, and quality of human capital. Population density, in turn, may be positively linked with economic growth as a result of increased specialization, knowledge diffusion, and so on. Migration would affect growth potential of both the sending and receiving countries.

10. Other factors of *sociocultural nature*, e.g., ethnic diversity or cultural diversity may affect growth although in a much more residual, indirect, and unclear manner. For instance, it may have a negative impact on growth due to emergence of social uncertainty or even of social conflicts or a positive effect since it may give rise to a pluralistic environment where cooperation can flourish.

References

Helpman E (2004) The mystery of economic growth. The Belknap Press of Harvard University Press, Cambridge, MA

Krugman P (1991) Increasing returns and economic geography. J Polit Econ 99:183–199

Lucas R (1988) On the mechanics of economic development. J Monet Econ 22:3–42

Myrdal G (1957) Economic theory and underdeveloped regions. Hutchinson Publications, London

North DC (1990) Institutions, institutional change, and economic performance. Cambridge University Press, New York

Romer P (1986) Increasing returns and long run growth. J Polit Econ 94(2):1002–1037

Romer P (1990) Endogenous technological change. J Polit Econ 98(5):S71–S102

Solow RM (1956) A contribution to the theory of economic growth. Q J Econ 70(1):65–94

Professional Guides

► Codes of Conduct

Prohibition

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

Abstract

A prohibition is a government decree against the production and exchange of a good or service. Recent studies on prohibitions, for a variety of goods and services, such as drugs, alcohol, and prostitution, suggest that prohibitions impose heavy costs and are extremely difficult to enforce.

Prohibition

Governments throughout history have utilized prohibitions on a multitude of goods and services throughout history for a variety of reasons, such as attempts to protect domestic producers from foreign trade, protect consumers, regulate morality, etc. Thus, a prohibition is a means to achieving a broader social end.

In theory, the economics of prohibition is simple. Ultimately, prohibitions are a supply reduction legislation designed to curtail the production, exchange, and consumption of a good through the use of penalties, such as fines, confiscation of assets, and even jail sentences. Prohibitions attempt to reduce supply by making it more difficult for suppliers to produce and distribute the good in question. Within the supply and demand model, this causes a shift of the supply curve up and to the left resulting in a higher price and decrease in quantity demanded.

If enforcement is successful, this results in lower consumption, which has several welfare implications. First, it leads to a loss for consumers resulting from the higher price and the substitution of consumption to lower-valued goods. Second, producers experience a loss from the higher production costs and risks or in a loss of income and utility in needing to shift to other occupations, which differ from their comparative advantage.

Finally, a further utility loss occurs for both consumers and producers because the decrease consumption creates an allocative inefficiency due to the lost gains from trade or what economists call a deadweight loss.

The level of prohibition enforcement is also determined by economic factors. In enforcing prohibitions the government faces scarcity. Thus, enforcement is unlikely to be increased until the goal of eliminating the supply is reached. Just as individuals must equate marginal costs with marginal benefits in decision-making, so too must policy makers. Enforcement is not costless and requires the use of resources. Every dollar spent on prohibition enforcement is a dollar not spent on some other policy or a dollar not spent on something else by the taxpayer. Thus, the law of diminishing marginal utility exists in the policy world as well as for the individual. Policy makers must find the optimal level of enforcement, which is unlikely to lead to zero consumption. Enforcement will lower consumption but whether or not this leads to a level of consumption that achieves the policy goal is determined by the opportunity costs perceived by the individuals who make up society.

From a policy perspective, the success of the prohibition should not stem from the evidence of reduced consumption but rather whether the reduced consumption leads to the desired outcomes or not. A further difficulty in regard to the enforcement of a prohibition is the dynamic effects, or the unintended consequences, of the law. First, in pushing the trade underground information is greatly restricted due to the necessary premium on secrecy, which in turn greatly reduces price competition. This leads many in the market to charge monopoly prices, which widens profit margins. These large profits create an incentive for many, normally quite law abiding, individuals to break the law. This will include not only those directly participating in the trade but also government officials who succumb to the temptation of corruption. This has an effect of increasing both supply and demand, which attracts the attention of government officials, who then attempts to crack down. As a result the amateur is pushed out of the market but with the continued profit opportunities for individuals who are skilled at evading the law remain.

Second, enforcement can also lead to consequences concerning the good itself, such as a lower quality. With fewer competitors, suppliers do not need to rely on quality as much as they would in a legal market. The principle consequence, however, is what Richard Cowan, in a 1986 article, called the "iron law of prohibition," which is "the more intense the law enforcement, the more potent the drugs become." Cowan's point was related to drug prohibition but the principle stands for other goods as well. There are two reasons for this. First, the prohibition makes smuggling a necessary activity in order to get the goods to the consumers, and smugglers will prefer to minimize the bulk of their goods. For example, typically a shot of liquor, glass of beer, and a glass of wine have roughly the same amount of alcohol. Thus, due to its smaller size, liquor becomes more profitable to smuggle, driving beer and wine out of the market. Second, from the demand side, the same considerations apply. Liquor is more potent than beer and wine; it takes less of the good to do the job, of getting drunk in this example, for the consumer. When taking a hip flask with you for a night on the town filling it with whiskey is more effective than filling it with your favorite Pinot Noir.

These unintended consequences of prohibition are a causal result of the relationship between the market process and the intervention into that process. While these consequences are not completely predictable in detail, they are not completely surprising either. Interventions into the market's discovery process alter that discovery process; it does not eliminate it. Economist Israel Kirzner categorizes the four different ways interventions can broadly alter the market's discovery process, namely, by creating an undiscovered discovery process, an unsimulated discovery process, a stifled discovery process, and a wholly superfluous discovery process. All four of these are important for an intervention like a prohibition.

The undiscovered discovery process refers to the demand for government intervention due to either, or both, an ignorance or impatience with the market's ability to achieve the desired end or

state of affairs. The market is a process not an instantaneous state of affairs. Further markets operate under imperfect knowledge. Thus, corrections take time and occur under a decentralized spontaneous process. Prohibitions are often implemented due to perceived market failures, but often these failures are a result of other government interventions that impaired the market process in the first place. In other words, the undiscovered discovery process matters because government interventions are further demanded because imperfections of interventions are not seen or are blamed on the market and/or individuals are ignorant and/or impatient with the markets progress toward a solution.

The unsimulated discovery process refers to the inability of the bureaucracies in charge of a regulation to simulate the market's discovery process. Prohibition enforcement is made difficult because the bureaucracies in charge of enforcement are unable to simulate the discovery, or success, of the market. Bureaucrats have little incentive to innovate in order to improve efficiency within their tasks because they lack the incentive structure created by a competitive environment. As a result enforcement tends to be inefficient relative to what the market could or would produce. And this obviously creates a difficult hurdle for enforcement.

Interventions into the market also help to stifle the market's discovery process. Innovations in new techniques, safety, product characteristics, sources of supply, etc. created through this process are no longer discovered. In the case of prohibitions, the discovery process can be completely destroyed and also curtail and distort the process for other goods, as innovations in one area may have implications in others. And understanding the magnitude of this is simply not possible.

Finally, the wholly superfluous discovery process is the creation of new economic profit opportunities in response to the government's intervention. These new discoveries in search of profits are not always desirable. The new discovery process often creates wholly unexpected and undesired outcomes that create tremendous change in terms of both the product and the supply chain. Essentially, the government's attempts at enforcement are undercut as entrepreneurs figure ways to supply the good to consumers, and the means they use may even be worse than what the regulation is trying to fix. As economist Milton Friedman, in his 1975 book an Economist's Protest, put it, "There is ample evidence that imagination and innovation are not stilled by restrictive legislation – only diverted to figuring ways around it."

The altered profit opportunities not only affect market participants but also government officials responsible for enforcing the prohibition. Since the bureaucrats are not the residual claimant of the bureaucracy they are a part of, black markets provide chances to gain by providing market participants either protection or selective enforcement. In other words, a problem of graft and corruption within their own ranks will further impede the government's efforts to enforce the law, and this is a direct result of the wholly superfluous discovery process.

While not all prohibitions will be the same, the economic theory of prohibition does tend to illustrate difficulties with the enforcement of restricting market activities for many goods and services.

Cross-References

► Alcohol Prohibition

References

Cowan R (1986) How the narcs created crack: a war against ourselves. Natl Rev 38(23):26–34

Friedman M (1975) An economist's protest. Thomas Horton and Daughters, Glen Ridge

Further Reading

Becker GS, Murphy KM, Grossman M (2006) The market for illegal goods: the case of drugs. J Polit Econ 114(1):38–60

Boettke PJ, Coyne CJ, Hall AR (2013) Keep off the grass: the economics of prohibition and U.S. Drug Policy. Oregon Law Rev 17(4):1069–1095

Miron JA (2004) Drug war crimes: the consequences of prohibition. Independent Institute, Oakland

Thornton M (1991) The economics of prohibition. University of Utah Press, Salt Lake City

Promotional Activities

▶ Promotional Effort

Promotional Effort

Nikolaos Georgantzis[1,3] and Christian Boris Brunner[2]
[1]Economic and Social Sciences, School of Agriculture Policy and Development, University of Reading, Reading, Berkshire, UK
[2]School of Agriculture, Policy and Development, University of Reading, Reading, Berkshire, UK
[3]LEE and Economics Department, Universitat Jaume I, Castellon, Spain

Abstract

The term promotional effort refers to all strategies aimed at broadening a firms' market scope through the establishment if a larger and more loyal consumer basis. Advertising, public relations, sales promotion, personal selling as well as price-related strategies affecting a firm's sales potential are addressed. Both positive and normative approaches are birefly reviewed, discussing the theoretical and empirical issues studied in the existing literature.

Synonyms

Product promotion strategy; Promotional activities

Definition

The term is used to refer to the qualitative and quantitative aspects of a firm's strategies aimed at broadening its market scope through the establishment of a larger and more loyal consumer basis. According to Kotler et al. (2013), such activities can generally be classified into product management, pricing, promotion, and distribution. Promotional activities include advertising, public relations, sales promotion, personal selling as well as database marketing, direct response marketing, sponsoring, social media, and other alternative marketing activities (Clow and Baack 2014). In formal economic models, promotional effort is treated separately from pricing, in which case it refers to investments enhancing a firm's potential market before pricing is taken into account. However, price-related strategies like price announcements, bundle pricing, or low price guarantees could be considered as part of a firm's promotional effort, rather than merely a pricing decision.

Impact Measurement and Responses to Promotional Effort

The measurement of a firm's promotional effort is a challenge for marketers. First, a problem arises due to the difficulty in identifying the costs specific to different activities (e.g., Chapman 1986). Second, a problem arises with the measurement of a direct causal relationship between a given promotional strategy and its outcome (Berger et al. 1964), like, for example, consumers' reactions such as their attitude toward the brand or the sales rate of products. Third, this relationship is moderated and/or mediated by unknown and uncontrolled factors (Kuehn 1964; Berger et al 1964). Fourth, the outcome of promotional effort is delayed, which causes the problem of accountability over time between promotional effort and outcomes (e.g., Mills 1959; Miller and Strain 1970). Finally, multiple promotional activities of a firm may support or weaken each other.

Similar problems arise with respect to various products of a multiproduct firm. To overcome the complexity of the resulting setup, several experimental studies have been performed in order to isolate promotional effort effects within each separate domain (e.g., Berger et al. 1964; Miller and Strain 1970). To find the optimum promotional effort of a firm, economic models have taken into account the multiperiod, multicompetitor, and multiproduct nature of the problem (e.g., Gupta and Krishnan 1967a, b). Furthermore, the price of the firm's product(s) (e.g., Carpenter 1987; Krishnan and Gupta 1967) as well as a firm's price changes (price reductions or price

promotions) have been analyzed more in detail, including consumer switching or variety-seeking behavior (Kahn and Raju 1991). In order to be effective and efficient with respect to its promotional activities, a firm needs to: (1) *choose* the *right* promotional activities, to which its target group pays attention and responds positively. In addition to this, the selected promotional activities must fit to the firm's positioning and image and differentiate the firm from competitor; (2) *execute* the selected promotional activities in an efficient and effective way; and (3) *integrate* all promotional activities to one "main picture" in order to create a unique brand image in the consumer's mind.

Qualitative and Quantitative Aspects of Promotional Effort

From a managerial point of view, qualitative and quantitative considerations define the two major domains along which promotional effort enters into a firm's decision-making process. Qualitative considerations regard the choice of the actual strategy mix and the nature of explicit or implicit messages and signals transmitted to the consumer, whereas quantitative considerations regard the firm's investment in the different promotional activities. Along the qualitative domain, a firm's promotional effort could aim at creating or reinforcing the consumer's or stakeholder's brand awareness, stimulating their interest in the firm's products, encouraging first or more frequent purchases of the firm's products and building a long-term relationship (Bester and Petrakis 1996; Moraga-Gonzalez and Petrakis 1999) between the firm and its customers (Clow and Baack 2014; Keller 2013). The marketing literature emphasizes the need for an integrated strategy in order for promotional effort to lead to a unified image of the firm and its brands in the consumer's mind (Keller 1993, 2013), so that all promotional activities should support each other, transferring the same meaning to the receiver (Clow and Baack 2014). The importance of these qualitative aspects of promotional effort in a firm's decision-making problem has contributed to the fact that the issue has attracted researchers from many different disciplines, like economics, management, psychology, sociology, neuroscience, media and communication science, and even sociolinguistics.

Thus, the use of promotion techniques is informed by all the aforementioned approaches, pointing clearly to a possibility and desire of firms to intervene and affect the consumer's decision-making context and overall attitude, beyond pure product-related information, so that along all touch-points between customer and brand, the firm needs to communicate the same information to create a strong brand image in the customer's mind and to reinforce customer's brand knowledge (Keller 2013; Rossiter and Bellman 2005). Despite that, mainstream approaches within both the economics and marketing disciplines have strongly insisted on the information-enhancing role of advertising and promotional effort in general, adopting a preference invariance approach, similar to what could be motivated by Stigler and Becker's (1977) *De gustibus non est disputandum* (see also Becker 1996; Becker and Murphy 1993). As a consequence, all relevant legislations are almost exclusively concerned with the truthfulness of messages contained in informational campaigns, rather than with the persuasive effects of advertising, aimed at the subconscious processes underlying the consumer's choices (except for a clear ban of advertising aimed at trapping children's preferences).

Regarding the channel through which promotional effort broadens a firm's market potential, promotional effort may target retailers and wholesalers or end consumers. In the former case, the firm uses a *push* strategy, offering incentives directly to retailers and wholesalers (e.g., through personal selling or promotion). The aim is that retailers and wholesalers are the ones who engage in specific marketing activities such as personal selling, sales promotion, or advertising directly to the end consumers (Kotler et al. 2013; Barreda and Georgantzis 2002). In the latter case, the firm uses a *pull* strategy, promoting its brands and products directly to end consumers, for example, through TV spots, print advertisements, mobile marketing, and mailings.

Finally, the quantitative approach to promotional effort focuses on the issue of advertising expenditure and its efficiency as a market-

enhancing mechanism. Both types of advertising, informative (Grossman and Shapiro 1984) and persuasive (Bloch and Manceau 1999; von der Fehr and Stevik 1998) have been studied in oligopolistic contexts to reach a rather robust conclusion that advertising expenditures may be excessive from a social point of view and may lead firms to a prisoner dilemma-type of situation in which firms are trapped into "run-to-stay-still" competition. In such a case, firms may also end up earning lower profits than they would in a world without advertising. In the case of persuasive advertising, enhancing consumer heterogeneity seems to be the desired effect of promotional effort due to its competition-reducing ability. Generally speaking, this can be in detriment of social welfare even in the extreme case in which heterogeneity per se is in favor of a prosocial or environmental-friendly consumer behavior (Garcia-Gallego and Georgantzis 2009). Most of this wisdom has not been translated sufficiently into specific measures of economic policy and is largely neglected by the existing laws, except for some regulation regarding limitations of the overall time and space that advertising should be allowed in certain media.

Within the scope of informative advertising, special attention has been paid recently to best-price guarantees (e.g., *if you could find it cheaper, we refund the difference*) advertised by many large retailers all over the world. While, at a first glance, such a strategy appears to be a signal of the firm's commitment to prices which are lower than any of its rivals, suspicion has been raised by some authors (Arbatskaya et al. 2004, 2006), arguing that price guarantees may be used by firms wishing to facilitate cartel sustainability and discourage price cuts, because the transparency achieved regarding rival prices makes deviations from collusive agreed prices easier to detect. Both the promotional and the cartel facilitating explanations of price guarantees have received some support. The controversy persists, calling for a case-by-case treatment (Fatas et al. 2013) when concerns arise regarding the true motivation and effects of such guarantees. Other subtle forms of informative advertising relate to signaling of the firm's attitude toward certain quality aspects of their products, like safety (ISO standards), chemical composition, fair trade rule compliance, and environmental performance achieved by certificates and labeling (Loureiro and Lotade 2005). In those cases, legislation is assisted by the certifying or label-awarding authority so that no issues arise regarding imperfect and asymmetric information.

Cross-References

- ▶ Cartels and Collusion
- ▶ Distance Selling and Doorstep Contracts
- ▶ Labeling
- ▶ Liability and Information
- ▶ Internet Governance
- ▶ Labeling
- ▶ Signalling

References

Arbatskaya M, Hviid M, Shaffer G (2004) On the incidence and variety of low price guarantees. J Law Econ 47:307–332

Arbatskaya M, Hviid M, Shaffer G (2006) On the use of low-price guarantees to discourage price cutting. Int J Ind Organ 24:1139–1156

Barreda-Tarrazona I, Georgantzís N (2002) Regulating vertical relations in the presence of retailer differentiation costs. Int Rev Law Econ 22:227–256

Becker G (1996) Accounting for tastes. Harvard University Press, Cambridge, MA

Becker G, Murphy M (1993) A simple theory of advertising as a good or bad. Q J Econ 108:941–964

Berger PK, Fraley GW, Tarpey LX (1964) The effects of advertising on the use of a train's diner. J Advert 9:25–29

Bester H, Petrakis E (1996) Coupons and oligopolistic price discrimination. Int J Ind Org 14:227–242

Bloch F, Manceau D (1999) Persuasive advertising in hotelling's model of product differentiation. Int J Ind Organ 17:557–574

Carpenter GS (1987) Modeling competitive marketing strategies: the impact of marketing-mix relationships and industry structure. Market Sci 6:208–221

Chapman RG (1986) Assessing the profitability of retailer couponing with a low-cost field experiment. J Retail 62:19–49

Clow KE, Baack D (2014) Integrated advertising, promotion, and marketing communications. New York: Pearson Education Ltd

Fatás E, Georgantzís N, Sabater-Grande G, Máñez J (2013) Experimental duopolies under price matching and price beating guarantees. Appl Econ 45:15–35

García-Gallego A, Georgantzís N (2009) Market effects of changes in consumers' social responsibility. J Econ Manage Strat 18:235–262

Grossman GM, Shapiro C (1984) Informative advertising with differentiated products. Rev Econ Stud 51(1):63–81
Gupta SK, Krishnan KS (1967a) Differential equation approach to marketing. Oper Res 15:1030–1039
Gupta SK, Krishnan KS (1967b) Mathematical models in marketing. Oper Res 15:1040–1050
Kahn BE, Raju JS (1991) Effects of price promotions on variety-seeking and reinforcement behavior. Market Sci 10:316–337
Keller KL (1993) Conceptualizing, measuring, and managing customer-based brand equity. J Market 57:1–22
Keller KL (2013) Strategic brand management: building, measuring, and managing brand equity. Prentice Hall, Upper Saddle River
Kotler P, Armstrong G, Harris L, Piercy NF (2013) Principles of marketing. Essex: Pearson Education Ltd
Krishnan KS, Gupta SK (1967) Mathematical model for a duopolistic market. Manage Sci 13:568–583
Kuehn AA (1964) The marketing concept in action. American Marketing Association, Chicago
Loureiro ML, Lotade J (2005) Do fair trade and eco-labels in coffee wake up the consumer conscience? Ecol Econ 53:129–138
Miller BR, Strain CE (1970) Determining promotional effects by experimental design. J Market Res 7:513–516
Mills HD (1959) A study in promotional competition. The Harlan D. Mills Collection. http://trace.tennesse.edu/utk_harlan/48
Moraga-González JL, Petrakis E (1999) Coupon advertising under imperfect price information. J Econ Manage Strat 8:523–544
Rossiter JR, Bellman S (2005) Marketing communications: theory and applications. New York: Pearson.
Stigler G, Becker G (1977) De Gustibus Non Est Disputandum. Am Econ Rev 67:76–90
von der Fehr NH, Stevik K (1998) Persuasive advertising and product differentiation. South Econ J 65:113–126

Property Rights: Limits and Enhancements

Geoffrey M. Hodgson
Hertfordshire Business School, University of Hertfordshire, Hatfield, Hertfordshire, UK

Definition

This essay considers the concept of "property right" as typically employed in "the economics of property rights." Some early uses of the term "property right" in economics are quoted before moving to the predominant usages in the standard "economics of property rights" today. This prevailing usage is compared with the more nuanced notion of "property right" in legal theory. It is argued that, for the purposes of understanding the role and impact of property rights on economic performance, it would be better for economists to return to the legal meaning of property rights. This legal meaning can cope better with important developmental phenomena such as the use of property as collateral to finance loans. Also the recognition of the legal character of property opens doors for a richer understanding of motivations to respect property rights, which is consistent with empirical data indicating that people are not simply motivated to respect laws out of instrumental calculations of costs and benefits.

Introduction

Nowadays, many economists concur that the development of secure and enforceable property rights is a vital factor in the process of modern economic development. But what are property rights? Different answers to this question focus on different kinds of social relationship or institution, which, in turn, can lead to different policy recommendations. On closer inspection, there are major differences of opinion on this issue.

These differences can be divulged by posing two connected queries. First, are property rights conceived as an eternal feature of the human condition, or are they historically specific, emerging (say) with states and legal systems? Many researchers in this field treat property as an evolved – spontaneously or endogenously emerging – response to human interactions over territory or other wealth, and do not assume that states or legal systems are necessary for its emergence.

The division of opinion on the first query often related to the divide in the history of economics between those that emphasise universal principles or laws, and those that claim that concepts such as exchange, supply, demand, and even calculative rationality are historically specific, depending on particular economic institutions. Much of modern

This chapter distils and extends some material from Hodgson (2015a, b).

economic theory emphasizes universal principles, whereas others, including the German historical school and the original American institutionalists, emphasized historical specificity.

The second, related, query is whether "property" is treated as synonymous with mere possession – i.e., de facto control – over a resource, or does it involve the acquisition of declared rights, backed up by institutions of legislation, adjudication, and enforcement? Again the rival choice of emphasis on universality or historical specificity is relevant here. Those that seek a universal concept of property must find it also in worlds where there is no state, where custom rules without institutionalized law. By contrast, those that make property more than possession and relate rights to legal institutions are obliged to abandon the notion that property is a universal or ahistorical concept.

This entry argues that this fundamental difference of outlook over the nature of property has major theoretical and practical consequences. It has divided opinion since the disciplines of law and economics began to interact in the nineteenth century. It has yet to be played out to an adequate solution.

Consider some examples. John R. Commons was an early institutional economist and a leading explorer of the overlapping territories of economics and law. He was clear: "in the end, the actual title to property rests on the sovereign power of the state to enforce its decrees. . . . There is, strictly speaking, no such thing as absolute, unlimited right of property, which law steps in as an afterthought to restrict" (Commons 1893, p. 110).

The institutional economist Thorstein Veblen also pointed out that property is historically specific: "no concept of ownership, either communal or individual, applies in the primitive community. The idea of communal ownership is of a relatively later growth" (Veblen 1898, p. 358).

A historically specific, state-related concept of property also pervades much analysis by legal scholars. For example, the leading legal theorist Antony Honoré (1961, pp. 107, 115) insisted that property is much more than possession or control:

> A people to whom ownership was unknown, or who accorded it a minor place in their arrangements, who meant by *meum* and *tuum* no more than "what I (or you) presently hold" would live in a world that is not our world. . . . To have worked out the notion of 'having a right to' as distinct from merely "having" . . . was a major intellectual achievement.

Legal scholars Daniel Cole and Peter Grossman (2002) have documented how the notion of "property rights" in prevailing contemporary approaches in economics departs radically from that notion in law. For other insights and dissenting voices on the treatment of property, see Hernando (2000), Steiger (2008), Fukuyama (2011, pp. 66–71), Arruñada (2012, 2017), and Heinsohn and Steiger (2013).

This entry explores this crucial divergence of usage and interpretation of a key concept. The following section lays out the concept of property in much of "the economics of property rights." Sections "Why Do People Obey the Law?" and "Property Rights and Economic Development" explore some implications of the discussion in terms of practical analyses of the roles of property. Section "Concluding Remarks" concludes the chapter.

The Concept of Property in the Contemporary "Economics of Property Rights"

At least in its early stages of its development, the now-conventional "economics of property rights" was influenced by the treatment of institutions in the Austrian school of economics, particular by Carl Menger, Ludwig Mises, and Friedrich Hayek.

Menger famously located the essence of some institutions in the spontaneous arrangements that engender or sustain them, rather than in acts of decree by a state or other public authority. Similarly, the standard economics of property rights strives to understand property as a spontaneous institution, which does not necessarily involve the state.

Mises made a sustained attempt to make core economic concepts as universal as possible. So when he considered the nature of ownership, he treated the legal aspect as merely a normative

("ought to have") justification of de facto "having" something:

> From the sociological and economic point of view, ownership is the *having* of the goods . . . This *having* may be called the natural or original ownership, as it is purely a physical relationship of man to the goods, independent of social relations between men or of a legal order. . . . Economically . . . the natural *having* alone is relevant, and the economic significance of the legal *should have* lies only in the support it lends to the acquisition, the maintenance, and the regaining of the natural *having*. (Mises [1932] 1981, p. 27)

Hence, for Mises, ownership was natural and ahistorical rather than legal or institutional. An individual-physical rather than a social relationship, it was deemed independent of law or any other social institution. He downgraded the institutions required for the legitimation, protection, and enforcement of the capacity to have, and neglected social aspects of ownership that may signal power or status.

Hayek's position was in some respects different from Mises, but by arguing that all law was basically custom, Hayek (1973, pp. 72–5) echoed Mises's argument by removing any historical specificity to the notion of law. In part, this is a matter of the definition of law. But there are good reasons for regarding law as a characteristic of systems with a fully institutionalized judiciary and legislature (Hodgson 2015a, b). Hayek not only reduced law to custom, but he treated common law as essentially spontaneous and customary, whereas in fact it was by large part codified and made consistent by the state (Fukuyama 2011, pp. 254–60).

It is possible to reject Hayek's extremely wide characterization of law without adopting the view that law emanates simply from the state, and without underestimating the crucial role of custom in legal development. Indeed, all law depends to a large degree on custom, for its origin or for its implementation. While law, in the narrower and historically specific definition, implies the state, it is not constituted simply by decree and it depends upon custom.

Notably, in contrast to Hayek, Ronald Coase never suggested that property in a modern market economy could exist or be understood without a state legal system. Coase (1988, p. 10) wrote:

> When the physical facilities are scattered and owned by a vast number of people with very different interests . . . the establishment and administration of a private legal system would be very difficult. Those operating in these markets have to depend, therefore, on the legal system of the State.

Subsequent writers on "the economics of property rights" departed from this maxim, enabling them to tackle the definition of "property right" in a very general way, ignoring its dependence on "the legal system of the state."

Armen Alchian was a key figure in the development of the contemporary "economics of property rights." Alchian (1965) defined private property rights in terms of assignments of the ability to choose the use of goods (without affecting the property of other persons). Following Mises, his definition was largely in terms of de facto powers of control rather than legal or moral rights.

Alchian removed morality from the notion of a "right." He downgraded the role of law, from being the source of legitimate authority concerning rights to one means among others for enforcing possession and control. But typically, we refer to moral or legal rights, we do not simply mean de facto ability to control resources. The *Oxford English Dictionary* defines "right" (as a noun) as "that which is morally correct, just, or honourable" or "a moral or legal entitlement to have or do something." It means much more than the ability to do something.

Later Alchian (1977, p. 238) defined the "property rights" of a person in universal and institution-free terms including "the probability that his decision about demarcated uses of the resource will determine the use." Alchian's definitions of property neglect the essential concept of legitimated, rightful ownership. This concept is important, even if what is rightful is contestable or difficult to identify. His definitions denote possession rather than property.

Similarly, the highly influential property-rights economist Yoram Barzel (1994, p. 394) defined property as:

> an individual's net valuation, in expected terms, of the ability to directly consume the services of the asset, or to consume it indirectly through exchange.

> A key word is *ability*: the definition is concerned not with what people are legally entitled to do but with what they believe they can do.

This explicitly removed the question of legal title from the definition of property. The upshot of this is that if a thief manages to keep stolen goods then he acquires a substantial property right in them, even if, on the contrary, legal or moral considerations would suggest that they remain the rightful property of their original owner. Elsewhere Barzel (1997, p. 3) argued:

> The term 'property rights' carries two distinct meanings in the economic literature. One . . . is essentially the ability to enjoy a piece of property. The other, much more prevalent and much older, is essentially what the state assigns to a person. I designate the first 'economic property rights' and the second 'legal (property) rights.' Economic rights are the end (that is, what people ultimately seek), whereas legal rights are the means to achieve the end. Legal rights play a primarily supporting role. . .

Barzel's distinction between "economic property rights" and "legal property rights" has proved influential in this genre of literature. Barzel made it clear that his version of "the economics of property rights" is not about legalities. But it is questionable to regard "the ability to enjoy" something as a "right." Enjoyment can exist without rights, and rights without enjoyment. Rights result from institutionalized rules involving assignments of potential benefit. They always involve relations between people as well as relations with things. The "ability to enjoy" may not involve more than an individual's relationship with an object.

One possible response is that all this is essentially about the choice of definitions and the uses of words. In particular, property rights economists might simply say that what some call "possession" they call "economic property rights" and what some call "property rights" they call "legal property rights." Consequently, we can go through and edit the relevant economics of property rights literature, making the appropriate terminological substitutions, and everything else would remain valid.

This is contestable. Words carry meanings that we cannot simply alter at will. While the foundational economics of property rights literature brings important and valid insights, it would still be impaired after these terminological substitutions. First, using the word "right" to describe something that is not a right but a matter of de facto control is misleading: it obscures the adopted legal meaning of rights in modern legal and economic systems (Cole and Grossman 2002).

Second, attention is diverted from the roles of moral sentiments and dispositions to obey authority: they are important parts of human motivation, alongside greed and self-interest. Third, even with the terminological substitutions, the foundational property rights literature would have an inadequate treatment of (legal) property rights as collateral to obtain loans, and of the historically specific legal institutions that make collateralization possible. These issues are explored in the following two sections.

Why Do People Obey the Law?

Consider the intrinsic role of law in human motivation. Intrinsic motivation stems from the perceived nature of law itself, rather than from the sanctions or rewards of the legal system. The "economics of property rights" often neglects this intrinsic motivation: it is often assumed that law impinges on behaviour principally as a cost or constraint. Obeying the law is simply a matter of expected costs and expected benefits, where no benefit is assumed from legal compliance itself.

Hence, having made his distinction between two kinds of "rights," Barzel (2002, pp. 16, 157) claimed that: "What individuals maximize (subject to their personal safety) is the value of their *economic rights*." These exclude "legal rights," which are defined as "*claims over assets delineated by the state*." In other words, individuals are ceteris paribus indifferent to "legal rights" and act solely to maximize their enjoyment of assets under their control, whether these assets are obtained legally or illegally. This assumes a particular form of maximizing behaviour where law itself has no direct input as an argument in the preference function.

Barzel accepted that the law may matter. It may be a major factor in determining outcomes in terms of control and enjoyment of resources. But Barzel (1997, p. 3) argued that law matters only insofar as it leads to "what people ultimately seek," namely the enjoyment of resources. Legal obedience has no intrinsic value, and it is simply treated as an instrumental "means" to that "end." Legal obedience would not appear as an argument in any presumed preference function. Law is regarded as simply an instrumental means to an end. According to Barzel, obeying the law is not an end in itself.

By contrast, the subfield of psychological jurisprudence has gathered considerable survey data to establish that people place some normative value on obeying the law, in addition to any instrumental consideration of expected personal costs or benefits. Tom R. Tyler is the leading authority in this area. Tyler (1990, p. 3) contrasted the "instrumental perspective" where "people are viewed as shaping their behaviour to respond to changes in the tangible, immediate incentives and penalties associated with following the law" with the "normative perspective" concerned with "what people regard as just and moral as opposed to what is in their self-interest."

Supported by evidence gained from a survey of several hundred citizens in Chicago, Tyler (1990) argued that citizen may be inclined to obey the law for noninstrumental as well as instrumental reasons. For some actors, the estimated costs (involving estimates of the chances of being caught and likely punishments) and benefits of breaking the law are weighed against the costs and benefits of compliance. But the evidence of Tyler and others suggests that such instrumental calculations have relatively little effect on compliance. More important is internalized obligation, stemming from other considerations.

First, citizens often comply with the law because they regard the legal authority as having a legitimate right to lay down rules that people must obey. The notion of legitimate authority can have several possible bases, but in modern democracies, it mainly derives from the belief that the popular election of a government makes such state authority legitimate. Max Weber ([1922] 1968, p. 215) saw legal authority and legitimation as "resting on a belief in the legality of enacted rules and the right of those elevated to authority under such rules to issue commands."

Second, citizens sometimes comply with a particular law because it is believed to be moral. This is particularly the case with laws against murder or rape, but may not be so with other laws, such as those governing traffic on roads. Tyler argued that moral beliefs sometimes help to sustain laws. But this does not work when people believe that a particular law lacks sufficient moral force.

Instrumentalists in general, and proponents of utility-maximization in particular, argue that the issues of legitimacy and morality can be incorporated in the calculus of costs, benefits or utility. Hence advocates of utility-maximization claim that when someone acts according to their perceptions of legitimacy or morality, they gain extra utility from a "warm glow" of self-satisfaction. A problem with this argument is that it undermines the very meaning of legitimacy and morality. The whole point of actions in accord with perceived legitimacy or morality is that they do not necessarily serve self-interest. Acting morally means "doing the right thing," even if it is costly to the actor. People obey legitimate legal authority saying "because it's the law." These deontic motivations cannot be reduced to convenience, convention, or cost-benefit calculation, in accord with a unidimensional calculus of utility or desire.

Both adults and children feel strong obligations to obey the law. Austin Sarat (1975) found that 70% of adults in his US sample agreed that a law "must always be obeyed." A survey of US high school students found that 77% of whites and 72% of blacks agreed that "people should always obey the law" (Rodgers and Lewis 1974). Tyler (1990, p. 178) summarized a major implication of his survey of Chicago citizens:

> People obey the law because they believe that it is proper to do so, they react to their experiences by evaluating their justice or injustice, and in evaluating the justice of their experiences they consider factors unrelated to outcome, such as whether they have had a chance to state their case and been treated with dignity and respect. . . . The image of the person resulting from these findings is one of a person whose attitudes and behavior are influenced

> to an important degree by social values about what is right and proper. This image differs strikingly from that of the self-interest models which dominate current thinking. . .

Our ingrained inclinations to respect those in authority were dramatized by the experiments of Stanley Milgram (1974), who invited members of the public to help in a laboratory study ostensibly about learning. A "scientist" asked these recruits to administer electric shocks to a subject, to punish wrong answers to questions. Milgram found that a majority of adults would administer shocks that were apparently painful, dangerous, or even fatal, if ordered to do so by the person in authority. In fact, there were no shocks and the subject was an actor, feigning agony or even death. This experiment shows that people can willingly accept the orders of perceived authority figures, even when their own moral feelings are violated.

Milgram (1974, pp. 124–5, 131) argued that our dispositions to respect authority emanate from the evolutionary survival advantages of cohesive social groups. While socialization and learning are clearly important in developing propensities to obey authority and the law (Engel 2008), Milgram also proposed that the human species has evolved an inherited, instinctive, propensity for obedience that is triggered by specific social circumstances. He suggested that dispositions to respect authority have both genetic and cultural foundations. This is in accord with Charles Darwin in his *Descent of Man*, who proposed that human tribes that developed systems of social obedience and cooperation, and a moral code to buttress these attributes, would survive in competition with other human tribes and in dealing with their environment.

These arguments undermine the view of Barzel and others that respect for the law – based on its perceived legitimacy or moral concordance – plays little or no part in attitudes toward property or other legal rights. The insistence that property is a legal right does not imply that people never break the law, or that law alone somehow predicts behavior. But the establishment of legal rights, through perceptions of moral legitimacy and the use of state power, can affect intentions or behavior. As Honoré (1961, pp. 107, 115) insisted, an economy involving mere possession is very different in nature and outcomes from one that has institutionalized rights of property.

The mistaken removal of legal rights from the definition of property cannot be justified on the ground that they are unnecessary to explain or predict behaviour. Any explanation of dispositions, choices, or preferences must take such factors into account. If economists are interested in predicting behavior on the basis of some scientific understanding of what causes it, then they must take matters of motivation, including the instrumental and the normative into account (Merrill and Smith 2007).

Property Rights and Economic Development

Arguments emphasizing the perceived legitimacy of the legal system have implications for establishing the rule of (state) law, and particularly installing just and secure property rights to help promote economic development.

China is an important test case for these arguments. China began its market reforms in 1978 and its systems of property, commercial, and corporate law are still relatively underdeveloped compared to Europe or North America. This fact, alongside its highly impressive economic growth since 1978, has led some economists to conclude that legally-enforced property rights are of lesser significance.

But, despite superficial appearances to the contrary, there is evidence that legal systems and legal property rights matter. China's explosive growth started when land-use (*usus fructus*) rights were widely conceded to the peasants after 1978 (Coase and Wang 2012). Relevant legislation concerning land leasing followed rather than preceded this concession. But this does not mean that legal land-use rights were unimportant. Local power from below tentatively established *de facto* powers, which spread widely and became *de jure* when it was legally ratified by the state. This endorsement, along with the institutional arrangements established from below, was vital to safeguard these rights.

Legalities matter and evidence suggests that they matter still more as capitalism develops. Further economic development in East Asia may depend in part on the installation of superior state legal and political systems governing and protecting property and contracts. Private ordering is vital but insufficient. The cross-country evidence of Robert J. Barro (1997) and others suggests that economic growth is correlated with the rule of law, among other factors.

By contrast, the Alchian-Barzel approach might see it as sufficient that Chinese entrepreneurs can control resources sufficiently to enable prosperity and rapid growth. Matters of legal infrastructure and enforcement would be secondary. Instead, the alternative perspective outlined here would point to the priority of developing legal institutions, their separation from political authority, and the related reform of the political system.

As Hernando (2000) and others have pointed out, the registration and enforcement of property rights (particularly in land and buildings) is a necessary condition for economic development. But Arruñada (2017) argued persuasively that property registries, while necessary, are insufficient. Among other things, they need to be backed up by well-functioning systems of law enforcement.

The exclusive focus on control in "the economics of property rights" overlooks the use of property as collateral for loans. The possibility of collateralization – which relies on legal and financial institutions – cannot be based on possession alone. It involves institutions: relations between individuals as well as relations between individuals and things. While emphasizing the importance of "property rights," much of this discourse side-lines the vital institutions that are required to sustain them and make them fully operational in a developed economy.

Another relevant historical case study is the English Glorious Revolution of 1688. Douglass North and Barry Weingast (1989) famously claimed that this event made property rights more secure by constraining the power of the monarch. But on closer inspection, it becomes clear that property rights in England were relatively secure centuries before 1688 (Fukuyama 2011, pp. 418–20). The problem instead was that these rights involved *entails*, which kept land and buildings in the family and prevented their use as collateral to secure loans.

Instead of making property rights secure, 1688 led to new alliances and a series of wars, requiring the state to reform its administration, gather more taxes, and help create a new financial system. This created conditions in the eighteenth century that eventually incentivized and facilitated the use of land as collateral for loans for agricultural, infrastructural, and industrial development (Hodgson 2017). The overly simplistic treatment of property rights as mere possession obscures the importance of different aspects of property (Honoré 1961) and the crucial role of potential collateralization in creating the conditions for the Industrial Revolution.

Attempts to make "property rights" analysis universal, so that it applies to societies without effective (state) legal authority draw our attention away from the importance of property rights, properly defined (Arruñada 2017). Of course it is important to understand extra-legal enforcement mechanisms, such as with pirates and mafias, but we should not pretend that might and right are the same. Of course, studies of worlds without (state) legal enforcement can help us devise policies that apply to developing countries where the state is weak or dysfunctional. But – following Coase (1988) among others – we should not assume that such spontaneous mechanisms are sufficient, or can apply to large-scale, complex economies. The reinstatement of a more genuine concept of "property rights" in development economics would lead to greater emphasis on the importance of an effective legal system that enjoys some autonomy from political or sectional power with some perceived justice in its proceedings.

Concluding Remarks

The success of capitalism has depended on systems of law enforcement. But these took a long time to establish. Even today, in much of the world, systems of law enforcement are weak,

expensive, corrupt, or inaccessible. In their absence, people fall back on other means of establishing obligations and ensuring compliance. Commerce then works through clan or family ties, shared religion, or ethnicity, bureaucratic cooption and corruption, or threats of violence to person or property. Systems of spontaneous enforcement show how commercial agreements can be maintained in the absence of adequate state systems of law. Such systems existed in history and persist today in some contexts. Hence they are important objects of analysis. But this should not mislead us into believing that fully-developed modern capitalist systems can rest on purely spontaneous or customary foundations.

A crucial argument here is that legal institutions have to be taken into account in "economic" analysis. Law is much more than a constraint: it matters too for an adequate understanding of human motivation and the financial dynamics of capitalism. An alternative approach – which takes the impact of legal institutions more seriously – is described as *legal institutionalism* (Hodgson 2015a).

An appreciation of noninstrumental motivations for legal compliance, which are part and parcel of the arguments here concerning the nature of property rights, challenges the use of utility-maximization as sufficient model of human behaviour. If "economics" is confined to utility-maximizing agents, then "economics" cannot adequately deal with the reality of property rights. But if economics takes on board the insights of Adam Smith, Ronald Coase, Amartya Sen, and several others, who all adopted more complex views of human motivation involving moral sentiments, then the "economics of property rights" can be greatly enriched.

Cross-References

- ▶ Austrian School of Economics
- ▶ Capitalism
- ▶ Constructivism, Cultural Evolution, and Spontaneous Order
- ▶ Economic Analysis of Law
- ▶ Economic Development
- ▶ Economics Imperialism in Law and Economics
- ▶ Hayek, Friedrich August von
- ▶ Institutional Economics
- ▶ Mises, Ludwig von

References

Alchian AA (1965) Some economics of property rights. *Il Politico* 30:816–829

Alchian AA (1977) Some implications of recognition of property right transaction costs. In: Brunner K (ed) Economics and Social Institutions: Insights from the Conferences on Analysis and Ideology. Martinus Nijhoff, Boston, pp 234–255

Arruñada B (2012) Institutional foundations of impersonal exchange: theory and policy of contractual registries. University of Chicago Press, Chicago

Arruñada B (2017) Property as sequential exchange: the forgotten limits of private contract. *J Inst Econ* 13(4):753–783

Barro RJ (1997) Determinants of economic growth: a cross-country empirical study. MIT Press, Cambridge, MA

Barzel Y (1994) The capture of wealth by monopolists and the protection of property rights. *Int Rev Law Econ* 14(4):393–409

Barzel Y (1997) Economic Analysis of Property Rights, 2nd edn. Cambridge University Press, Cambridge

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

Coase RH (1988) The firm, the market, and the law. University of Chicago Press, Chicago

Coase RH, Wang N (2012) How China became capitalist. Palgrave Macmillan, London/New York

Cole DH, Grossman PZ (2002) The meaning of property rights: law versus economics? *Land Econ* 78(3):317–330

Commons JR (1893) The Distribution of Wealth. Macmillan, London/New York. Reprinted 1963 (New York: Augustus Kelley)

Engel C (2008) Learning the Law. *J Inst Econ* 4(3): 275–297

Fukuyama F (2011) The Origins of Political Order: From Prehuman Times to the French Revolution. Profile Books/Farrar, Straus and Giroux, London/New York

Hayek FA (1973) Law, legislation and liberty; volume 1: rules and order. Routledge/Kegan Paul, London

Heinsohn G, Steiger O (2013) Ownership economics: on the foundations of interest, money, markets, business cycles and economic development. Routledge, London/New York. Translated and edited by Frank Decker

Hernando DS (2000) The mystery of capital: why capitalism triumphs in the west and fails everywhere else. Basic Books, New York

Hodgson GM (2015a) Conceptualizing capitalism: institutions, evolution, future. University of Chicago Press, Chicago
Hodgson GM (2015b) Much of the "economics of property rights" devalues property and legal rights. *J Inst Econ* 11(4):683–709
Hodgson GM (2017) 1688 and all that: property rights, the glorious revolution and the rise of British capitalism. *J Inst Econ* 13(1):79–107
Honoré AM (1961) Ownership. In: Guest AG (ed) Oxford Essays in Jurisprudence. Oxford University Press, Oxford, pp 107–147. Reprinted in *J Inst Econ,* 9(2), 2013, pp. 227–55
Merrill TW, Smith HE (2007) The morality of property. William Mary Law Rev 48(5):1849–1895
Milgram S (1974) Obedience to authority: an experimental view. Harper and Row/Tavistock, New York/London
Mises L (1981) Socialism: An Economic and Sociological Analysis. Liberty Classics, Indianapolis. Translated from the second (1932) German edition
North DC, Weingast BR (1989) Constitutions and commitment: the evolution of institutions governing public choice in seventeenth-century England. *J Econ Hist* 49(4):803–832
Rodgers HR, Lewis E (1974) Political support and compliance attitudes: a study of adolescents. *Am Polit Q* 2:61–77
Sarat A (1975) Support for the legal system. *Am Polit Q* 3(1):3–24
Steiger O (ed) (2008) Property economics: property rights, Creditor's money and the foundations of the economy. Metropolis, Marburg
Tyler TR (1990) Why people obey the law. Yale University Press, New Haven
Veblen TB (1898) The beginnings of ownership. *Am J Sociol* 4(3):352–365
Weber M (1968) Economy and Society: An Outline of Interpretative Sociology. University of California Press, Berkeley. 2 vols, translated from the German edition of 1922

Proportionality Test

Tomáš Sobek[1] and Josef Montag[2]
[1]Faculty of Law, Masaryk University, Brno, Czech Republic
[2]International School of Economics, Kazakh-British Technical University, Almaty, Kazakhstan

Abstract

Proportionality test is a legal method used by courts, typically constitutional courts, to decide hard cases, which are cases where two or more legitimate rights collide. In such cases a decision necessarily leads to one right prevailing at the expense of another. In order to decide such cases correctly, the court must balance the satisfaction of some rights and the damage to other rights resulting from a judgment. This entry overviews the proportionality test and the four steps of implementing the test. We also discuss the incommensurability problem, which is the main criticism of the balancing approach.

Proportionality Approach to Hard Cases

Courts, constitutional courts, in particular, sometimes face cases that cannot be decided by referring to the letter of the law or an interpretation thereof. Such cases, henceforth "hard cases," are characterized by a collision of legitimate rights, or public interests, that are protected by the law, typically by the constitution. Adjudicating such hard cases inevitably results in a decision as to which right will prevail at the expense of limiting the other right. That is to say, in such cases courts face a trade-off situation.

In order to decide such cases in a rational manner, the adjudicators need a criterion to distinguish good and bad trade-offs. After the Second World War, the German constitutional theory developed the proportionality test as its main methodological tool for reconciling conflicting rights and values. Proportionality test is a method that allows courts to structure the decision dilemma at hand in a way that facilitates the decision to be carried out in a rational and transparent manner by balancing the benefits and costs – broadly understood – of alternative courses of action. And this approach proved to be viable. The proportionality test, indeed, became a dominant technique of adjudication of hard cases around the world (Aleinikoff 1987).

Proportionality test can be viewed as a set of rules for determining the necessary and sufficient conditions for a limitation of a constitutionally protected right and ascertaining whether these conditions are satisfied. The proportionality test consists of four steps or sub-tests (Alexy 2014).

A limitation of a constitutional right by a legal act is constitutionally permissible if and only if (i) the act pursues a legitimate aim (legitimacy test), (ii) the act is capable of achieving this aim (suitability test), (iii) the act impairs the affected right as little as possible (necessity test), and (iv) the importance of achieving the aim outweighs the importance of preventing the limitation on the affected right (balancing test or proportionality *stricto* sensu).

If these conditions are satisfied, the test concludes that the right under consideration has greater weight and should prevail over the conflicting right. The resulting preference relation, however, is specific to the case at hand only and its context. The outweighed right is not to be considered invalid; it may itself outweigh some other right in different circumstances.

From the economic point of view, the proportionality framework is interesting because it embodies the requirement for efficiency. The rules of proportionality require that the tradeoffs that courts make are efficient in the sense that the prevailing right is of higher social value and that its protection is achieved with minimal costs for the "losing" right. The test postulates that there should be a reasonable relationship between a particular purpose to be achieved by law and the legal means used to achieve that purpose. Judges must pay attention to the social importance of achieving the law's purpose as well as the social importance of preventing the harm to other constitutionally protected rights. Put simply, they must consider not only the benefits associated with the proposed decision but also the damage to other rights caused by it.

The Steps

Legitimacy Test

The legitimacy test is simply a requirement that the legal principle, right, or public interest which the decision would uphold is lawful and reasonable, and so it should not be dismissed right away. This step simply aims to filter out cases, which are to be decided on different grounds and for which the proportionality test would be therefore obsolete. The second, third, and fourth sub-tests express the idea of optimization relative to the factual and legal possibilities, and so we will discuss them in more detail.

Suitability Test

The stage of suitability review examines whether there is a reasonable connection between the act that interferes with a right and its legitimate aim. The interference must be capable of contributing to the achievement of the legitimate aim at least to a small extent. The compliance with this requirement excludes the adoption of means which obstruct some right without promoting any other right or interest for which it has been adopted. The point of this stage is to sort out those cases where there does not actually exist a conflict of rights (or a conflict of a right and public interest). Naturally, a conflict of two rights means that one cannot realize one right without detriment to the other. If the interference with one right does not contribute to the realization of any other right then this interference would be merely an obstruction without any conflict of rights justifying it.

Necessity Test

The stage of necessity reviews examines whether the proposed legal act impairs an affected right as little as possible. The act is necessary if there are no alternative measures that may achieve the same purpose with a lesser degree of impairment. If there is an equally suitable means that interferes with the affected right less intensively, then things can be improved without any costs. The necessity requirement postulates that no such free lunch is left on the table. So, this sub-test requires that of two means promoting the same aim (a certain right or public interest) to the same degree, the one that interferes less intensively with a conflicting right, i.e., generates lower costs, has to be chosen. This shows that the necessity stage is an expression of the idea of efficiency.

Balancing Test

While the stages of suitability and necessity essentially involve optimization relative to the factual possibilities, the stages of legitimacy and balancing (proportionality in the narrower sense)

refer to what is legally possible. Balancing, the last stage of the proportionality test involves optimization relative to the competing rights. The greater the degree of non-satisfaction of, or detriment to, one right, the greater is the required importance of satisfying the other right for the balancing test to tilt in its favor (Alexy 2002, p. 102). This requirement excludes an intensive interference with one right that is justified only by a low importance assigned to the satisfaction of the colliding right.

The process of balancing can be broken down into three steps. (i) In the first step, the degree of non-satisfaction of or detriment to a right is established. (ii) The second step consists of establishing the importance of satisfying the competing right. (iii) And in the last step, it is examined whether the importance of satisfying the latter right justifies the detriment to or non-satisfaction of the former.

These three steps of balancing are more explicitly captured by the weight formula:

$$W_{i,j} = \frac{I_i \cdot W_i \cdot R_i}{I_j \cdot W_j \cdot R_j},$$

where i and j index the colliding principles, I_i and I_j capture the intensity of interference in the respective rights in case of a decision in favor of the other right, W_i and W_j represent the abstract weights of the two colliding principles, and R_i and R_j capture the reliability of empirical assumptions on both sides of the argumentation (Alexy 2014, pp. 55–56). If the weight is greater than one, principle i should prevail; if it is less than 1, the court should rule in favor of principle j.

The data entering the formula is necessarily imprecise and based on court's judgment. This, however, stems from the very nature of the decision problems that courts face and does not invalidate the balancing approach itself. To illustrate how such balancing can be approached, Alexy (2014) suggests to use the triadic scale often used in the constitutional discourse: light (l), moderate (m), and serious (s). The levels of this scale can then be mapped to a numeric scale with more than proportional increments, such as the geometric sequence such $2^0, 2^1, 2^2$. This captures the idea that the power of principles increases with the intensity of interference more than proportionally (the "marginal cost" of interference is increasing).

Incommensurability Criticism

At a first glance, it looks like the proportionality test requires the judges to reconcile irreconcilable interests: the freedom of expression versus the right to good reputation, the protection of public safety versus the right to personal privacy, the protection of natural environment versus the freedom of business, the protection of public order versus the freedom of religion, the right to a fair trial versus the economic efficiency of judicial process, and so on.

An important objection against the proportionality test is that the idea of balancing presupposes comparing rights as value bearers without a common metric. The main point of this line of criticism is that the balancing of rights (or the public interest) compares items which are "incommensurable" (see Chang 1997, 2013; Frank 2008; Raz 1986, Chap. 13). The late US Supreme Court Justice Antonin Scalia, who may count as one of the pronounced proponents of the incommensurability objection, put it bluntly: "[T]he scale analogy is not really appropriate since the interests on both sides are incommensurate. It is more like judging whether a particular line is longer than a particular rock is heavy" (Bendix Autolite Corp. v. Midwesco Enterprises 1988).

This variant of the incommensurability objection is based on intuition that we can only compare two things of the same kind so that it is impossible to perform a rational cardinal comparison between gains and losses for colliding rights or the public interest that are incommensurable. How convincing is this objection?

We suggest that the famous idiom about the comparison of apples with oranges should not be used in a mechanical way because comparing things of different kinds can, indeed, be meaningful. Incommensurability need not imply incomparability (Da Silva 2011, see also Chang 2002). Balancing can work in a satisfactory manner as long as comparability among the colliding rights

is rationally justified, no matter whether they are commensurable or not. And for this we would need to know what exactly we are actually comparing during the balancing exercise.

The answer to that may be relatively straightforward. We compare the marginal social importance in fulfilling one right and the marginal social importance in preventing the harm to another right (Barak 2012, p. 484). However, the problem here is that "[i]dentifying a single criterion does not eliminate incommensurability if the application of the criterion depends on considerations that are themselves incommensurable" (Endicott 2014, p. 311). An so one can still be legitimately worried that if the judges are required to balance things that cannot be compared, then the conclusion of such a balancing is nothing more than the outcome of their arbitrary preferences or subjective feelings (Alder 2001, p. 717).

The severity issues raised by the incommensurability criticism notwithstanding the problem with this argument as, we see it, is twofold: First, if one needs to make a decision, it is not of much help to point out that the decision problem is difficult or impossible to crack. One can contemplate indefinitely what is the proper course of action in the trolley dilemma (Edmonds 2014) as long as this happens in a thought experiment. When, however, the trolley is actually hurtling down a track toward the five people tied to it, a decision has to be made whether to push the fat man over the bridge, saving the five, or let them die, saving the fat man. Courts are not in the business of thought experimentation; they need to make the actual decisions.

The second problem with the incommensurability criticism can be seen when we bring it *ad absurdum*. One may argue that individuals make decisions based on comparisons of incommensurable courses of actions on a daily basis. Should I order a steak or a salad? Should I snooze the alarm or get up right now? Should I stay in the office or give my dog a proper walk? It is hard to see how these decisions are made using some common objective scale. Indeed, all of these choices are different from one another along different dimensions, some of which they may not share, making it hard to argue that there is a common "objective" denominator in any of these dilemmas. If the incommensurability argument was valid, all peoples' choices could be regarded as purely incidental. Such an argument would be problematic not least from the legal point of view as the bulk of law presupposes individual rationality and the ability of people to make sensible choices for themselves (see Chang 2012).

Conclusion

When individuals choose the desired course of action, they do so based on their preferences or feelings. The courts are not in a fundamentally different position in this respect, save for two important aspects: (i) unlike individuals who decide for themselves, judges sometimes must make decisions on behalf of the society, and (ii) unlike individuals who face the consequences of their own decisions, the consequences of judges' decisions are born by the society at large or other persons. This makes adjudicating hard cases more delicate, difficult, and prone to error. Indeed, judges are entrusted with decision tasks that people seldom face in their private lives. Proportionality approach forces the judge to explicitly articulate whether and why she evaluates the desired effects of her decision as proportional with respect to the undesired limitations the decisions implied for the other rights, using a common, if abstract, metric for such a comparison and to do so in a manner that is methodical and consistent across cases (see Klatt and Meister 2012).

Proportionality test does not need to deliver a rock solid solution in every case in which it is conducted. Indeed, the test may sometimes yield an ambiguous result. Referring to our weighting formula, this would happen when the resulting weight is equal to one or when alternative but plausible assumptions behind the individual entries (i.e., uncertainty) produce different outcomes of the test. In such stalemate situations, there is no way out other than court's discretion to decide either way (Alexy 2014, see also Chang 2002; Da Silva 2011; Raz 1999, Chap. 3). It goes without saying that where the test is valuable is precisely in cases where it points one direction

with enough certainty. It is these cases where it helps the court to make a sound decision and avoid a bad trade-off.

Incommensurability criticisms raise valid points about the fundamental challenges associated with the proportionality test. Indeed, detailed awareness about the problems concerning the determination of the relative value of rights and infringements may improve the quality of courts' deliberation. However, pointing out that two rights are not commensurable is not very helpful when the court has to decide between them somehow. In the absence of availability of a better alternative decision algorithm, the proportionality appears to be the go-to solution. The value of the proportionality project seems to be in that it forces courts to articulate their arguments in an explicit, structured, and transparent manner so that the court's deliberation and final decision can be understood (Klatt and Meister 2012, see also Winter 2016), even if one might disagree with the weights assigned to individual elements in the weighting exercise.

Cross-References

- ▶ Conflict of Laws
- ▶ Cost–Benefit Analysis
- ▶ Judicial Decision-Making

References

Alder J (2001) Incommensurable values and judicial review: the case of local government. Public Law 4:717–735

Aleinikoff TA (1987) Constitutional law in the age of balancing. Yale Law J 96:943–1005

Alexy R (2002) A theory of constitutional rights. Oxford University Press, Oxford

Alexy R (2014) Constitutional rights and proportionality. Revus 22:51–65

Barak A (2012) Proportionality: constitutional rights and their limitations. Cambridge University Press, Cambridge, UK

Chang R (1997) Introduction. In: Chang R (ed) Incommensurability, incomparability, and practical reason. Harvard University Press, Cambridge, MA

Chang R (2002) The possibility of parity. Ethics 112:659–688

Chang R (2012) Are hard choices cases of incomparability? Philos Issues 22:106–126

Chang R (2013) Incommensurability (and incomparability). In: LaFollette H (ed) International encyclopedia of ethics. Blackwell Publishing, Malden, pp 2591–2604

Da Silva VA (2011) Comparing the incommensurable: constitutional principles, balancing and rational decision. Oxf J Leg Stud 31:273–301

Edmonds D (2014) Would you kill the fat man? The trolley problem and what your answer tells us about right and wrong. Princeton University Press, Princeton

Endicott T (2014) Proportionality and incommensurability. In: Proportionality and the rule of law: rights, justification, reasoning. Cambridge University Press, Cambridge, UK, pp 311–342

Frank RH (2008) Why is cost-benefit analysis so controversial? In: Hausman DM (ed) The philosophy of economics: an anthology. Cambridge University Press, New York, pp 251–269

Klatt M, Meister M (2012) The constitutional structure of proportionality. Oxford University Press, Oxford

Raz J (1986) Morality of freedom. Oxford University Press, Oxford

Raz J (1999) Engaging reason: on the theory of value and action. Oxford University Press, Oxford

Winter J (2016) Alexyho vážící formule. Právník 5:446–461

Prosocial Behaviors

Sébastien Roussel
Department of Economics, CEE-M, Université Montpellier, CNRS, INRA, SupAgro, Université Paul Valéry Montpellier 3, Montpellier, France

Abstract

Prosocial behaviors contribute to the well-being of others following activities like charitable giving or volunteering. Theories that explain prosocial behaviors are pure altruism and impure altruism, inequality aversion reciprocity and conditional cooperation. These theories are linked to a system of motivations, that is intrinsic motivation, extrinsic motivation, and image motivation. Social norms play a key part in one's motivations to behave prosocially, in particular with regard to image motivation and likely sanctions when one's deviates from these norms. The expressive power of law leads to norms legitimacy. Prosocial behaviors are mainly studied in the literature in economics and psychology, especially in behavioral and experimental economics.

Definition

Prosocial behaviors refer to behaviors contributing to the interests of others and not to its own self-interest. Antisocial behaviors stand for the opposite of prosocial behaviors.

Introduction

From the seminal example of blood gift (Titmuss 1970) to the field of charitable giving (Bekkers and Wiepking 2011), many situations show that people act in a prosocial manner. Prosocial behaviors contribute to the well-being of others following activities like charitable giving (monetary giving or sharing for example), volunteering, voting, etc. List (2011) highlights that, in a regular year, total charitable gifts of money represent 2% of the US Gross Domestic Product (GDP), e.g., worth $314 billion in 2007. This latter figure has been increasing to reach consecutive records in 2014 and 2015, with an estimated $373.25 billion in 2015 (Giving USA 2016).

In terms of the history of economic thought, prosocial behaviors have been originally considered through Adam Smith's visionary overview of one individual's rationale. Smith (1759) stated that: "How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it, except the pleasure of seeing it" (cited by Meier 2007) with regards to "*the passions and motives which influence it*" (cited by Bénabou and Tirole 2006). This interpretation of rationality already suggested that decision-makers are not always egoistic. Consequently, one may analyze the role of social preferences/other-regarding preferences and motivations leading to prosocial behaviors (section "Social Preferences and Motivations"), as well as the part played by the institutions and the legal framework through social norms and the expressive function of law (section "Social Norms and Laws"), and how economics model and assess these behaviors (section "Modeling and Assessing Prosocial Behaviors").

Social Preferences and Motivations

In a detailed survey, Meier (2007) underlines the economic theories that explain prosocial behaviors through social preferences. These theories are respectively: pure altruism (Eckel and Grossman 1996) and impure altruism (Andreoni 1990), inequality aversion (Fehr and Schmidt 1999), reciprocity (Rabin 1993), and conditional cooperation (Frey and Meier 2004). While pure altruism and inequality aversion refer to situations where individuals act themselves selflessly, impure altruism, reciprocity, and conditional cooperation suggest that individuals expect to get something back from their prosocial activities.

The theories involving social preferences are linked to a system of motivations that result in prosocial behaviors in different ways. Ariely et al. (2009) describe the set of motivations that leads people to adopt prosocial behaviors. These motivations are in a broad sense: intrinsic motivation, extrinsic motivation, and image motivation. Intrinsic motivation refers to prosocial preferences like altruism, while extrinsic motivation indicates how external factors like monetary rewards matter in acting prosocially. Image motivation deals with the impacts of the others' perception and the individual's self-esteem on prosocial behaviors, and how this may frame one's behavior. However, the interplay between these motivations is critical. For example, many studies in economics and psychology show how intrinsic and extrinsic motivation may be complement or substitute, leading to crowding in – crowding out phenomena (Frey and Oberholzer-Gee 1997). This is also linked to the key role played by social norms.

Social Norms and Laws

Social norms play a key part in one's motivations to behave prosocially, in particular with regard to image motivation and likely sanctions when one's deviates from these norms. Social norms rely on a consensus towards acceptable behaviors (Elster 1989; Fehr and Schmidt 1999). Viscusi et al. (2011) characterize social norms "in terms of what is normatively appropriate rather than what

is the conventional mode of behavior," and study how personal and external norms matter in recycling behavior. Another classical distinction refers to injunctive norms and descriptive norms, where the former characterizes the perception of what should be socially approved, and the latter characterizes the perception of what is effective. What is at stake is to conform or not to these norms, along with the individual valuation of what is morally good or wrong, and how these norms are enforced. Besides, identity plays a meaningful role while individuals follow prescriptions or social norms to secure their self-concepts (Akerlof and Kranton 2000).

The institutionalization of social norms into laws has been underlined by McAdams (2015) through the expressive powers of law, allowing for the recognition of good behavior versus bad behavior, leading to injunctive norms legitimacy. Moreover, he shows that law contributes to a coordination function, discloses information and beliefs, and provides compliance rules as explicit incentives. This was previously highlighted in the literature by the expressive function of law (Sunstein 1996a, b). Public authorities disclose the social meaning of one's activities to help people make desired choices in a so-called norm management. In addition, Funk (2007) provides an empirical analysis of voting turnout and identifies the greater impact on behavior when law targets at the civic duty.

Modeling and Assessing Prosocial Behaviors

Prosocial behaviors are mainly studied in the literature in economics and psychology, especially in behavioral and experimental economics. Participation in prosocial activities is then traditionally modeled as a contribution to a public good (Andreoni 1990). Regarding the expressive content of laws, it has been shown how this can be modeled in the presence of norms or reputational rewards following Bénabou and Tirole (2006, 2011). In the literature, economic games are used in experiments to test hypotheses on social preferences and motivations that lead to prosocial behaviors. In these experiments, a prosocial behavior like giving more is desirable (monetary or non-monetary contribution), and economic incentives relying (or not) on intrinsic, extrinsic, and image motivations can be implemented.

Traditional economic games used in the experiments on prosocial behaviors are mainly the ultimatum game, the (repeated) public good game, or the dictator game. For example, the dictator game allows testing individual generosity (List 2007), where the recipient must accept the sum offered by the dictator; consequently, the dictator's behavior is then not strategic. It has also been shown that a social link between the dictator and the recipient increases the incentive to donate (Engel 2011), in particular when the recipient is a nongovernmental organization (NGO) like the Red Cross (Eckel and Grossman 1996). Furthermore, psychological and biological factors drive prosocial behaviors in expressing the motivations stated above. Moods and emotions do influence prosocial behaviors in various ways. For instance, gratitude plays a key role in building relationships and shaping costly prosocial behaviors (Bartlett and DeSteno 2006), awe increases donation amounts (Ibanez et al. 2017), whereas anger can impact negatively these behaviors (Drouvelis and Grosskopf 2016). With regards to the impact of laws on behavior, Galbiati and Vertova (2014) show from alternative public good games that obligations (laws and other formal rules) and economic incentives are complementary to incentivize cooperative behaviors.

Cross-References

- Altruism
- Intrinsic and Extrinsic Motivation
- Public Goods

References

Akerlof GA, Kranton RE (2000) Economics and identity. Q J Econ 115(3):715–753

Andreoni J (1990) Impure altruism and donations to public goods: a theory of warm-glow giving. Econ J 100: 464–477

Ariely D, Bracha A, Meier S (2009) Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. Am Econ Rev 99(1):544–555
Bartlett MY, DeSteno D (2006) Gratitude and prosocial behavior. Psychol Sci 17(4):319–325
Bekkers R, Wiepking P (2011) A literature review of empirical studies of philanthropy: eight mechanisms that drive charitable giving. Nonprofit Volunt Sect Q 40(5):924–973
Bénabou R, Tirole J (2006) Incentives and prosocial behavior. Am Econ Rev 96(5):1652–1678
Bénabou R, Tirole J (2011) Laws and norms. The National Bureau of Economic Research (NBER) working paper no. 17579
Drouvelis M, Grosskopf B (2016) The effects of induced emotions on pro-social behaviour. J Public Econ 134:1–8
Eckel CC, Grossman PJ (1996) Altruism in anonymous dictator games. Games Econ Behav 16:181–191
Elster J (1989) Social norms and economic theory. J Econ Perspect 3(4):99–117
Engel C (2011) Dictator games: a meta-study. Exp Econ 14(4):583–610
Fehr E, Schmidt K (1999) A theory of fairness, competition, and cooperation. Q J Econ 114(3):817–868
Frey BS, Meier S (2004) Social comparisons and pro-social behavior: testing conditional cooperation in a field experiment. Am Econ Rev 94(5):1717–1722
Frey BS, Oberholzer-Gee F (1997) The cost of price incentives: an empirical analysis of motivation crowding-out. Am Econ Rev 87(4):746–755
Funk P (2007) Is there an expressive function of law? An empirical analysis of voting laws with symbolic fines. Am Law Econ Rev 9:135–159
Galbiati R, Vertova P (2014) How laws affect behavior: obligations, incentives and cooperative behavior. Int Rev Law Econ 38:48–57
Giving USA (2016) The annual report on philanthropy for the Year 2015. The Indiana University Lilly Family School of Philanthropy
Ibanez L, Moureau N, Roussel S (2017) How do incidental emotions impact proenvironmental behavior? Evidence from the dictator game. J Behav Exp Econ 66:150–155
List JA (2007) On the interpretation of giving in dictator games. J Polit Econ 115(3):482–493
List JA (2011) The market for charitable giving. J Econ Perspect 25(2):157–180
McAdams RH (2015) The expressive powers of law: theories and limits. Harvard University Press, Cambridge, MA
Meier S (2007) A survey of economic theories and field evidence on pro-social behavior. In: Frey BS, Stutzer A (eds) Economics and psychology: a promising new cross-disciplinary field. MIT Press, Cambridge, MA, pp 51–88
Rabin M (1993) Incorporating fairness into game theory and economics. Am Econ Rev 83(5):1281–1302
Smith A (1759) The theory of moral sentiments. Prometheus Books, Amherst
Sunstein CR (1996a) On the expressive function of law. Univ Pa Law Rev 144:2021–2053
Sunstein CR (1996b) Social norms and social roles. Columbia Law Rev 96:903–950
Titmuss RM (1970) The gift relationship: from human blood to social policy. Allen and Unwin, London
Viscusi WP, Huber J, Bell J (2011) Promoting recycling: private values, social norms, and economic incentives. Am Econ Rev 101(3):65–70

Prostitution

► Sex Offenses

Prostitution, Demand and Supply of

Øystein Hernæs[1,2], Niklas Jakobsson[3] and Andreas Kotsadam[4]
[1]Department of Economics, European University Institute, Florence, Italy
[2]Norwegian Social Research, Oslo, Norway
[3]Oslo and Akershus University College of Applied Sciences, Norwegian Social Research and Karlstad University, Oslo, Norway
[4]Department of Economics, University of Oslo, Oslo, Norway

Abstract

The market for sex is a contentious one, and has often been subject to heavy regulation. This chapter goes through factors that are important with regards to the demand for and supply of prostitution. A particular focus is on the relationship between laws and the quantity of sex bought and sold. The most common way that laws have been used to affect the demand for prostitution is by outright criminalization, which may lead to less prostitution, but may also drive the activity further underground. The effect of criminalizing prostitution on trafficking is ambiguous since criminalization may also lead to a substitution effect towards more trafficked prostitutes. Scarcity of reliable data is one of the main challenges for the study of prostitution.

Definition

The demand and supply of prostitution

Introduction

Buying and selling of sex can be analyzed as any regular economic exchange. The demand side is made up of customers who compensate the suppliers, prostitutes, for the commodity, or service, of sex. This all takes place within some regulatory framework.

The market for sex is a contentious one and has often been subject to heavy regulation. Both these facts have spurred academic research, as this chapter will show. However, the field has until recently not been very well developed, mostly because it has been plagued by a scarcity of reliable data. The lack of data is due to both the fact that prostitution in many societies is illegal or just at the margin of being illegal and the social stigma attached to being associated with the trade.

This chapter goes through factors that have found to be important with regard to the demand for and supply of prostitution. A particular focus is on the relationship between laws and quantity of sex bought and sold. It ends with an identification of gaps in the literature and a discussion of some underexplored questions and fruitful avenues for further research.

Demand

Amenable to Change Through Laws

The most common way that laws have been used to affect the demand for prostitution is by outright criminalization. The possibility of a penalty is an unambiguous (expected) cost for a potential buyer and will lead to less prostitution. This mechanism is confirmed empirically by Jakobsson and Kotsadam (2012), who investigate the effect of criminalization of buying sex on the amount of sex bought. Using longitudinal data from Norway and Sweden and employing a difference-in-differences approach, they estimate that the criminalization of buying sex in Norway in 2009 leads to a downward change in the amount of sex bought of 3.5% points. They do not find an effect on attitudes in this sample and conclude that the most likely mechanism is an increased risk of getting caught. A serious problem in this type of studies is that they rely of self-reporting, in particular self-reporting of activities that are not only highly stigmatized but also change legal status during the period investigated. The authors also use different types of measures, one direct whereby people are asked if they had bought sex during the last 6 months and one more indirect where the respondents are asked if they know someone having bought sex during the last 6 months. While it is likely that respondents are more likely to report on people they know, it is still an inherent problem that the people they know perhaps keep such information more to themselves after it has become illegal.

If demand is relatively inelastic, criminalization of clients may to a large extent drive the activity further underground, with a corresponding increase in risk and worsened working conditions for prostitutes.

Laws may also have a normative function. In particular, it may have an expressive function in signaling a society's beliefs about what is considered acceptable behavior. Jakobsson and Kotsadam (2011) analyze whether the criminalization of buying sex in Norway in 2009 affected attitudes to prostitution. Again using longitudinal data from Norway and Sweden and employing a difference-in-differences approach, they find that the law made people in the capital more negative toward buying sex, but no effect on moral attitudes in the aggregate. They explicitly relate this to a framework whereby people that are most affected by legal changes are most likely to change their attitudes and argue that since people in the capital area were more exposed to street prostitution before the law and saw a large immediate reduction afterward, they should be more likely to be affected. They also find that young people were more affected in their attitudes by the law, which they link to socialization theory arguing that young people are more adept to changing their moral values. On the other side, passing laws that are difficult to enforce or against which there

is significant disagreement may weaken the respect for laws in general.

Della Giusta et al. (2009b) model reputation and stigma explicitly. Loss of reputation may be incurred by both buyers and sellers of sex and reduce the amount of sex exchanged. As described in the studies above, moral attitudes may not be very responsive to the law, but criminalization certainly carries social stigma in itself. Although Jakobsson and Kotsadam (2012) do not find an effect on moral attitudes of Norway's prostitution law, it is possible that the law increased the stigma related to buying sex and thereby added to the formal penalty. Della Giusta (2010) suggests that politicians adopt overly repressive regimes to please conservative swing voters.

Della Giusta et al. (2009a) delve empirically into the demand side and find that stigma does play an important role. They also find that demand may be multifaceted, e.g., in terms of risk preferences, and stress the importance of recognizing this when designing policy.

Important Factors Less Responsive to Legal Change

Insofar as sex is a normal good, the amount bought depends positively on an individual's income. However, prostitution is more common in poor countries. This will often have to do with supply, discussed below, but higher income of the customers may also decrease prostitution. Edlund and Korn (2002) propose a theoretical model in which increases in male (customer) income decrease the number of prostitutes because men prefer women as wives rather than as prostitutes. Relatedly, men's marriage opportunities may to a large extent determine their demand for prostitution and itself be determined by income, although married men constitute a large share of the demand. This suggests that policies and regulations promoting stable employment opportunities and a social safety net are relevant tools for authorities concerned with prostitution.

Noncommercial nonmarital sex and commercial nonmarital sex are typically thought of as substitutes for marital sex (Edlund and Korn 2002), but the relation between commercial and noncommercial nonmarital sex is ambiguous. On factors such as pornographic material, a society's sociosexual culture, and the distinctness of gender roles, there is also little consensus.

Sex ratios in the population are also thought to be important. Places where men have outnumbered women to a significant degree have largely seen more prostitution. As Edlund and Korn (2002) point out, these places have tended to consist of men in transit, who participate only in a locality's sex market and not in its marriage market.

Supply

In some countries, selling sex is illegal while buying is legal, but a number of countries maintain the illegality of both sides of the market. When selling sex is illegal, the potential prostitute faces formal sanctions in addition to social stigma. The occupation of prostitute is thus made less appealing, and the woman in question is perhaps led to consider other even less desirable options. If prostitution does not entail any negative externalities and criminalizing selling sex is bad for women in the business, lawmakers should of course think twice before enacting such laws. What women do, and even if they do change occupation, after it becomes illegal is, however, unknown, but there are indications of severe externalities in terms of sexual trafficking, so a firm policy conclusion cannot be reached at this stage.

The most obvious factor affecting the supply of prostitution is income relative to other occupations and in general employment opportunities for women. As with any other type of work, the better potential labor market elsewhere, the less supply there will be of workers in a particular trade. As with the demand side, this suggests a role for providing employment opportunities and a social safety net in reducing prostitution. Other factors that impact the profitability of prostitution are sex ratios and inequality. To the best of our knowledge, there are no empirical studies investigating these issues.

Another standard factor in occupational choice is working conditions. As an effect of being

undertaken outside the formal economy, prostitution has often lacked regulation and oversight. In general, prostitutes face a more risky working environment than other laborers. In Europe, a significant minority of countries have legalized prostitution and tried to regulate the industry (Jakobsson and Kotsadam 2013). Safe working conditions, access to health services, and the rights to public benefits are all factors that make prostitution relatively more tempting. Regulations denying these to an occupation, as is largely the case for prostitution, can therefore be expected to lead to less selection into that profession. Any form of criminalization will likely lead the mere advocation of basic rights and labor standards more difficult.

Prostitution is distinct from most other professions in that it carries a significantly higher opportunity cost in terms of social stigma. Della Giusta et al. (2009b) model reputation explicitly. Criminalization will mostly affect the stigma attached to the side or activity that is criminalized; however, it might spill over to other sides as well. For instance, outlawing buying of sex could increase the stigma not only of buying sex but also of selling. Edlund and Korn (2002) stress foregone marriage opportunities as the most important opportunity cost of prostitution. This is due to limited mobility between the marriage and prostitution sectors, a feature that also explains prostitutes' relatively high (up-front) wages. However, Arunachalam and Shah (2008) find no empirical support for the marriage-based explanation of the earnings premium in sex work.

There have been many instances in which women have been coerced to work as prostitutes. Forced labor is now outlawed in almost all countries – prostitution legislation aimed at alleviating this problem has therefore often been aimed at restricting the market, in the hope that this will hurt the potential profits of traffickers. Related to this, there has recently been developed a separate literature on trafficking. Although any kind of criminalization will reduce prostitution, the effect on trafficking is ambiguous, since criminalization may also lead to a substitution effect toward more trafficked prostitutes (Cho et al. 2013). Jakobsson and Kotsadam (2013) and Cho et al. (2013) investigate this question using cross-country data and case studies and find that harsher sex laws are associated with less trafficking.

Notwithstanding the problem of forced labor in the sex trade, a significant part of international sex labor migration can be seen as voluntary and economic in motivation. Thus, policies that promote freer movement of labor and fewer restrictions on travel can be expected to increase the amount of prostitution. On the other hand, better opportunities to freely travel and search for other jobs may reduce the possibilities of human traffickers to exploit or deceive vulnerable migrant and force them into the sex trade.

A general issue with regard to trafficking is the tension between the welfare of potential trafficking victims and that of current prostitutes. Restricting the market for prostitution, either through the denial of benefits to prostitutes or some form of criminalization, is likely to lead to prostitution being less profitable and thus less tempting for traffickers but is also likely to hurt those currently in the profession. Lee and Persson (2013) propose a regulatory framework that provides a solution to this conundrum: licensing prostitutes and criminalizing buying sex from unlicensed prostitutes. This model deters demand for nonvoluntary prostitutes and thus reduces the potential profit for traffickers while at the same time safeguarding the interests of licensed prostitutes.

Technological change has impacted the prostitution market. Cunningham and Kendall (2011) examine the use of the Internet in the prostitution business. They find that online soliciting has displaced some of the off-line market but that the net effect is an increase in supply. They also find that prostitutes soliciting online are largely engaging in lower-risk behaviors than their street-level counterparts, although former street workers may retain their higher-risk behavior also online.

Gaps in the Literature

Scarcity of reliable data is one of the main challenges for the study of prostitution. Collection of

new, reliable data is one of the most worthwhile initiatives with regard to prostitution research. This could be in the form of surveys, anonymous if necessary, but in societies where prostitution is legal, the ambitions should be even higher. For instance, if the framework of mandatory registration of legal prostitutes proposed by Lee and Persson (2013) is adopted, data from those registration procedures should be collected, kept safe, and be available for researchers. In general, regulation will help data collection, due to the fact that one is not collecting data on something criminal and that there will be data emanating from the enforcement of regulations, whether they are about working conditions, health, or taxes.

Even where prostitution is not legal, there are opportunities to collect more and better data and develop better measures, for instance, through the use of list experiments (also called list treatment or the item count technique).

There has been little research on how different forms of criminalization or regulation affect the working conditions of prostitutes. Di Tommaso et al. (2009) use data from the Anti-Trafficking Unit of the International Organization for Migration and find that the well-being of trafficked women is (further) worsened when having to work in secluded spaces.

Criminalization legislation is often the result of an alliance between one side consisting of people who are opposed to prostitution as a matter of principle and another that is concerned about coercion of involuntary prostitutes. The stability of this alliance remains an open question.

Whether demand for voluntary and nonvoluntary prostitutes is uniform is an important question that has not been much studied. For instance, if a substantial share of demand is for the voluntary product, that would be valuable to take into account when designing regulation, as shown theoretically by Lee and Persson (2013). Della Giusta et al. (2009a) take some steps toward disaggregating the demand side, but they do not specifically focus on if demanders care about voluntariness.

There are also ways in which the supply side can be disaggregated further. Globalization has made movement across borders easier. Cho et al. (2013) and Jakobsson and Kotsadam (2013) investigate the relation between sex laws and trafficking, but it would be valuable to know also about the more direct relation between the enforcement of laws against trafficking and coerced labor on the one side and involuntary trafficking on the other.

Topics on which some hold strong moral attitudes can easily become politicized and subject to limited inquiry. So also with prostitution; it is therefore particularly important that researchers in this field are allowed to remain independent when addressing the topics described in this chapter.

Cross-References

- ▶ Crime and Punishment (Becker 1968)
- ▶ Gender Diversity
- ▶ Liberty
- ▶ Sex
- ▶ Sex Offenses
- ▶ Shadow Economy
- ▶ Underground Economy

References

Arunachalam R, Shah M (2008) Prostitutes and brides? Am Econ Rev 98(2):516–522

Cho S-Y, Dreher A, Neumayer E (2013) Does legalized prostitution increase human trafficking? World Dev 41:67–82

Cunningham S, Kendall TD (2011) Prostitution 2.0: the changing face of sex work. J Urban Econ 69(3):273–287

Della Giusta M (2010) Simulating the impact of regulation changes on the market for prostitution services. Eur J Law Econ 29(1):1–14

Della Giusta M, Di Tommaso ML, Shima I, Strøm S (2009a) What money buys: clients of street sex workers in the us. Appl Econ 41(18):2261–2277

Della Giusta M, Di Tommaso ML, Strøm S (2009b) Who is watching? the market for prostitution services. J Popul Econ 22(2):501–516

Di Tommaso ML, Shima I, Strøm S, Bettio F (2009) As bad as it gets: well-being deprivation of sexually exploited trafficked women. Eur J Polit Econ 25(2):143–162

Edlund L, Korn E (2002) A theory of prostitution. J Polit Econ 110(1):181–214

Jakobsson N, Kotsadam A (2011) Do laws affect attitudes? an assessment of the norwegian prostitution law using longitudinal data. Int Rev Law Econ 31(2):103–115
Jakobsson N, Kotsadam A (2012) Shame on you, john! laws, stigmatization, and the demand for sex. Eur J Law Econ 37(3):1–12
Jakobsson N, Kotsadam A (2013) The law and economics of international sex slavery: prostitution laws and trafficking for sexual exploitation. Eur J Law Econ 35(1):87–107
Lee S, Persson P (2013) Human trafficking and regulating prostitution. New York University School of Law, Law & economics research paper series, working paper no. 12–08

Protective Factors

Norbert Hirschauer[1] and Sebastian Scheerer[2]
[1]Agribusiness Management, Institute of Agricultural and Nutritional Sciences, Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany
[2]Institute for Criminological Social Research, Faculty of Economics and Social Sciences, University of Hamburg, Hamburg, Germany

Abstract

Protective factors help people achieve positive outcomes despite exposure to potentially negative influences (in medicine, good health in spite of smoking cigarettes; in the social sciences, productive adult lives as law-abiding citizens despite adverse conditions during childhood and adolescence). In Law and Economics, the concept, in conjunction with control theories, contributes to a better understanding of why economic actors obey the law and comply with rules despite exposure to material incentives to the contrary. Protective factors can result from external sources (social control) and internal sources (internalized conventional values). They generate nonmaterial benefits in the case of compliance and/or nonmaterial costs in the case of noncompliance. The inclusion of protective factor into regulatory analysis – in combination with the consideration of risk factors that work in the opposite direction – helps to understand why some regulatory regimes work better than others. The concept can be especially useful when it comes to designing regulatory innovation based on a better understanding of why some economic actors obey the law and others don't.

Definition

Protective factors are individual and/or environmental characteristics that reduce people's vulnerability to adversities such as natural, health, or moral hazards.

Protective Factors in Law and Economics and Criminology

Conceptual Background

The concept of *protective factors* owes its existence to the attempt to explain the curious fact that some people experience positive outcomes in the face of highly adverse conditions. If some heavy smokers, for instance, manage to live in good health until their 90th birthdays and beyond, then it is plausible to assume that there is something – some kind of protective factor – that shielded them from contracting any of the diseases commonly associated with inhaling the toxic mixture of over 4,000 chemicals called cigarette smoke. Analogous phenomena of *resilience* can be found in relation to delinquent life trajectories. The quest to explain surprisingly positive life courses of people from dysfunctional families and highly delinquent neighborhoods will reveal characteristics and/or resources that worked in favor of these individuals and guarded them against negative influences. Protective factors may also play a role in the context of *Law and Economics* and, more specifically, rational choice theories. Here, the concept can explain why (some) people obey rules in spite of *moral hazards* and an abundance of profitable and convenient opportunities to break them, i.e., despite an environment with a high degree of adversity to *compliance* with rules.

When using the term "protective factor" in *Law and Economics*, one implicitly resorts to

principal agent theory and adopts the dichotomous perspective of a *principal* (e.g., a lawmaker or regulator) who has defined two types of *agents*' behavior with regard to a rule: noncompliance (undesired behavior) as opposed to compliance (desired behavior). The concept's analytical potential is especially useful when the term "protective factors" is limited to designate nonmaterial factors associated with social control and internalized norms that encourage compliance in an environment where material factors would favor noncompliance. This requirement is met by the following definition:

> **Protective factors** are characteristics in individuals and/or their socio-economic environments that discourage actors from rule-breaking by causing nonmaterial benefits (utility) in the case of compliance and nonmaterial costs (disutility) in the case of non-compliance.

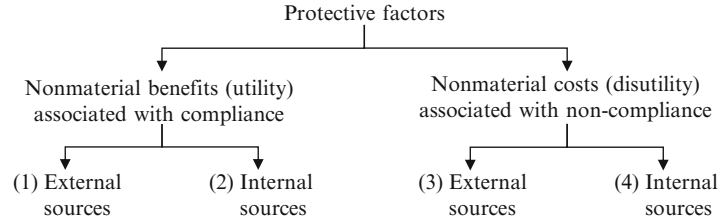
This definition is a useful tool to transcend an all too *narrow rational choice* conception with its restrictive assumption of *utility* hinging exclusively on material wealth. It turns our attention to the fact that not only law works "as a means for changing relative prices attached to individual actions" (Parisi 2004, p. 262), but that there are also nonmaterial factors influencing those "prices" for the individual actor. It corresponds with the understanding that people often pursue multiple goals and strive not only for wealth but also for *social recognition* and distinction as well as for consistency with their *internalized values* and identity (Lösel and Bender 2003; Akerlof and Kranton 2010). Depending on the situation, utility gains from complying with rules may, or may not, outweigh temptations to break them (Pinstrup-Andersen 2005). In other words, the utility of multi-goal decision-makers with *bonds to social norms* is not in all cases monotonically increasing in expected material wealth. Instead, such actors are prepared to pay an *ethical premium* that shields them from deviance if – but only if – it exceeds the economic temptation to break the rules. The ethical premium is the utility outcome produced by protective factors. It provides an operational definition of resilience which, in turn, is the antonym of *vulnerability*: The stronger an individual's protective factors, the higher her/his ethical premium and the higher (lower) her/his resilience (vulnerability) to moral hazards.

Figure 1 illustrates that we can distinguish four classes of protective factors: On the one hand, there are factors that provide nonmaterial benefits to the individual in case of compliance. These benefits may result from (1) external sources (e.g., social respect) or (2) internal sources (e.g., self-respect). On the other hand, there are factors that cause nonmaterial costs in the case of noncompliance. These costs may also have (3) external sources (e.g., ostracism) or (4) internal sources (e.g., conflict with self-image).

Adopting the perspective that, besides economic wealth, people may seek social respect and inner contentment necessarily requires juxtaposing protective factors with their opposite that has been labeled *risk factors* (and is sometimes also referred to as *criminogenic factors*). Risk factors can be defined as follows:

> **Risk factors** are characteristics in individuals and/or their socio-economic environments that encourage actors to break rules by causing nonmaterial benefits (utility) in the case of non-compliance and nonmaterial costs (disutility) in the case of compliance.

Analogous to protective factors, four classes of risk factors can be distinguished: An individual's integration into a deviant subculture, for instance, may provide (1) nonmaterial external benefits for



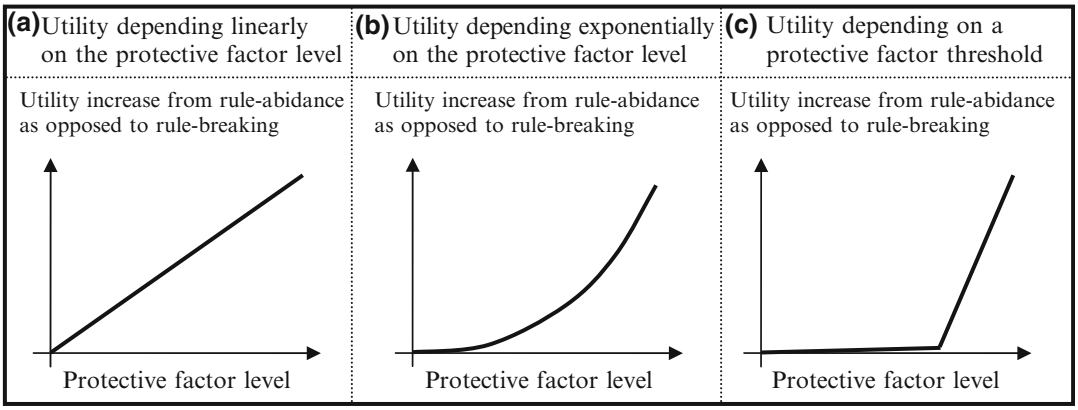
Protective Factors, Fig. 1 A decision-theoretic classification of protective factors

the rule breaker (e.g., respect by deviant peers). A lack of conformity with the values behind the rules (termed conventional in the rest of this essay) or an outright internalization of deviant norms may generate (2) nonmaterial internal benefits for the deviant individual (affirmation of the deviant self). Prominent examples are illicit behavior due to *reactance* (Brehm 1966; Miron and Brehm 2006) and *altruistic rule breaking*. (3) Compliance in such settings, in contrast, may give rise to ostracism by the deviant peer group and (4) cause conflict with the individual's deviant self-image. In brief, nonmaterial costs for obeying rules and nonmaterial benefits for disobeying can be conceived as individual-level consequences of risk factors that would cause illicit behavior even in the absence of economic temptations to break the rules (hence, the alternative label "criminogenic factors").

Including the term "risk" into the label that has been attached to the nonmaterial factors encouraging deviance reflects three important conceptual aspects: First, risk factors are conceived as random variables. Second, they impact the probability distribution of the behavioral outcome (i.e., the *behavioral risk*). Third, there is an unambiguous direction of influence: Risk factors increase the probability of the undesired behavior on the part of the agents. While not semantically reflecting the aspect of risk, protective factors share the first two qualities of risk factors: They are random variables and they impact the behavioral risk. They work in the opposite direction, however, and decrease the probability of the undesired behavior.

In empirical research on how protective factors and risk factors jointly impact behavioral risks, we will encounter three unknowns: the protective factors and risk factors that are at work in a specific context, the level of these factors, and, equally important, the interactions and functional forms that link them to utility. It will thus be doubtlessly difficult to quantify how much utility people derive from nonmaterial sources in given contexts. We know, however, that they will only abstain from profitable rule breaking if their willingness to renounce profits (ethical premium), as resulting from the balance of protective factors and risk factors, exceeds the monetary benefits from illicit behavior.

Adopting a ceteris paribus perspective and thus abstracting once again from risk factors (and their interaction with protective factors), Fig. 2 illustrates how protective factors work: An individual's utility for rule abidance as opposed to rule breaking increases monotonically in the level of some upstream variable acting as a protective factor. The precise form of the functional relationship varies depending on factor and context. In the simplest case, it will be linear as depicted in (a). If the functional relationship is to reflect a self-reinforcing mechanism, it will be exponential as depicted in (b). If a threshold needs to be exceeded before a protective factor generates a significant



Protective Factors, Fig. 2 Functional relationships linking protective factors to utility and behavior

impact, the function will include a clear discontinuity and may be divided into two linear segments as depicted in (c).

Finally, it should be noted that even a *broad rational choice* conception that includes utility derived from social recognition and inner harmony is often insufficient to explain and predict – with reasonable success – the behaviors of people facing specific rules and specific contextual conditions. Instead, one needs situation-specific models or "minitheories" (Korobkin and Ulen 2000) that facilitate a reconstructing understanding of idiosyncratic or group-specific decision-making processes in particular real-life contexts. Such models do not have to be isomorphic with respect to the "objective" rules of the world but to the world as subjectively perceived and evaluated by the individual or specific subpopulation under consideration (Rubinstein 1991). In other words, one needs models that tackle the gap between narrow rational choice predictions and actual behavior (Garoupa 2003). Such models need to reflect behavior as the result of *multi-goal* decision-making by (potentially) rule breaking and *bounded rational* actors (cf. Simon 1957; Gigerenzer and Selten 2001) who *subjectively* form expectations and evaluate likely outcomes at the time they make their choices.

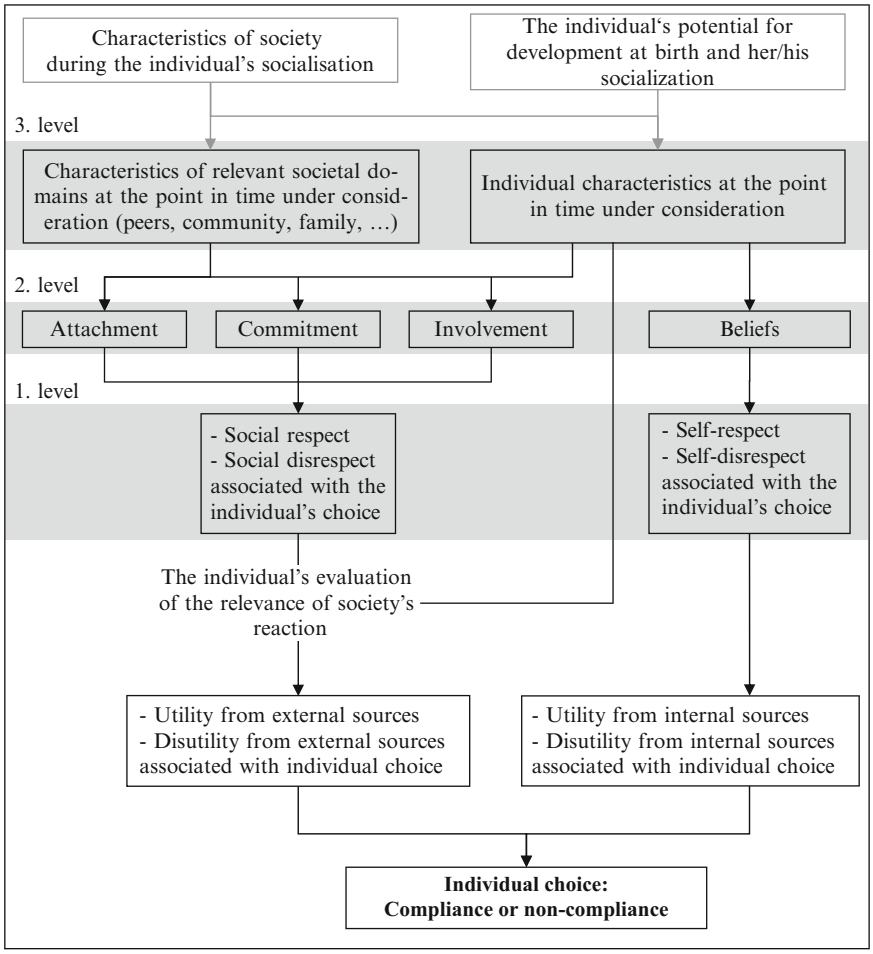
Levels of Analysis

The decision-theoretic definition of protective factors provided above facilitates an operational understanding of their utility outcomes and a clear separation from risk factors. Focusing on the utility outcomes of the individuals under consideration, however, does not yet tell us which real-life characteristics produce these outcomes and thus make people resilient to the seductive potential of illicit behavior. Figure 3 describes the different levels of the *cause-effect chain* on which these factors can be found.

On a first upstream level in the cause-effect chain, we can distinguish external causes of utility (social respect and recognition, social disrespect and ostracism) from internal causes (self-respect and self-affirmation, self-disrespect and self-loathing). Regarding external utility sources, a conceptual distinction between a given level of respect/disrespect shown by others and the individual's evaluation of that level may be useful. It does not seem to make sense, however, to attempt an operational distinction between a given level of one's own self-respect/self-disrespect and one's own evaluation of that level. At best, it would require a schizophrenic separation of the self in two parts.

On a second upstream level of the cause-effect chain, four major types of factors acting as protective mechanisms and determining the factors on the first level have been distinguished (Hirschi 1969): first, social integration into conventional society and emotional bonds to relevant others (elders and peers) who cherish rule abidance and repudiate rule breaking (*attachment*); second, hostage posting to conventional society through reputation as an honest business person or holder of public office that has been accumulated from earlier "investments" in conventional life and that is lost if outlawed behavior is disclosed (*commitment*); third, engagement in conventional activities that consume time and mental energy and thus help the individual abstain from the often complex and time-consuming actions of rule breaking (*involvement*); and fourth, identification with the conventional system of values that are "behind the rules," including the recognition of authority, the emulation of role models, and an appreciation of rule abidance in itself (*beliefs*).

Going further back the causal chain, attempts have been made to systematize protective factors "behind" the factors on the second level. They all list and classify variables believed to be positive influences still further upstream. Commonly used classifications categorize these upstream factors into domains such as *peers, community*, *family*, and *individual*. Accordingly, variables such as long-lasting relationships and trust in the (business) community, a functioning social fabric in neighborhoods and families, low existential stress, and emotional stability are listed as protective factors. All these compilations face the difficult task to avoid gaps, duplications, and inconsistent listings of variables that, while representing behavioral correlates such as gender



Protective Factors, Fig. 3 Protective factors: cause-effect chain and levels of analysis

or ethnic affiliation, are either not causes of behavior in the sense of this word or situated on different levels of the cause-effect chain. As indicated in Fig. 3, one could go even further back in time and causation and compile protective factors on still further upstream levels. Whatever the focal level of analysis, inconsistent listings are futile descriptive exercises with little practical potential to indicate the variables that preventive strategies could, and should, tackle to produce more compliant behavior.

Relation to Control Theories

The decision-theoretic concept of protective factors can be linked to the view of human nature in criminological *control theories* (cf. Hirschi 1969; Gottfredson and Hirschi 1990; Tittle 2000). Contrary to conventional criminological theories, which used to explain crime and deviance by searching for criminogenic factors in the delinquent subject and her/his environment, control theories contend that it is not the breaking of rules that needs explanation in the first place but rather compliance. In the view of control theorists, humans are basically free to deviate from rules following their own interest. It is therefore the challenge of every society to gradually and effectively "deprive" individuals of their original freedom to deviate by creating bonds that make it increasingly difficult for a well-socialized member of society to act against social and legal restraints.

In their attempts to understand the processes by which societies deprive people of their original freedom to deviate, control theorists analyze how social bonds are created through both the attachment to others and the internalization of value systems in line with social expectations. While the former generates external (social) control and rewards/sanctions, the latter leads to internal (self-)control and rewards/sanctions. If society fails to create sufficient social bonds, man's self-interest and original freedom to deviate, in conjunction with economic temptations, continue to induce rule breaking. Table 1 summarizes the core concept of control theories using an illustrative numerical example.

Our point of departure is an individual who faces a misdirected incentive of –80 (economic temptation) to break some rule. Due to protective factors – both in the form of bonds to relevant others and internalized values – the individual is prepared to pay an ethical premium totaling 55. Since this premium does not exceed the economic temptation, we would conclude that the individual is not sufficiently shielded from the natural inclination of self-interested rule breaking. Reflecting the basic concern of control theories with the gaps and holes in the system of social bonding and control, the numerical example in Table 1 is based on the assumption that deviant behavior (e.g., *white-collar crime*) is caused by economic temptations that coincide with lacking protective factors rather than by risk factors. This is why nonmaterial cost from obeying and nonmaterial benefits from disobeying, while being mentioned, are assumed to be zero.

Assuming risk factors to be inexistent may be an adequate analytical simplification in some situations; it may be completely inadequate in others. The degree of resistance to a given material temptation results, in principle, from the balance of protective factors and risk factors. An individual may, for instance, simultaneously be a member of social groups with conflicting value systems. As a result, she/he will strive not only for the social recognition within conventional society but also for the respect of deviant elders and peers, thus possessing external protective factors and external risk factors at the same time. Even two souls may be dwelling in an individual's breast. An example would be a generally law-cherishing individual who, for some reason or another, feels an internal pressure to engage in altruistic rule

Protective Factors, Table 1 An illustrative example for the concept of control theories[a]

	Utility for obeying		Utility for disobeying		Utility differential
Material benefits (+) and costs (−)		**+700**		**+780**	**−80**[b]
External benefits (+) and costs (−) from social respect and disrespect in conventional society	*Protective factors*	+10	*Protective factors*	−25	+35[c]
External benefits (+) and costs (−) from social respect or disrespect in deviant subculture	*Risk factors*	0	*Risk factors*	0	0
Internal benefits (+) and costs (−) from consistency and conflict with internalized conventional values	*Protective factors*	+5	*Protective factors*	−15	+20[d]
Internal benefits (+) and costs (−) from reactance or internalized values of deviant subculture	*Risk factors*	0	*Risk factors*	0	0
Expected benefits (+) and costs (−) from nonmaterial sources		**+15**		**−40**	**+55**

[a]For risk-averse actors who deduct risk premiums from risky benefits and add risk premiums to risky costs, the figures are to be conceived as units of utility. For risk-neutral actors, they can be directly interpreted as payoffs and payoff equivalents.
[b]The negative sign of the differential indicates a misdirected incentive (economic temptation).
[c]Ethical premium caused by external sources.
[d]Ethical premium caused by internal sources.

breaking in a certain context. Another example would be a situation of reactance in which an actor finds inner satisfaction by breaking an imposed rule while simultaneously feeling uneasy or even guilty for breaking the law. In both cases, internal protective factors and internal risk factors are at work at the same time. While risk factors are conventionally considered in criminological research on juvenile delinquency and illicit drug use, they are often omitted from the analysis of economic deviance. However, reactance, deviant cultures, and the internalization of deviant norms can be important in business contexts too. Unless we have contrary contextual knowledge, risk factors should thence not be a priori ignored in the analysis of economic deviance.

The adequacy of a behavioral model depends on the context and the level of analysis (cf. Hess and Scheerer 2004). On the *macro-level* where one considers an industry's response to (a change in) rules, a narrow rational choice approach with its exclusive focus on economic incentives may, or may not, facilitate a realistic view. For an analysis on the *meso-level* of company decisions made by respected members of the business community, applying the rationale of control theories and abstracting from risk factors can be adequate but may not always be so. There may be more causes of white-collar crime, for instance, than just economic temptations and gaps in social bonds and social control (Katz 1988, pp. 310–324). A priori disregarding risk factors, however, is most plausibly inadequate not only when trying to understand juvenile delinquency but also in a *micro-level* analysis of individual employees who may face strong group cohesion and deviant corporate subcultures with informal social pressures and rewards to bend or break the law.

The extension of the narrow rational choice conception provided by the inclusion of protective factors and risk factors into the set of potential behavioral determinants can be related to Ostrom's (2005) "Institutional Analysis and Development Framework." Adopting an institutional economics focus on rule-governed social life, Ostrom attaches the label *delta parameters* to the nonmaterial benefits and costs that an individual derives from her/his choices. That is, besides the utility derived from economic incentives (payoffs), emphasis is put on the utility (payoff equivalents) that people gain from external sources in terms of social approval or disapproval (positive and negative external deltas) and internal sources in terms of affirmation of, or conflict with, internalized values and self-image (positive and negative internal deltas). Given the obvious communication problems caused by the lack of a generally acknowledged terminology between the various behavioral science disciplines (here as well as in many other cases), one can but agree with Ostrom's own words, "If every social science discipline or subdiscipline uses a different language for key terms and focuses on different levels of explanation as the 'proper' way to understand behavior and outcomes, one can understand why discourse may resemble a Tower of Babel rather than a cumulative body of knowledge" (Ostrom 2005, p. 11).

Prevention of Rule Breaking

How people respond to rules is one of the core concerns of *Law and Economics* and regulatory theory. First of all, regulators have to understand existing compliance problems as resulting from the joint effects of economic incentives, protective factors, and risk factors that are at work at a given point in time. Regarding prevention and the *management of behavioral risks*, a reconstructing understanding of the status quo is but a starting point, however. To successfully manage behavioral risks and reduce the probability of first-time rule breaking as well as of recidivism, regulators need to identify effective and cost-efficient regulatory strategies. Their success in this endeavors – be they directed at certain target groups within risk-based preventive schemes (secondary prevention) or at individual offenders within personalized reintegration approaches (tertiary prevention) – depends on the ability to forecast behavioral changes that are likely to be brought about by hypothesized regulatory action.

While the protective factor concept is geared toward the analysis of compliance with given rules, one should keep in mind that, depending on the type of the behavioral determinant they are

primarily aimed to impact, four ideal types of regulatory regimes can be distinguished in the first place:

1. *Hierarchical command and control*, such as tight law enforcement, impacts the set of choices that are available to regulatees (e.g., through license withdrawal) as well as their relative economic competitiveness (e.g., through sanctioning).
2. *Incentive-oriented regimes*, such as steering taxes and payments handed out to those who undertake specified actions, increase the relative economic competitiveness of the socially desired behavior without imposing mandatory rules.
3. *Value-oriented regimes*, such as the promotion of corporate social responsibility and professional ethics, are based on persuasion and finally aimed to increase protective factors that shield regulatees from socially undesired behavior.
4. *Human capacity building*, such as counseling and training, is to enhance the actors' abilities to perform socially desired behavior.

The law- and control-based systems that criminology is traditionally concerned with can clearly be associated with the first regulatory type. However, even within regimes based on mandatory rules, regulators can go beyond command and control and target the full set of behavioral determinants to produce compliant behavior. Accordingly, regulatory enforcement can be related to the three basic regulatory strategies known in criminology: (1) *incapacitation and target hardening*, (2) *deterrence*, and (3) *accommodation* (Picciotto 2002). Two definitional settings are required for this relational positioning: First, we need to equate sanctioning and deterrence not only with punitive measures but with the production of incentives that reduce the relative competitiveness of rule breaking (including damage compensations, recall costs, contractual fines, litigation costs, losses of sales, etc.). Second, we need to understand accommodation as not only addressing personal values but including other ways of human capacity building.

Regulatory and criminological scholars from different normative schools attribute different degrees of importance to material and nonmaterial motivations. Some believe in deterrence only; others believe that accommodation and gentle persuasion (i.e., the promotion of protective factors) work to secure compliance (Ayres and Braithwaite 1992). Contrary to this antithetic pairing, evidence from a wide range of fields – from occupational safety (Scholz and Gray 1990) over standards in nuclear plants (Rees 1994) to tax compliance (Braithwaite 2009) – indicates that successful strategies avoid the dysfunctional effects of pure deterrence, such as defiance and reactance, and the negative effects of pure accommodation and leniency, such as a blurring of standards. Successful strategies are able to simultaneously reduce economic temptations and strengthen bonds to social norms by generating *value correspondence* between regulator and regulatees.

Focusing on tertiary prevention and the reduction of recidivism, Braithwaite (2002) provides a related systematization of regulatory strategies and stresses that they need to be combined and applied in a systematic progression. He distinguishes *restorative justice coaching* (persuasion/counseling) from increasing levels of *deterrence* (warning letters → civil penalties → criminal penalties) and finally from different levels of *incapacitation* (license suspension → license revocation). Accounting for the pros and cons of persuasion as opposed to the various degrees of punishment, he advocates that one should attempt to reintegrate offenders into the rule-abiding society by using transparent and graduated responses to misconduct, with escalating/de-escalating measures contingent on the regulatee's degree of bad/good conduct (*responsive regulation*). According to his *enforcement pyramid*, noncompliance should be met with a clear disapproval of the fact and increasingly punitive measures, but regulators should always start softly by using the cooperative measures of persuasion and counseling aimed at reintegrating offenders into the law-abiding community. This is not to be confused with naïve trust. According to the concept of responsive regulation, the harsher the available ultimate sanctions, the more likely compliance will be achieved through persuasion. The recommendation to regulators would therefore be to "speak softly, while carrying very big sticks" (Ayres

and Braithwaite 1992, p. 40, paraphrasing a well-known quotation from Theodore Roosevelt).

Technically speaking, the responsive regulation approach moves away from an overly simplistic partial perspective according to which one might ostensibly conclude, for instance, that the more controls and sanctions, the higher the actors' inclination to abide by the rules. Instead, responsive regulation represents a holistic approach geared to generate and exploit positive interactions between nonmaterial and material behavioral determinants and, above all, to avoid so-called ironic or perverse effects of sanctions and control, including *self-fulfilling prophecies of distrust* (Luhmann 2000). It can thus be related to the concern with motivation in law and behavioral science. When regulators introduce new measures to steer behavior, be they new rules or new enforcement approaches, a ceteris paribus assumption with an exclusive focus on economic incentives is hardly adequate to predict the behavioral impacts. This is due to the fact that the following interactions may occur (cf. Frey 1997):

- Even primarily incentive-oriented regulatory action – be it based on monetary rewards for desired or on sanctions against undesired behavior – may have a positive impact on internal protective factors as well. This will be the case, for instance, if incentives/disincentives and accompanying controls are being appreciated by the individual as a clear affirmation of what is right and wrong in a society. Such desirable synergies of regulatory action have been termed *motivation crowding in*.
- Regulatory change such as tightening of controls and/or an increase of sanctions may reduce the economic temptation to break rules but impair internal protective factors. In such a case, the positive impact on the behavior in question will be lessened, and the desired motivational change could even be reduced to (or below) zero. Such dysfunctional dynamic shifts have been termed *motivation crowding out*.
- Regulatory change aimed at reducing economic temptations might even generate risk factors. When regulatees experience a new regulation as an illegitimate interference with their basic values and/or freedom of action, they may derive a genuine nonmaterial internal benefit from standing up against the new rule. This reactance makes them even accept an economic disadvantage of rule-breaking (a *negative ethical premium*) in exchange for the satisfaction of asserting their autonomy in the face of what they see as an outrageous interference.

Frey and Jegen (2001, p. 590) describe motivation crowding out as "one of the most important behavioral anomalies in economics, as it suggests the opposite of the most fundamental economic 'law,' that raising monetary incentives increases supply." One would certainly have to integrate reactance into this statement. The *behavioral anomaly* obviously dissolves as soon as one adopts a more realistic and broad utilitarian view instead of maintaining narrow rational choice assumptions. Regulators should realize that regulatory change may cause interconnected impacts and consequently search for strategies that are "smart" in that they avoid dysfunctional effects and, at best, generate synergies. The conception of economic man underlying the change from the famous *get incentives right* to the more adequate *get utilities right* – as subjectively expected by those who are subject to rules – is the key to the understanding of what behavioral economic analysis and the regulatory issue are essentially about (Hirschauer et al. 2012).

Last not least, looking beyond the enforcement of given rules, it should be noted that producing compliant behavior makes only sense if the rule itself makes sense. If a rule is not meaningful in the first place (i.e., if the social costs associated with compliance exceed the social benefits of compliance), then even the most effective production of compliance will not contribute to a positive outcome in the real world. If a rule makes sense, however, adverse social outcomes resulting from malpractice can be conceived as *negative externalities* that are caused by the breaking of rules that had been designed to prevent them in the first place. In this latter case, we can indeed limit our concerns to the question of how society can cost-efficiently make individuals abide by its (presumably social efficient) rules.

Cross-References

- ▶ Criminal Sanctions and Deterrence
- ▶ Externalities
- ▶ Rationality

References

Akerlof G, Kranton R (2010) Identity economics. Princeton University Press, Princeton

Ayres I, Braithwaite J (1992) Responsive regulation: transcending the deregulation debate. Oxford University Press, New York

Braithwaite J (2002) Restorative justice and responsive regulation. Oxford University Press, New York

Braithwaite V (2009) Defiance in taxation and governance: resisting and dismissing authority in a democracy. Edward Elgar, Northampton

Brehm JW (1966) A theory of psychological reactance. Academic, New York

Frey BS (1997) Not just for the money: an economic theory of motivation. Edward Elgar, Cheltenham

Frey BS, Jegen R (2001) Motivation crowding theory: a survey of empirical evidence. J Econ Surv 15:589–611

Garoupa N (2003) Behavioral economic analysis of crime: a critical review. Eur J Law Econ 15:5–15

Gigerenzer G, Selten R (eds) (2001) Bounded rationality: the adaptive toolbox. MIT Press, Cambridge

Gottfredson MR, Hirschi T (1990) A general theory of crime. Stanford University Press, Stanford

Hess H, Scheerer S (2004) Theorie der Kriminalität. In: Oberwittler D, Karstedt S (eds) Kölner Zeitschrift für Soziologie und Sozialpsychologie. Sonderheft 43. Soziologie der Kriminalität, Wiesbaden, pp 69–92

Hirschauer N, Bavorová M, Martino G (2012) An analytical framework for a behavioural analysis of non-compliance in food supply chains. Brit Food J 114:1212–1227

Hirschi T (1969) Causes of delinquency. University of California Press, Berkeley

Katz J (1988) Seductions of crime. Basic Books, New York

Korobkin RB, Ulen TS (2000) Law and behavioral science: removing the rationality assumption from law and economics. Calif Law Rev 88:1051–1144

Lösel F, Bender D (2003) Resilience and protective factors. In: Farrington DP, Coid J (eds) Prevention of adult antisocial behaviour. Cambridge University Press, Cambridge, pp 130–204

Luhmann N (2000) Vertrauen, 4th edn. Lucius & Lucius, Stuttgart

Miron AM, Brehm JW (2006) Reactance theory: 40 years later. Z Sozialpsychol 37:9–18

Ostrom E (2005) Understanding institutional diversity. Princeton University Press, Princeton

Parisi F (2004) Positive, normative and functional schools in law and economics. Eur J Law Econ 18:259–272

Picciotto S (2002) Introduction: reconceptualizing regulation in the era of globalisation. In: Picciotto S, Campbell D (eds) New directions in regulatory theory. Blackwell, Oxford, pp 1–11

Pinstrup-Andersen P (2005) Ethics and economic policy for the food system. Am J Agric Econ 87:1097–1112

Rees JV (1994) Hostages of each other: the transformation of nuclear safety since three Mile Island. Chicago University Press, Chicago

Rubinstein A (1991) Comments on the interpretation of game theory. Econometrica 59:909–924

Scholz JT, Gray WB (1990) OSHA enforcement and workplace injuries: a behavioral approach to risk assessment. J Risk Uncertainty 3:283–305

Simon HA (1957) Models of man: social and rational. Wiley, New York

Tittle CR (2000) Theoretical developments in criminology. Crim Justice 1:51–101

Further Reading

Braithwaite J (2008) Regulatory capitalism: how it works, ideas for making it work better. Edward Elgar, Northampton

Coleman JS (1987) Norms as social capital. In: Radnitzky G, Bernholz P (eds) Economic imperialism. The economic method applied outside the field of economics. Paragon House Publisher, New York, pp 133–155

Englerth M (2010) Der beschränkt rationale Verbrecher: Behavioral Economics in der Kriminologie. LitVerlag, Berlin

Thomas N, Baumert A, Schmitt M (2012) Justice sensitivity as a risk and protective factor in social conflicts. In: Kals E, Maes J (eds) Justice and conflicts: theoretical and empirical contributions. Springer, Berlin/Heidelberg, pp 107–120

Tittle CR (1995) Control balance. Towards a general theory of deviance. Westview Press, Boulder

Tyler TR (1990) Why people obey the law. Princeton University Press, Princeton

Public Choice: The Virginia School

Rosolino A. Candela

Political Theory Project, Department of Political Science, Brown University, Providence, RI, USA

Abstract

Public Choice, broadly defined, is the economic analysis of nonmarket decision-making. My primary focus in this chapter will be on the central importance of the "Virginia School" of

Public Choice, specifically in two aspects. First, Public Choice was central to countering the presumption of market failure among economists, beginning in the 1950s and 1960s. Second, yet less emphasized, is the role in which Public Choice played in analyzing the economics of anarchy, beginning in the 1970s. Although seemingly disconnected, both aspects of the Virginia School are linked by the fundamental question of political economy: Under what institutional conditions can social order emerge unintendedly from the self-interest of individuals? Therefore, in both respects, the Virginia School has been central to the rearticulation of 19th political economy during the twentieth and twenty-first centuries.

Introduction

Public Choice, broadly defined, is the economic analysis of nonmarket decision-making. More specifically, James Buchanan, who cofounded the field of Public Choice with Gordon Tullock, referred to Public Choice as "politics without romance" (Buchanan 1979). Public Choice first emerged as a discipline within the broader tradition of political economy at the Thomas Jefferson Center for Studies in Political Economy at the University of Virginia (UVA), which Buchanan co-founded in 1957 together with G. Warren Nutter. Since then, several other branches of Public Choice have emerged. These include the "Bloomington School" of Vincent and Elinor Ostrom; the "Chicago School" developed by Gary Becker, Sam Peltzman, and George Stigler; and the "Rochester School" of William Riker. To compare and contrast each of these branches of Public Choice, let along provide a comprehensive survey of the entire field, goes beyond the scope of this entry (see Mitchell 1988 and Mueller 1976).

My primary focus in this chapter will be on the central importance of the "Virginia School" of Public Choice, specifically in two aspects. First, more than any other branch of public choice, its explicit motivation "is the rejection of all of the pillars of the Samuelsonian revolution" (Boettke and Marciano 2015: 54), which had begun to emphasize the tendency for markets to be a suboptimal mechanism of resource allocation. By applying the logic of economic decision-making to political settings, Public Choice economists countered the Samuelsonian presumption of market failure by pointing out that governments may also be suboptimal in their attempt to address market failures. Second, yet less emphasized, is the role in which Public Choice played in analyzing the economics of anarchy. Beginning in the 1970s, the Virginia School began to analyze the mechanisms by which institutions can emerge to define and enforce property rights and contractual arrangements without government. This has led to a flowering of theoretical and historical case studies providing evidence of social cooperation without government.

At first glance, these two aspects of the Virginia School seem to be disconnected. However, both the study of government failure and anarchy are flips sides of the same question grounding political economy: Under what *institutional* conditions can social order emerge unintendedly from the self-interest of individuals? Therefore, in both respects, the Virginia School has been central to the rearticulation of 19th political economy during the twentieth and twenty-first centuries. In both the study of government failure and anarchy, the uniqueness of the Virginia School has been its ability to "marry the property-rights, law-and-economics, public-choice, Austrian subjectivist approaches" of economics (Buchanan 2015: 260). For the remainder of this chapter, I will trace out the origins and contributions of the Virginia School to the broader tradition of political economy.

A Brief Overview of the Virginia School

The origins of Public Choice can be traced back to Adam Smith and the classical political economists of the nineteenth century. What all branches of Public Choice, in particular the Virginia School, shared with classical political economy is their comparative analysis of alternative processes of decision-making and their respective results under

alternative institutional arrangements, particularly between market and nonmarket institutional settings. Public Choice, and the Virginia School in particular, bases its analysis of political decision-making on three basic assumptions: (1) rational choice; (2) methodological individualism; and (3) politics as an exchange process.

To say that individuals choose rationally implies simply the following: When an individual faces a set of alternatives, he or she will choose the alternative they expect to give them the greatest satisfaction. In other words, the pursuit of one's rational self-interest implies simply that individuals choose more rather than less of whatever they prefer, or that the individual *strives* to maximize utility. Moreover, Public Choice assumes *behavioral symmetry*, meaning that individuals are utility maximizers in both market and political settings; the only difference between market and political actors is the manner in which utility-maximizing behavior manifests itself. For example, in the marketplace, firm owners strive to maximize monetary profit, but in nonprofit setting, such as politics, there are four groups of decision-makers, each of which are maximizing utility as well. They include voters and interest groups, who want goods and services from elected officials, politicians (i.e., elected officials), who strive to maximize votes and financial support from interest groups and voters, and bureaucrats, who wish to maximize their budgets from elected officials (Simmons 2011: 52).

Unlike in the market place, where individuals express their preference by "voting" with their dollars for different goods sold simultaneously by competing firms, the same individual in politics must vote on a single, bundled policy, which constitutes many different goods, "sold" to them by political officials they've elected. Given the bundled nature of public policy, and its implementation, voters will be rationally ignorant, given the high costs of gathering information about the different "goods" in a policy bundle, which may include defense, health care, import tariffs, etc. Moreover, the bundle nature of public policy also implies that individuals will prefer different "goods" *within the bundle* with different intensities. For example, given the concentrated benefit from implementing tariffs on imported sugar, domestic sugar producers will no doubt express greater preference for such a policy, in the form of votes as well as campaign contributions to elected officials, unlike the common voter, who will tend to remain relatively ignorant of such a policy. Therefore, the logic of political decision-making will be for vote-maximizing politicians to concentrate benefits on well-organized and well-informed interest groups, which include budget-maximizing bureaucracies, and disperse costs on unorganized and ill-informed voters. Such a political outcome is rational, once we've traced such an outcome back to the choices made by utility-maximizing individuals under a nonmarket setting, where individuals do not bear the full costs and benefits of their decision-making. This brings us to methodological individualism.

By rejecting an organic holistic vision of the state, Public Choice employs methodological individualism, meaning that the outcomes of collective action are traced back to the choices and interactions of rational, utility maximizers. This does not imply that individual preferences can be aggregated into a social welfare function, which the state maximizes on behalf of the electorate. Rather, collective action under a methodologically individualist view will take place if two or more individuals find it mutually beneficial to accomplish certain common purposes jointly with others, rather than separately through bilateral market interactions, an example being the drainage of a mosquito-infested swamp (Buchanan 1964: 219–220). This example reveals that, in the Virginia School, social phenomena are not simply the result of individuals passively responding to constraints.

Unlike in a constrained maximization problem, defined by fixed constraints, Public Choice theorists of the Virginia School direct their analytic attention to *choice among constraints*, where politics is the artifact of an exchange process, whereby individuals strive to agree to mutually beneficial rules that constrain their behavior in a Pareto-optimal manner. The Virginia School political economists therefore take a *constitutional perspective*, which focuses on analyzing "the rules of game" that governs political

interaction. Relevant political choices, according to the study Constitutional Political Economy, are not choices among alternative distributions or allocations of resources within a set of political rules. Rather, it postulates that relevant political choices are among an alternative set of political rules that generate different patterns of distribution and allocation of resources. Constitutional Political Economy is tied to Public Choice, broadly speaking, in two respects. Assuming rational self-interest on the part of political officials, alternative decision-making rules, such as majority rule or unanimity, will generate different policy outcomes within those sets of rules. For example, requiring more than majority rule for road repair will tend to reduce public expenditure on roads, since more people will have to be included in the political exchange. Increasing the number of individuals as a decision-making rule will consequently increase the cost to the beneficiaries of road repair, and, therefore, low costs will be spilled over and dispersed to parties not agreeing with the exchange. Second, political reform cannot occur by changing the "players" within the political game, but only by changing the rules by which the game is played. It is no accident that the study of Public Choice, and the Virginia School, in particular, emerged when the rule level of analysis began to recede to the background of political and economic analysis.

The birth of the Virginia School cannot be understood outside the historical context it inherited. The Samuelsonian revolution changed the tacit presupposition of economists with regard to the government's role in a market economy, from a "laissez-faire presumption" to a "market failure presumption" (Boettke and Lesson 2015: xiii–xviii). This presumption can best be understood as *epistemological*, rather than political or ideological. Under the laissez-faire presumption, the analytical description of the market economy was complemented by a default presumption to limit government's role in the marketplace, generally speaking, to the protection of private property and contract enforcement. However, under the market failure presumption, economists assume that the presence of macroeconomic instability, monopoly power, externalities, and public goods will merit, by default, government correction of the market's allocative process. The distinction here is not simply a presumption of "more government" or "less government," but fundamentally a presumption of what knowledge government officials possess to correct market failures. Buchanan best states this distinction as follows:

> The classical (Smithean) argument for control (or depoliticization) and the welfare economists' argument for control (for politicization) are on all fours ***only if we presume the existence of the same underlying evaluative standard in the two cases***. To suggest, with the welfare economists, that market failure supports politicization, there must be not only departures from the necessary conditions for efficiency, but also some presumption that political action is informed by a knowledge of what the allocatively efficient solution is, quite apart from the operation of politics itself. By contrast, to suggest, with Adam Smith, that regulatory failure supports market liberalization does not require any presumptive knowledge about what particular outcome is likely to produce maximal value. ***There is a categorical epistemological difference between the two comparative exercises, a difference that many modern economists still do not understand***. (emphasis added, Buchanan 1996 [2001]: 292)

Central to this "market failure presumption" and the Samuelsonian paradigm was the presumed existence of an objective social welfare function, which government officials would maximize by delivering an optimal amount of public goods, eliminating monopoly power and externalities, and establishing macroeconomic stability in the economy through aggregate demand management. From this context, Buchanan, Tullock, and the Virginia School established Public Choice, based on the notion of *homo economicus* and politics as an exchange process.

Public Choice and Government Failure

Beginning in the late 1940s and early 1950s, Public Choice began to counter the presumption of market failure by challenging the existence of an objective social welfare function. In his 1949 article, "The Pure Theory of Government Finance: A Suggested Approach," Buchanan contrasted between an "organismic" theory of

the state and an "individualistic" theory of the state. The basis of this contrast would not only define Buchanan's intellectual trajectory (Wagner 2017), but also that of Public Choice as well. According to the organismic theory, the state is considered as a single decision-making unit, or "a fiscal brain" as Buchanan puts it, which acts for society as a whole by seeking to maximize "general welfare" or "social utility" (1949: 497). In the individualistic theory of the state, the state embodies no ends other than those of individuals in society, who find it in their self-interest to pursue a certain portion of their wants collectively. The individualistic theory typifies the nature of interaction between individuals and the state as an "exchange" of government services paid out of the economic resources of individuals. The basis of the individualistic theory of the state, and the Public Choice perspective, is an extension of two related aspects from economic theory: (1) based on the incentive structure of an institutional setting, individuals will act according to their own self-interest (i.e., *homo economicus*); and (2) individuals pursue their self-interest by engaging in mutually beneficial exchange. By applying the behavioral postulate of *homo economicus* and the principle of mutually beneficial exchange to political decision-making, Public Choice challenges the main pillars of the Samuelsonian paradigm by arguing threefold: (1) there is no objective welfare function that a benevolent despot maximizes; (2) even if such a social welfare function existed, only individuals choose, not "society" or "the state"; and (3) individuals acting in a market setting or a political setting will pursue their self-interest based on their subjective assessment of costs and benefits (Boettke 2012: 249).

In countering the paradigm of Samuelsonian economics, Public Choice emerged "as a 'theory of government failure' that offsets 'the theory of market failure' that emerged from theoretical welfare economics" (Buchanan 1983 [2000]: 113). Under market failure theory, perfect competition became the normative benchmark from which the inefficiencies of real-world market outcomes were compared. Any deviations from perfectly competitive equilibrium, such as the presence of public goods, externalities, monopoly power, or asymmetric information, imply government intervention as a public policy prescription to correct for such market failures. Consistent with the evaluative standard of perfect competition, political actors are presumed to possess perfect information to address market failures. However, this was a categorical epistemological difference in understanding market processes as well as political processes, not only for the Virginia School, but also for economists prior to the mid-twentieth century. Ronald Coase, a pioneer in the field of law and economics, and one of the early faculty members of the Thomas Jefferson Center at the UVA, best articulated this presumption of government failure:

> This "novel theory" (novel with Adam Smith) is, of course, that the allocation of resources should be determined by the forces of the market rather than as a result of government decisions. Quite apart from the malallocations which are the result of political pressures, an administrative agency which attempts to perform the function normally carried out by the pricing mechanism operates under two handicaps. First of all, it lacks the precise monetary measure of benefit and cost provided by the market. Second, it cannot, by the nature of things, be in possession of all the relevant information possessed by the managers of every business. . . to say nothing of the preferences of consumers for the various goods and services (Coase 1959: 18).

Rather than view economics in terms of states of equilibrium, the Virginia School has always regarded economics as a science of exchange and the institutions within exchange takes place. By anchoring political choices in price theory and methodological individualism, Buchanan, Tullock, and the Virginia School as a whole were able to turn the presumption of market failure on its head, namely, by marrying the best insights of the property-rights economists, law-and-economics, and Austrian subjectivism. For example, following the subjectivism of Austrian economists Carl Menger, Ludwig von Mises, and F.A. Hayek, Buchanan saw the notion of cost as inherently tied to the act of individual choice and understood a subjective assessment of trade-offs by individuals if it would have any meaning in a theory of decision-making (Buchanan 1969 [1999]; see also Buchanan and

Thirby 1981). Although a seemingly elementary insight into price theory, its public policy implications for market failure theorists are devastating. For example, in the face of externalities, such as pollution, market failure economists would propose, a lá Pigou, the use of taxes so that individual polluters would bring the full social costs of polluting into their decision-making. However, returning to the earlier quote by Coase, this presumes that political actors have both the incentives as well as the knowledge to address these market failures in accordance with ideal conditions of general competitive equilibrium. However, under such conditions, such a policy is either possible and redundant, or impossible to set because the institutional conditions presupposed for their establishment either eliminate their necessity or preclude the availability of the knowledge necessary to calculate the correct tax to levy.

In the field of public finance, Buchanan challenged the Samuelsonian orthodoxy in fiscal policy by questioning the incentives that political actors face in balancing budgets, namely by demonstrating the intergenerational transfer of government debt (Buchanan 1958 [1999]). In his 1948 edition of his seminal text *Economics*, Paul Samuelson argued that the "interest on internal debt is paid by Americans to Americans; there is no *direct* loss of goods and services" (Samuelson 1948: 427). In other words, government debt is not a burden because "we owe it to ourselves." By disregarding the notion of a social welfare function, let alone its maximization, what Public Choice shows is that political actors will only assess costs as subjective trade-offs in the maximization of their own utility functions. Understood this way, politicians will only have the incentive to run ever increasing budget deficits, not to smooth government spending over the business cycle. This is because the economic logic that political officials face in their decision-making is to concentrate benefits among small and well-organized interest groups in their constituency, and disperse costs among the ill-informed masses of the population. Alternatively speaking, if we regard fiscal responsibility as a public good, political actors face a free-rider problem in the elimination of budget deficits, namely because of the cost of eliminating a spending program will be concentrated on a specific interest group, while the benefits of eliminating fiscal irresponsibility will only be dispersed throughout the population. Given the absence of property rights in a political institutional setting, it only makes political sense that a "tragedy of the fiscal commons" will prevail (see Wagner 2012). Thus, it is the interest of political actors to shift the debt burden to future generations. This fiscal insight, however, has broader public policy implications. Given the logic of political decision-making, government attempts to address alleged market failures, not only in the form of macroeconomic instability, but also in the presence of monopoly, externalities, and public goods, are prone to rent-seeking and regulatory capture for the private benefit of special interest groups, which also include government bureaucracies who benefit from the expansion of the scale and scope of government. As we show in the next section, this general concern regarding the ever increasing size and scope of government due to government failure naturally evolved into a presumption of anarchy in the Virginia School (see Powell and Stringham 2009).

Public Choice and Anarchy

Beginning in the 1970s, Public Choice economists began to analyze the capability of individuals to engage in peaceful social cooperation without government. Although seemingly new and radical, this was an inquiry in political economy dating back to the nineteenth century. Carl Menger proposed this question in his 1883 book, *Investigations into the Method of the Social Sciences*, where he asked the following: "*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*?" (emphasis original, 1883 [1985]: 146). Due to the civil unrest that emerged during the Vietnam War and the Civil Rights movement, James Buchanan, Gordon Tullock, and Winston Bush undertook a

radical re-examination of alternative institutional arrangements for governing society at Virginia Polytechnic Institute and State University, resulting in the publication of *Explorations in the Theory of Anarchy* (1972) and *Further Explorations in the Theory of Anarchy* (1974). As Winston Bush stated, "[i]t is not surprising that 'anarchy' and 'anarchism' have reemerged as topics for discussion in the 1960s and the 1970s, as tentacles of government progressively invade private lives and as the alleged objectives of such invasions receded yet further from attainment" (1972 [2005]: 10). However, the early investigations into the prospects of anarchy were generally more pessimistic that they would later become among scholars of the Virginia School. Given the historical context in which they were writing, Buchanan, Bush, and the other contributors regarded anarchism with skepticism. "The anarchists of the 1960s," according to Buchanan, "were enemies of order, rather than proponents of any alternative organizational structure" (2005: 192). Anarchy, as it was understood in *Explorations in the Theory of Anarchy* and *Further Explorations in the Theory of Anarchy*, referred to a state in society characterized by the absence of law, leading to banditry, violence, and social disorder. The common assumption held by these scholars was an identification of government with governance itself.

Since the 1970s, however, Public Choice scholars have extended the original work on government failure to show in fact that anarchy operates more effectively than previously believed. Applying the same logic developed by the earlier generation of Public Choice economists to develop the theory of government failure, later these scholars simply pushed the theory to its logical conclusion, arguing that governments would not always improve upon conditions of anarchy. The work of Public Choice economist Bruce Benson illustrates this well. For example, Benson has shown historically how international commercial law, known as the Law Merchant, emerged spontaneously in Medieval Europe to facilitate international trade, and operated without government enforcement. Disputes between merchants were settled in private merchant courts, and these court decisions were accepted by winners and losers because they were backed by the discipline of repeated dealings as well as the threat of ostracism by the merchant community at large (Benson 1989). In *The Enterprise of Law* (1990), Benson also illustrates how the centralization of law enforcement by government in England later crowded out private mechanisms of law enforcement for the purpose of using the legal system to collect revenue. The failure of anarchy to provide a private market for law enforcement can be attributed then to government failure.

Another important publication in the transition from a pessimistic presumption of anarchy to a more optimistic presumption of anarchy was *Anarchy, State and Public Choice* (2005), which was ostensibly written in response to the contributions written in *Explorations in the Theory of Anarchy* and *Further Explorations in the Theory of Anarchy*. From *Anarchy, State and Public Choice*, a burgeoning literature has further developed the economic analysis of anarchy. Whereas the earlier generation of Virginia School economists saw the gains from trade and innovation being limited by the extent to which governments secured property rights and enforced contracts, the empirical challenge of this newer generation was to show that the existence of such potential gains from trade and innovation presented an entrepreneurial profit opportunity for the endogenous formation of norms and rules.

The economic analysis of anarchy as it evolved in the 1990s and 2000s, like the economic analysis of government failure in the 1950s and 1960s, cannot be understood outside the historical context in which such scholarship emerged. Specific events, such as the collapse of communism in Eastern and Central Europe, ethnic and religious fractionalization in the Balkans and the Middle East, and the exportation of liberal democracy to failed and weak states in the developing world, have demonstrated that governance requires the endogenous formation of rules, rather than their imposition by governments exogenously (Coyne 2008). Moreover, the economic analysis of anarchy, just like the theory of government failure that preceded it, emerged to counter the policy

implications of utilizing perfectly competitive equilibrium as a normative benchmark of analyzing markets in the developing world. The standard neoclassical model populated by fully informed and homogenous agents, in which property rights are well-defined and well-enforced, is unreliable to understanding the situation of failed and weak states in the developing world for two reasons. First, governments in the developing world provide poor enforcement of property rights, or are outright predatory. Second, precisely because the situation in failed and weak states is one in which individuals are heterogeneous, have imperfect information, and exhibit high discount rates, collective action problems that may exist under anarchy may prove to be even worse under a dysfunctional government. For example, the Virginia School scholars have looked at Somalia as a case study, analyzing development economic indicators before and after the collapse of the Barre regime in 1991. While it may be the case that Somalia under anarchy remains one of the poorest parts of the world, it does not automatically follow that the re-establishment of government would be an ideal solution for the provision of governance. In Somalia, data on standard indicators of economic development suggests that statelessness has improved Somali development substantially in terms of lower rates of infant mortality, higher life expectancy, and lower percentage of individuals living on less than one dollar per day (See Leeson 2007: 697). Scholars of anarchy in the Virginia School, like their predecessors analyzing government failure, have attempted to overturn the existence of a "Nirvana Fallacy" (Demsetz 1969) by comparing imperfect and real institutional alternatives in history between the existence of anarchy and the state. Such analysis can be traced back not only to the roots of Public Choice broadly defined as the economic analysis of nonmarket decision-making, but also more specifically as the study of the economic role of the state without romance. In both respects, Public Choice, particularly the Virginia School, has progressively brought the central inquiry of political economy back to the foreground of economics in the twentieth and twenty-first centuries.

Conclusion

Given the breadth and development of Public Choice since the 1950s, the author has focused on the intellectual origins and evolution of the Virginia School of Public Choice. The author has highlighted the uniqueness of the Virginia School in developing a theory of government failure to counter the presumption of market failure in economics. Moreover, in developing this presumption of government failure, the author has highlighted that the economic analysis of anarchy has followed logically from the initial skepticism among Public Choice economists to use public policy measures to address market failures. Although seemingly separate enterprises, they are both rooted in comparative institutional analysis and an understanding of the price mechanism as offering the possible approach to understanding the emergence of peaceful social cooperation without government command.

Cross-References

- ▶ Anarchy
- ▶ Constitutional Political Economy
- ▶ Gordon Tullock: A Maverick Scholar of Law and Economics
- ▶ Government Failure
- ▶ Market Failure: Analysis
- ▶ Market Failure: History
- ▶ Political Economy
- ▶ Public Interest

References

Benson BL (1989) The spontaneous evolution of Commerical law. South Econ J 55(3):644–661

Benson BL (1990) The Enterprise of law: justice without the state. Pacific Research Institute for Public Policy, San Francisco

Boettke PJ (2012) Living economics: yesterday, today, and tomorrow. The Independent Institute, Oakland

Boettke PJ, Lesson PT (2015) Introduction. In: Boettke PJ, Leeson PT (eds) The international library of critical writings in economics 304: the economic role of the state. Edward Elgar, Northampton

Boettke PJ, Marciano A (2015) The past, present, and future of Virginia political economy. Public Choice 163(1–2):53–65
Buchanan JM (1949) The pure theory of government finance: a suggested approach. J Polit Econ 57(6):496–505
Buchanan JM 1958 (1999) The collected works of James M. Buchanan Vol. 2, Public principles of public debt: a defense and restatement. Liberty Fund, Indianapolis
Buchanan JM (1964) What should economists do? South Econ J 30(3):213–222
Buchanan JM 1969 (1999). The collected works of James M. Buchanan Vol. 6, Cost and choice: an inquiry in economic theory. Liberty Fund, Indianapolis
Buchanan JM 1979 (1999) Politics without romance: a sketch of positive public choice theory and its normative implications. In: The collected works of James M. Buchanan Vol. 1, The logical foundations of constitutional liberty. Liberty Fund, Indianapolis
Buchanan JM 1983 (2000) The achievement and the limits of public choice in diagnosing government failure and in offering bases for constructive reform. In: The collected works of James M. Buchanan volume 13: politics as public choice. Liberty Fund, Indianapolis
Buchanan J 1996 (2001) Adam Smith as inspiration. In: The collected works of James M. Buchanan volume 19: ideas, persons, and events. Liberty Fund, Indianapolis
Buchanan JM (2005) Reflections after three decades. In: Stringham E (ed) Anarchy, state and public choice. Edward Elgar, Cheltenham
Buchanan JM (2015) Notes on Hayek – Miami, 15 February 1979. Rev Austrian Econ 28(3):257–260
Buchanan JM, Thirby GF (1981) L.S.E. Essays on cost. New York University Press, New York
Bush W 1972 (2005) Individual welfare in anarchy. In: Stringham E (ed) Anarchy, state and public choice. Edward Elgar, Cheltenham
Coase RH (1959) The Federal Communications Commission. J Law Econ 2:1–40
Coyne CJ (2008) After war: the political economy of exporting democracy. Stanford University Press, Stanford
Demsetz H (1969) Information and efficiency: another viewpoint. J Law Econ 12(1):1–22
Leeson PT (2007) Better off stateless: Somalia before and after government collapse. J Comp Econ 35(4):689–710
Menger C 1883 (1985) Investigations into the method of the social sciences. New York University Press, New York
Mitchell WC (1988) Virginia, Rochester, and Bloomington: twenty-five years of public choice and political science. Public Choice 56(2):101–119
Mueller DC (1976) Public choice: a survey. J Econ Lit 14(2):395–433
Powell B, Stringham EP (2009) Public choice and the economic analysis of anarchy: a survey. Public Choice 140(3–4):503–538
Samuelson PA (1948) Economics. McGraw-Hill, New York
Simmons RT (2011) Beyond politics: the roots of government failure. The Independent Institute, Oakland
Tullock G (ed) (1972) Explorations in the theory of anarchy. Center for the Study of Public Choice, Blacksburg
Tullock G (ed) (1974) Further explorations in the theory of anarchy. University Publications, Blacksburg
Wagner RE (2012) Deficits, debt, and democracy: wrestling with tragedy on the fiscal commons. Edward Elgar, Cheltenham
Wagner RE (2017) James M. Buchanan and liberal political economy: a rational reconstruction. Lexington, Lanham

Public Enforcement

Iljoong Kim
Department of Economics, SungKyunKwan University (SKKU), Jongno-Gu, Seoul, South Korea

Abstract

This essay starts with discussions regarding what public enforcement is and why it is necessary. We explain economic rationales under which public enforcement becomes a superior sanctioning mode, in controlling many undesirable acts, to a wide variety of non-public sanctioning counterparts. Nonetheless, given that a large portion of the literature considers the high-cost aspect of public enforcement, the essay emphasizes the importance of lowering administrative costs and overcoming bureaucracy. From a similar perspective, we also examine the combination of public and non-public enforcement as well as the joint use of different modes of public enforcement.

Synonyms

Law enforcement by government

Definition

Public enforcement (PE) is a sanctioning mode involving a wide variety of government people such as police, prosecutors, and regulators.

Although society can rely on a range of nonpublic sanctioning modes to control undesirable acts, certain situations occur where economic rationales exist for PE. The social harm from the act, the probability of detection, and the severity of the sanction are critical for its optimality. It is also important to lower administrative costs and to overcome bureaucracy for successful PE.

Introduction

Public enforcement is a sanctioning mode sometimes with the use of physical force, involving a wide variety of people in the government sector such as police, prosecutors, and various regulators. In fact, the phenomenon of public enforcement playing such a predominant role is very recent in human history terms.

Believing that an efficient sanctioning mode varies across different situations, a number of scholars have examined the issue of enforcement for some time under themes such as "system of social control," "structure of enforcement," "modalities of regulation," or "methods of public control" (Ellickson 1973; Shavell 1993; Posner 2011). In particular, since Becker (1968), scholars have paid special attention to public enforcement and have produced numerous articles.

Why Public Enforcement and How Is It Undertaken?

Society relies on a range of nonpublic sanctioning modes to control many undesirable acts. A representative example is "self-help" such as reputation, self-protection, and purchase of insurance (Ehrlich and Becker 1972). Another example is civil ligations associated with torts, contracts, and other private-law doctrines, sometimes called "judicial regulations" (Posner 2011). Society has also developed remedial methods to support these nonpublic sanctioning modes, such as the property rule, the liability rule, or inalienability (Calabresi and Melamed 1972). However, certain situations occur where these private sanctioning modes do not work or become very inefficient in coping with undesirable actors.

Since Becker and Stigler (1974), Shavell (1984a), and others, it has been well established that at least four economic rationales exist for public enforcement. Public enforcement becomes a superior mode as government is better equipped in gathering information about the production of harm. Also, upon being caught, injurers may have insufficient assets to compensate the victim (i.e., the judgment-proof problem). It is also superior when the potential injurer seldom faces the threat of civil suits and can thus escape liability. Finally, public enforcement should not incur too high administrative costs to secure its superiority.

Given these situations, based on Polinsky and Shavell (2000) and others, let us consider briefly how public enforcers could set the severity of the sanction ($\boldsymbol{s}$) at an optimal level. A risk-neutral injurer (**A**) obtains a benefit ($\boldsymbol{b}$) from an act which can incur social harm ($\boldsymbol{h}$), with a probability ($\boldsymbol{q}$). If $\boldsymbol{h}$ occurs, **A** is caught with a probability ($\boldsymbol{p}$). $\boldsymbol{p}$ is less than one, reflecting the reality of imperfect enforcement. $\boldsymbol{h}$ differs across acts and is assumed to be fixed in the short run.

The optimality requires that the expected sanction to **A**, $\boldsymbol{ps}$, should equal the expected social harm, $\boldsymbol{qh}$. Since **A** then performs the act only when $\boldsymbol{b}$ exceeds $\boldsymbol{ps}$, the optimality condition warrants the social efficiency of the act in question (i.e., the condition whereby $\boldsymbol{b}$ should exceed $\boldsymbol{qh}$). In other words, **A**'s private incentive is consistent with that of the social planner. Further, since $\boldsymbol{p}$ is less than one, the optimal level of $\boldsymbol{s}$ is $\boldsymbol{qh}$ divided by $\boldsymbol{p}$. Thus, the inverse of $\boldsymbol{p}$ plays the role of a multiplier, so that **A**'s private decision can be internalized appropriately. Note that the optimal level of $\boldsymbol{s}$ should be reduced if **A** is risk averse. Otherwise, **A**'s act will be over-deterred because the expected net benefit (i.e., $\boldsymbol{b}$ contracted by $\boldsymbol{qh}$) is discounted by risk aversion. Finally, noteworthy is that this model is independent of **A**'s benefit, $\boldsymbol{b}$. In fact, a model heavily focusing on $\boldsymbol{b}$ inevitably forces public enforcers to focus on the information associated with $\boldsymbol{b}$ that can inherently be more easily fabricated by **A**, consistently making the arrow of public enforcers land wide on the target of optimal sanctioning.

Additional comments on this simple model might shed beneficial insights, even for more complicated models that were later developed. First, public enforcers, of the three variables in the model, need to have precise information particularly regarding h across different acts, in order for this system to work (p and q are assumed to be fixed at least in the short run, and information on b is unnecessary). Therefore, the aforementioned superiority of government in gathering information about the harm is a core prerequisite for public enforcement.

Second, although the (low) level of p has often been treated as exogenous, public enforcers must maintain a certain level of p; otherwise, the system will suffer from the ex post equity problem between the detected injurers and those who escaped successfully. Further, a prohibitively high s with a very low p (close to zero) will too easily induce the judgment-proof problem, leading to nullification of public enforcement. More importantly, such a system will hamper the essential (constitutional) principle in enforcement of "marginal deterrence," which postulates that the severity of sanction should increase proportionally to the harm level. This notion, in fact, was emphasized even in earliest writings such as Bentham's *Principles* in 1789. For example, if the death penalty or a one million dollar fine is imposed to a driver who hits a pedestrian and causes a slight wound, the driver would have an incentive to run away or even to kill the wounded pedestrian to escape the penalty. Therefore, public enforcers are required to invest in raising p to an acceptable level, which would consume real resources.

Lowering High Costs of Public Enforcement and Overcoming Bureaucracy

As theoretical inquiries have shown, public enforcement certainly offers an advantage under many circumstances. Nonetheless, as with many other government operations, it incurs generally high cost. If the assumptions about the government as the benevolent and omniscient planner are released, the cost further increases. A large portion of the literature considers this cost aspect of public enforcement.

Firstly, scholars realized that, while public enforcement is needed for a certain number of undesirable acts, for various reasons, such acts cannot be efficiently controlled by it alone. For example, the marginal information cost increases with respect to the degree of enforcement sophistication. In fact, such recognitions motivated scholars to examine the allocation of public enforcement resources, even from the early stage of research. It is perhaps in this context that research on the combination of public and nonpublic enforcement captured their attention. In retrospect, these early studies were meaningful contributions, particularly in the sense that the researchers were developing more hands-on normative theory built on positive observations.

Consider the case of using traffic signals to maintain order on public roads. Enforcement by police in terms of whether drivers violate them can be done reasonably easily. However, police cannot assign a tailor-made speed limit that reflects each driver's value of time, level of urgency, driving skill, etc. The limit is uniformly enforced (e.g., 100 km/h on highways and 30 km/h in downtown areas). If a driver mistakenly hits another car while driving over the limit and causes harm, liability is imposed on top of issuing a ticket (i.e., "per se negligent"). However, the opposite rule (i.e., "compliance defense") does not hold. That is, even if the injurer is driving under the limit, liability is not automatically exempted.

This joint use of public enforcement and liability is used in most jurisdictions. Shavell (1984b) attempted the first theorization to draw its efficiency implication. He highlighted the imperfect nature of public enforcement (e.g., enforcing the speed limit) due to imperfect information particularly about different magnitudes of harm. Also, the insufficient precaution when using liability alone was emphasized. Thus, a popular proposition followed that it is socially cheaper to employ regulation and liability jointly, with a lower regulatory standard than if liability were not used. Later, the role of liability to

ameliorate the high costs of public enforcement, primarily in terms of reducing information costs, was further confirmed. Also, the role of public enforcement to support liability was sometimes underlined instead. Overall, these studies were endeavors in search of efficient public enforcement, given its high costs (Kolstad et al. 1990; De Geest and Dari-Mattiacci 2007; Bhole and Wagner 2008).

Scholars subsequently applied these theoretical studies to broad arenas, ranging from product safety to transportation, hazards, environment, health, etc. La Porta et al. (2006) provided an exemplary inquiry. The authors demonstrated that the proper use of civil liability standards together with public enforcement is necessary for the successful operation of securities markets. Meanwhile, extensions were made to explore the joint use of different modes of public enforcement, such as that of administrative penalties and criminal punishment (Garoupa and Gomez-Pomar 2004; Bowles et al. 2008). Given that such joint use is ubiquitous across countries, this approach is differentiated from the earlier dichotomous-choice models. Furthermore, research on the combined criminal sanctions of fines and imprisonment was launched already in the 1980s (Polinsky and Shavell 1984). A major implication is that the use of imprisonment, the highest-cost sanction, should be confined to cases where the insufficient-asset problem or the need to incapacitate offenders prevails.

Finally, a brief but significant caveat deserves mention. Although overall features of actual public enforcement are roughly consistent with the theories, substantial discrepancy routinely occurs, i.e., "bad equilibrium," primarily due to the public enforcer's incentives which differ from the public-interest mindsets that ordinary citizens would expect them to have (Becker and Stigler 1974). Public enforcement can be characterized as being dominated by "entrepreneurial competition," wherein bureaucrats pursue their subjective goals such as wealth, promotion, and discretion (Breton and Wintrobe 1982). Numerous researchers have already examined public enforcement from this perspective. Nonetheless, in summary, it should be emphasized that the details of public enforcement must be steadily scrutinized through these critical lenses in order to facilitate a change in such bad equilibria. Also, in attempting to further lower high costs, incentive-compatible rules must be implemented that require much more use of "pricing" in the allocation of enforcement resources.

Cross-References

- ▶ Becker, Gary S.
- ▶ Criminal Sanctions and Deterrence
- ▶ Government

References

Becker G (1968) Crime and punishment: an economic approach. J Polit Econ 76:169–217

Becker G, Stigler G (1974) Law enforcement, malfeasance, and compensation of enforcers. J Leg Stud 3:1–18

Bhole B, Wagner J (2008) The joint use of regulation and strict liability with multidimensional care and uncertain conviction. Int Rev Law Econ 28:123–132

Bowles R, Faure M, Garoupa N (2008) The scope of criminal law and criminal sanctions: an economic view and policy implications. J Law Soc 35: 389–416

Breton A, Wintrobe R (1982) The logic of bureaucratic control. Cambridge University Press, New York

Calabresi G, Melamed D (1972) Property rules, liability rules, and inalienability: one view of the Cathedral. Harv Law Rev 85:1089–1128

De Geest G, Dari-Mattiacci G (2007) Soft regulators, tough judges. Supreme Court Econ Rev 15:119–140

Ehrlich I, Becker G (1972) Market insurance, self-insurance and self-protection. J Polit Econ 80:623–648

Ellickson R (1973) Alternatives to zoning: covenants, nuisance rules, and fines as land use controls. Univ Chic Law Rev 40:681–714

Garoupa N, Gomez-Pomar F (2004) Punish once or punish twice: a theory of the use of criminal sanctions in addition to regulatory penalties. Am Law Econ Rev 6:410–433

Kolstad C, Ulen T, Johnson G (1990) *Ex-post* liability for harm vs. *ex-ante* safety regulation: substitutes or complements? Am Econ Rev 80:888–901

La Porta R, Lopez-de-Silanes F, Shleifer A (2006) What works in securities laws? J Financ 61:1–32

Polinsky M, Shavell S (1984) The optimal use of fines and imprisonment. J Public Econ 24:89–99

Polinsky M, Shavell S (2000) The economic theory of public enforcement of law. J Econ Lit 38:45–76

Posner R (2011) Economic analysis of law, 8th edn. Aspen Publishers, New York
Shavell S (1984a) Liability for harm versus regulation of safety. J Leg Stud 13:357–374
Shavell S (1984b) A model of the optimal use of liability and safety regulation. RAND J Econ 15:271–280
Shavell S (1993) The optimal structure of law and enforcement. J Law Econ 36:255–287

Public Goods

Tim Friehe[1] and Florian Baumann[2]
[1]Public Economics Group, Philipps-University Marburg, Marburg, Germany
[2]Center for Advanced Studies in Law and Economics (CASTLE), University of Bonn, Bonn, Germany

Definition

A public good is a good that simultaneously features both nonexcludability and nonrivalry in consumption.

Delimitation and Examples

A pure public good is a good that simultaneously features both nonexcludability and nonrivalry in consumption. Nonexcludability implies that excluding individuals from making use of the good is prohibitively costly, such that all individuals can benefit from the provision of the good. Nonrivalry means that the consumption of the good by one individual does not reduce the consumption possibilities of other individuals. Classic examples include national defense or a lighthouse. In contrast, a private good is characterized by both excludability and rivalry in consumption. For example, if an apple is consumed by one individual no other individual can make use of the apple. In addition, excluding other individuals from the benefits of the apple is associated with only minor costs. There are also goods that show rivalry in consumption but nonexcludability. Such goods are denoted commons and are exemplified by fish in a given area. While it is difficult and costly to exclude individuals from fishing, every fish caught by one person can no longer be caught by any other individual and may endanger the sustainability of the stock of fish. Finally, pure public goods have to be distinguished from club goods. Club goods are characterized by nonrivalry in consumption and the possibility of exclusion. Broadcasting can be regarded as a club good, because watching the program does not influence the option for others to do the same while exclusion is possible through the use of scrambling.

Theory

From an economic point of view, the property that the consumption of the public good by one individual does not infringe on the consumption benefits of other individuals leads to a first question: how is the socially optimal supply of a pure public good determined? Moreover, given nonrivalry and nonexcludability, it is also interesting whether markets or voluntary contributions by individuals can implement the optimal supply of the public good. The basic problem may best be illustrated by a simple example resembling the prisoners' dilemma.

Suppose there are two individuals, A and B, who both hold an endowment that generates private utility equal to one. Each individual decides once whether or not to use the endowment to provide one unit of a public good. Due to the nonrivalry of the public good, both individuals gain from the provision of one unit of the public good by either of the individuals. Suppose that the utility derived per unit of the public good equals 0.75 for each individual. Table 1 summarizes the possible choices and resulting utility levels, where the first (second) entry refers to individual A's (B's) utility level.

Public Goods, Table 1 Actions and payoffs in a simple public-good game

Individual A/Individual B	Don't supply	Supply
Don't supply	1;1	1.75;0.75
Supply	0.75;1.75	1.5;1.5

The outcome that maximizes the sum of utilities is the one in which both individuals supply one unit of the public good (in which case the joint payoff equals three). The costs for each unit supplied are equal to one while the joint payoff is given by 1.5 (0.75 + 0.75). However, if the individuals A and B decide independently, each of them fares better by not supplying one unit of the public good given any decision by the other individual. The supply of one unit of the public good only yields a private benefit of 0.75 for each individual and is not sufficient to balance the utility cost equal to one. As a result, the equilibrium in strictly dominant strategies is that neither A nor B supplies a unit of the public good; consequently both end up with a utility level of one. This outcome is Pareto inferior to the optimal solution.

The example illustrates two ideas regarding public goods: (1) To decide on the socially optimal level of supply of the public good, the costs of supplying an additional unit should be compared to the sum of gains achieved by all individuals through the additional supply (due to the property of nonrivalry). (2) Without cooperative decision-making, the voluntary supply of the public good may not yield the optimal outcome.

More generally, the seminal condition that characterizes the Pareto-efficient supply of a public good was first established by Samuelson (1954, 1955). For any two private goods, a Pareto-efficient allocation is achieved when the marginal rates of substitution between the two goods (i.e., the rates at which individuals are willing to exchange the two goods without a change in utility) are equalized for all individuals and, in addition, are equal to the marginal rate of transformation (i.e., the rate at which the two goods can be exchanged according to the production possibility set). In contrast, the optimal allocation regarding a public good and a private one requires that the sum of the marginal rates of substitution over all individuals is equal to the marginal rate of transformation. As in the example above, increasing the supply of the public good is socially desirable as long as the marginal costs (i.e., the marginal rate of transformation) are weakly less than the sum of the gains enjoyed by all individuals (i.e., the sum of the marginal rates of substitution).

An important contribution examining the theory of the noncooperative provision of public goods is provided by Bergstrom et al. (1986). They confirm that public goods are underprovided and establish that a redistribution of wealth among individuals will change the amount of the public good supplied only if it changes the set of individuals actually contributing to the public good. Furthermore, additional supply of the public good by the state (and financed by taxes) will at least partly (if not fully) be offset by reductions in private contributions.

A cooperative provision of public goods may be attained under some circumstances, especially when the group considered is small. For example, a Lindahl equilibrium constitutes an allocation resulting from bargaining which results in a Pareto-efficient supply of the public good with individual-specific contributions based on each individual's valuation of the public good (Lindahl 1919). However, it must be acknowledged that individuals might have an incentive to misrepresent their preferences. The later literature related to the topic of cooperative public good provision has been greatly influenced by the work on collective action by Olsen (1965).

An alternative to the private provision of public goods is the provision by the state that can collect contributions by force in the form of taxes. In this case, if only distortionary taxes can be used to raise revenue, the additional costs from these distortions should be taken into account when deciding about the optimal level of the public good (see, e.g., Atkinson and Stern 1974). An even more fundamental problem arises in the case with asymmetric information about individuals' preferences. If individual payments are to be based on the stated preferences for the public good, individuals might understate their willingness to pay. Otherwise they might overstate it if they know that the burden of financing the additional supply will be shared by the society as a whole. Clarke (1971) and Groves (1973) present a mechanism which leads to a truthful statement about preferences. The amount of the public good may also be determined by majority voting

instead of by a benevolent government, a scenario for which Bowen (1943) offers an early analysis (Bergstrom and Goodman 1973 provide an empirical investigation).

The theory of public good provision based on standard preferences has been complemented by predictions derived from alternative specifications for individual preferences. For example, in order to better understand the sizable contributions to charities, nonstandard motives such as impure altruism may be considered. In the case of impure altruism, individuals obtain utility by increasing their contribution not only from a higher total level of the public good but also from having a higher individual contribution since these create a so-called warm glow (Andreoni 1989, 1990). The predictions that follow from frameworks enriched in such a way are often more intuitive and more in line with empirical observations, such as the result that government contributions to charity will only incompletely crowd out private donations.

Experiments

For decades, the so-called public-good game has been a workhorse for experimental economists interested in social dilemmas and their potential resolution through institutions (e.g., Ledyard 1995). Classically, individuals in groups of n subjects simultaneously determine the split of their symmetric endowments between a public account and a private one, where the return from the public account per subject is determined by the product of the marginal per capita return ($MPCR$) and the sum of contributions and $0 < MPCR < 1 < MPCR*n$ applies. This is the so-called linear public good game and has a unique Nash equilibrium in which no subject (with standard preferences) contributes. The experimental finding that subjects contribute on average 40–60% of their endowment to the public good in the one-shot version or in the first round of the repeated variant of the game is in striking contrast to the prediction based on standard preferences. However, the initially high contribution rates, which decline over time, are consistent with the idea that there are many conditionally cooperative subjects who are willing to contribute more when they expect others to contribute, although they do not match an increase in contributions by the others in full (e.g., Croson 2007; Fischbacher et al. 2001). Turning to institutions that remedy free-riding incentives, allowing participants to punish peers based on their contributions seems effective (although not necessarily efficient when punishment costs are taken into account). Intuitively, costly peer punishment is on average predominantly chosen by subjects with above-average contributions and addressed at subjects with below-average contributions. Such a mechanism seems to be associated with not only higher overall contributions but in the case of low costs of punishment and/or a long time horizon with real efficiency gains (e.g., Chaudhuri 2011). Interestingly, there is experimental evidence that the possibility of peer punishment attracts subjects when they can migrate between a regime with the possibility to sanction and one without and that sanctioning institutions emerge endogenously (e.g., Gürerk et al. 2006; Kosfeld et al. 2009). Other means to improve contributions to the public good include communication and ostracism (e.g., Chaudhuri 2011; Maier-Rigaud et al. 2010).

Applications

The theory of public goods is relevant to a number of applications in the field of law and economics. First, the private enforcement of norms, for example, by assigning punishment points to peers in experimental settings (as just described) or by ostracizing and shaming offenders in the field may become subject to free-rider incentives, as the deterrence benefit is diffused (e.g., Posner 1997). Relatedly, consider the case of unobservable private precautions against crime. When a household increases investments, this lowers the expected return for the potential offenders, decreasing the involvement of thieves. The benefit accrues to all households equally, whereas the full costs are borne by the household in question. This implies that decentrally determined private precautions

against crime fall short of what is optimal for the group of households (e.g., Shavell 1991). In another realm, it is often argued that the knowledge created by innovative activity is nonrival and may present difficulties regarding excludability. Patents ensure excludability for some time, thereby incentivizing the creation of knowledge while limiting the extent of the use of the knowledge. Accordingly, there is a discussion of the relative merits of different instruments in this context (e.g., Shavell and van Ypersele 2001). In addition, the divergence of private marginal benefits and social marginal benefits that commonly arises with public goods may also induce socially inefficient private incentives to proceed to trial. Court decisions may be socially valuable by creating public information about plaintiffs or precedents, aspects which usually do not feature prominently in the private trade-off (e.g., Hua and Spier 2005; Shavell 1999).

Cross-References

- ▶ Altruism
- ▶ Commons, Anticommons, and Semicommons
- ▶ Efficiency
- ▶ Experimental Law and Economics
- ▶ Externalities
- ▶ Market Failure: Analysis
- ▶ Market Failure: History

References

Andreoni J (1989) Giving with impure altruism: applications to charity and Ricardian equivalence. J Polit Econ 97:1447–1458

Andreoni J (1990) Impure altruism and donations to public goods: a theory of warm-glow giving. Econ J 100:464–477

Atkinson A, Stern N (1974) Pigou, taxation and public goods. Rev Econ Stud 41:119–128

Bergstrom T, Goodman R (1973) Private demands for public goods. Am Econ Rev 63:280–293

Bergstrom T, Blume L, Varian H (1986) On the private provision of public goods. J Public Econ 29:25–49

Bowen H (1943) The interpretation of voting in the allocation of economic resources. Q J Econ 58:27–48

Chaudhuri A (2011) Sustaining cooperation in laboratory public good experiments: a selective survey of the literature. Exp Econ 14:47–83

Clarke E (1971) Multipart pricing of public goods. Public Choice 11:17–33

Croson R (2007) Theories of commitment, altruism and reciprocity: evidence from linear public good games. Econ Inq 45:199–216

Fischbacher U, Gächter S, Fehr E (2001) Are people conditionally cooperative? Evidence from a public good experiment. Econ Lett 71:397–404

Groves T (1973) Incentives in teams. Econometrica 41:617–631

Gürerk Ö, Irlenbusch B, Rockenbach B (2006) The competitive advantage of sanctioning instititutions. Science 312:108–111

Hua X, Spier KE (2005) Information and externalities in sequential litigation. J Inst Theor Econ 161:215–232

Kosfeld M, Okada A, Riedl A (2009) Institution formation in public good games. Am Econ Rev 99: 1335–1355

Ledyard O (1995) Public goods: a survey of experimental research. In: Kagel J, Roth A (eds) Handbook of experimental economics. Princeton, Princeton University Press, pp 111–194

Lindahl E (1919) Die Gerechtigkeit der Besteuerung. Gleerup, Lund

Maier-Rigaud FP, Martinsson P, Staffiero G (2010) Ostracism and the provision of a public good: experimental evidence. J Econ Behav Organ 73:387–395

Olsen M (1965) The logic of collective action. Harvard University Press, Cambridge, MA

Posner R (1997) Social norms and the law: an economic approach. Am Econ Rev Pap Proc 87:365–369

Samuelson PA (1954) The pure theory of public expenditure. Rev Econ Stat 36:387–389

Samuelson PA (1955) Diagrammatic exposition of a theory of public expenditure. Rev Econ Stat 37:350–356

Shavell S (1991) Individual precautions to prevent theft: private versus socially optimal behavior. Int Rev Law Econ 11:123–132

Shavell S (1999) The level of litigation: private versus social optimality of suit and settlement. Int Rev Law Econ 19:99–115

Shavell S, van Ypersele T (2001) Rewards versus intellectual property rights. J Law Econ 44:525–547

Further Reading

Batina RG, Ihori T (2005) Public goods: theories and evidence. Springer, Berlin

Cornes R, Sandler T (1996) The theory of externalities, public goods and clubs. Cambridge University Press, Cambridge

Oakland WH (1989) Theory of public goods. In: Auerbach AJ, Feldstein M (eds) Handbook of public economics, vol 2. North Holland, New York, pp 485–535

Sandmo A (2008) Public goods. In: Durlauf SM, Blume LE (eds) The new Palgrave dictionary of economics, 2nd edn. Palgrave Macmillan, Basingstoke

Public Interest

Michael Hantke-Domas
Centre for Water Law, Policy and Science, University of Dundee, Scotland, United Kingdom
Third Environment Court, Valdivia, Chile

Abstract

The public interest is a concept that can be traced back to the late XVIII century. Ever since the concept has been used to refer to a goal to be obtained by actions of governments and public officials alike. Such view has been strongly contested by important economic schools such as Public Choice and the Chicago School of Regulation. Based on neoclassical economics, they contend the altruism needed to deliver in the public interest, as public officials pursue their own interest and regulations are captured to benefit regulated industries but recognizes the existence of equilibria that might balance the interests at stake (like Pareto or Kaldor-Hicks criteria). In contrast, the law is more optimistic when using public interest either to proceduralize collective interest or for courts adjudication. Even more, public interest serves as justification for government intervention in economic affairs. In the last decades, the promotion of integrity and prevention of corruption have tackled self-interest of public officials, leaving space for the public interest to be pursued by governments and public officials.

Definition

Public interest is the motivation of public officials in exercising their functions to achieve the common good of the collective, by improving democracy and social and economic welfare.

Introduction

Abraham Lincoln in his Gettysburg address (1863) coined a famous phrase defining democracy as a "government of the people, by the people, for the people." The idea that governments – or States, in continental European tradition – must serve the people is an expectation common to many people in modern history.

Modern States are inspired in the principles of the French Revolution that changed the Old Regime to a new social organization based on the values of liberty, equality, and fraternity. The State as a moral entity is – or should be – the guarantor of such values. The State has grown to become one of the most important institutions in many aspects of societies. Indeed, meanwhile in the nineteenth century, government expenditure amounted to 10% of GDP, and by 1996 it amounted to 45% in developed OECD countries (see Middleton 1996; Tanzi and Schuknecht 2000).

The "common good" – or the general welfare – is an important ideal for governments, as they must act in pursuance of public interest. However, what is the substance or content of such ideal?

John Rawls, in his book *A Theory of Justice* (1971), does not treat public interest as a common good but rather delineates a corresponding ethical ideal that is "justice as fairness" (Weisbrod and Benjamin 1978).

> "For us the primary subject of justice is the basic structure of society, or more exactly, the way in which the major social institutions [political constitution and the principal economic and social arrangements] distribute fundamental rights and duties and determine the division of advantages from social cooperation. . .the major institutions [for example, the legal protection of freedom of thought and liberty of conscience, competitive markets, private property in the means of production, and the monogamous family] define men's rights and duties and influence their life prospects, what they can expect to be and how well they can hope to do." (p. 7)

Public interest or justice as fairness hinges upon an assessment made in the "initial situation" where an interpretation is made of the moment and the problem of choice it poses, and an agreement is entered into on a set of principles (Rawls 1971). These principles illuminate what public interest is as they serve as guidance for the initial

position, in the sense that (a) each person has rights equal "to the most extensive basic liberty compatible with a similar liberty of others" (p. 60) and "social and economic inequalities are to be arranged so that they both (a) reasonably expect to be to everyone's advantage, and (b) attached to positions and offices open to all" (p. 60). Restating Rawls' proposal, public interest or justice as fairness might refer to equal rights and fair access.

Economists have different takes, which range from equating public interest to general welfare to those that see no benefit to anyone except public officials. Lawyers feel more comfortable with the concept, and the concept of public interest is widely used with less skepticism.

Not surprisingly the concept of public interest has attracted criticism, as some think that it is too broad to mean something specific (Lewis 2006). Notwithstanding, the concept remains important, probably because it is an analytic tool or a heuristic device present in normative public administration theory, economic regulation, administrative law, and judicial procedures. It is also important in governmental and professional standards of practice (Lewis 2006). Furthermore, it is of relevance for law, as interests are objects of protection by constitutions and laws, and adjudicated by the judiciary (Schmidt-Assmann 2003). Public interest is also important in economics as market failures – and the role of the State to correct them – find in public interest a rationale for economic regulation.

Despite the controversy, what remains the same is the existence of a space where citizens can get together to, for example, entrust their protection either from foreign aggression or for getting medical assistance.

As public interest can be interpreted from a variety of disciplines, I will devote the rest of the entry to reviewing the discussions from the standpoints of economics and law.

Public Interest and Economics

The concept of public interest holds the attention of economists in at least two different areas. First, it is a criterion used in collective choice, an area of economics that cares about the relationships between the preferences of individuals and the choices made by the government (Brown and Jackson 1998). Second, it is used for State regulation of economic activities when markets fail.

In liberal democracies, citizens select their representatives (e.g., members of parliaments or presidents) to make decisions and choices on their behalf. For economists, the aggregation of all individual preferences produces multiple combinations that are efficient to adopt by governments. However, this aggregation does not specify how societies formulate and express collective value judgments (Brown and Jackson 1998).

This is the point where disagreements start to appear, as ethical issues arise regarding collective choice rules (is the rule ethically acceptable?). Economists have argued over the idea of the rational, disinterested, and benevolent public official, which altruistically takes office for the good of the whole community.

Public choice scholars blow off the foundations of the commonly accepted idea that government officials pursue the interests of the whole community. They sustain that human beings act on a rational, self-interested, and utility maximizing basis (Mueller 1989) meaning that government officials act pursuing their own interest rather than promoting the general welfare.

Anthony Downs (1957), in *An Economic Theory of Democracy*, sustained that human beings are selfish. This is simply the application of a tenet of neoclassical economic theory of consumers. As individuals maximize their own utility, Downs could not fathom why government officials should depart from such truism to be altruists devoting their work to enhance the general welfare. If everyone is selfish, then government officials should be the same. Downs also studied political parties and concluded that they were mere vehicles to foster private interests (Lewin 1991).

In the same line as Downs, Buchanan and Tullock in *The Calculus of Consent* (1962) sustained that politics is an exchange between parties that aim to maximize individual interest. Common good was discarded as naive, since

public interest could only mean the satisfaction of many individual desires, facilitated by political activity (Lewin 1991).

Later, Gordon Tullock elaborated on government officials and concluded that an expression of their self-interest was the expansion of public sector irrespective of costs. According to Tullock, bureaucrats wanted to enhance their position either by increasing their own salaries, esteem, or influence. Bureaucrats disregarded the intentions of legislature – similar to public interest – as they wanted to foster their own private interests (Lewin 1991).

Public interest, viewed from public choice theory, is simply the aggregation of individual interests. Paradoxically, it is impossible, under mild conditions, to aggregate individual interests as Kenneth Arrow (1951) demonstrated in his *general possibility theorem*. In short, it is impossible to create a social order of preferences based on individual interests that meet certain criteria: unrestricted domain, non-dictatorship, Pareto efficiency, and independence from irrelevant alternatives (Sola 2004).

Leif Lewin (1991) adds that such impossibility is reinforced by the prisoner's dilemma, which shows that self-centered individuals drove by their own individual preferences usually end up worse off not better off from a collective point of view.

Even if it were possible to aggregate individual preferences, economics only presents a narrow frame of reference, as it does not include in its analysis different social and political values inherent to our democracies, which usually are part of the legal basis for the existence of the State, represented in their constitutions (Feintuck 2010). One clear example is the environment, as it has legal recognition providing for regulation that stretches beyond economic analysis, for example, the precautionary principle that allows governments to adopt decisions despite the existence of incomplete scientific data to warrant them, in order to protect the environment (Feintuck 2010).

Economic regulation is connected to collective choice as both are motivated by public and/or private interests. Government decisions are present in the regulatory process, from ascertaining the need for State intervention to its implementation. Thus, the collective choice discussion is part of economic regulation. The assessment of whether a market fails or not and the decisions to correct externalities, to complete information, to promote competition, to solve principal-agent problems, and to provide public goods are all governed by collective choice.

Economic regulation studies have shown that its implementation might not satisfy general welfare, as regulation in itself might fail (for reviews, see Hägg 1997; Hantke-Domas 2003). The Chicago Theory of Regulation, originated in the works of George Stigler (1971) and further developed by Sam Peltzman (1976) and Gary Becker (1983), holds that regulation is captured by industry and consequently is designed and operated for its benefit. Richard Posner (1974) added that empirical evidence showed the poor performance of the regulatory process usually benefited influence groups. In addition, regulation encourages rent-seeking behavior because interested parties are willing to invest in order to yield favorable decision-making by agencies or courts (Harnay and Marciano 2011).

Equally, economists have shown mechanisms allowing to correct market failures through the market without the need for governmental regulation, i.e., self-centered agents still maximize welfare even in case of market failures such as externalities or monopolies, for example, Coase's theorem regarding property rights and externalities or Demsetz's contestable markets showing that even if competition *within* the market is not possible, it still is possible to motivate competition *for* the market (Hägg 1997).

Institutional economics has shown that regulatory institutions (i.e., governments) can replace or assist private bargaining in the presence of externalities or risks born by groups facing high transaction costs or asymmetry of information endured by consumers prior to transactions. This insight might be associated with public interest, but it is not contradictory with public choice or the Chicago theory of regulation. Indeed, regulatory institutions might well be the result of pressure groups or the correction of a market failure (Hägg 1997).

Agency theory sustains that economic regulation fails as information is asymmetrical between regulators and regulated industry; hence, the latter can extract economic rents in the regulatory process. This phenomenon leads to agency capture. In other words, correction of market failures – aiming at enhancing general welfare – may not be possible as regulated industries drag out the regulatory process (Laffont and Tirole 1993). Again, this insight does not trump the public choice of the Chicago theory of regulation, but sheds light on the complexities of economic regulation, leaving space for public interests to be a driving force of regulation (Hägg 1997).

The discussion on collective choice shows that public decisions cannot satisfy every individual interest and the quest for public interest might be more complex than assuming that governments will promote it seamlessly. Notwithstanding, it is possible to find efficiency in cases of State intervention, despite the limitations annotated. Pareto criterion might be satisfied if at least one person is better off and none of the rest are worse off. Even in the former case, public interest decisions might be efficient if the well off can compensate the worse off, using the Kaldor-Hicks criterion (Coleman 1980).

Public Interest and Law

Public interest is ubiquitous within the realm of law. It proceduralizes collective interests, and courts adjudicate disputes to achieve justice. Public officials assess their actions against the interests of the whole community in implementing public policy by means of the law.

Traditionally, the law has used the concept of public interest as justification for government intervention in economic affairs. Black's Law Dictionary (2004) defines it as "The general welfare of the public that warrants recognition and protection" or "Something in which the public as a whole has a stake; esp., an interest that justifies governmental regulation."

In its regulatory justification, public interest originally appeared in the work of Lord Mathew Hale, *The Portibus Maris* (1787), inspiring two important decisions, one in England (*Allnutt v. Inglis*) and one in the United States (*Munn v. Illinois*). In essence, Lord Hale sustained that private businesses become *juris publici* if they were licensed or chartered by the King to act as a monopoly and those services were available to the public (Hantke-Domas 2003). The importance of the concept lies in the fact that some economic activities were in the interest of the public, and correspondingly it was necessary to balance expectations with the economic activity to obtain mutual benefit (Craig 1991).

The law embraces the concept of public interest in many ways besides being a warrant for governmental regulation. The UK Enterprise Act (2002) introduced a "public interest test" in cases of takeovers and mergers, by which the Secretary of State may intervene in the latter if interests of the public are involved as, for example, with issues of national security, media quality, plurality and standards, and financial stability ("Takeovers," n.d.). The UK Public Interest Disclosure Act (1998) bears a test that grants protection to whistle-blowing employees if the matter brought to light is in the public interest. Another important UK law is the Freedom of Information Act 2000 that provides for access to government-held information only if the public interest in disclosing it outweighs the public interest in not disclosing it. In these three laws, public interest is liberally used either to justify by the government a ban on a merger or to withhold or publicize information.

Civil service is another area where public interest is present ("Public Interest in UK Courts" 2011). In the United Kingdom, Lord Nolan's 7 Principles of Public Life are part of the remit of the Committee on Standards in Public Life advising the British government. For example, "selflessness" principle affirms, "Holders of public office should act solely in terms of the public interest." In Spain, its constitution (Section 103.1) obliges its civil service ("*administración pública*") to work for "general interests."

British courts may deal with public interest in different opportunities as defined by the Public Interest in UK Courts Project (2014). First, the law refers expressly to the concept, as in the case

of the English Freedom of Information Act 2000. Second, it refers to legal proceedings brought by the Attorney General in the "public interest." Third, the courts may invoke the concept to justify new developments in the law. Fourth, public interest immunity may be argued in court by the Crown to withhold documents from parties when such disclosure may be contrary to the public interest. Fifth, European Union law and the law of the European Convention on Human Rights recognize rights of European citizens but that these rights can be limited in the public interest in cases of national security, public health, and the prevention of crime.

Public officials' decisions in England are subject to judicial review regarding lawful exercise of their statutory powers. Public interest may be argued by courts to extend this review to non-statutory decisions. In continental Europe administrative law holds public officials accountable for the same reason, but with a comparative more developed corpus of law known as contentious administrative.

In Spain concepts such as "just economic and social order" or "quality of life" are written in her constitution and are part of the common good; hence "...public interest happens when the aim that must be served by a political or administrative organisation on its entirety, can only be achieved by that entirety" (Parejo 2003). More precisely, the general interest is fused together with the aims of the State. In the Spanish Constitution general interest refers to the protection of legal interests belonging to the community, a safeguard duty that must be assumed by the State as it is in charge of managing the common interest (Parejo 2003).

From a German perspective, the idea of common good is equally present, and for Prof. Eberhard Schmidt-Assmann (2003) it is the guidance for Parliament, public officials, and judges, but to be defended as well by private entities, interest groups, and individuals. Public interest strives to ensure the general interest. Both concepts are not interchangeable, but since public interest is promoted by the community, it tends to become general interest. Private interests are not opposed to general ones and in their evolution blend with the general interest. Public interest is not predefined and static, but it evolves throughout the administrative process. The administration – or government – is conceived as a structure representing multiple public interests. For Prof. Schmidt-Assmann general welfare is equivalent to common good, and public interest is the feasible fulfillment of common interests by means of the law. "The determination of what is general welfare is a matter depending above all on positive law, which normally offers for its achievement procedures and material criteria" (2003, p. 167). This proceduralization of public interest should be understood under the light of fundamental rights and the rules regarding the State function. Common good is the government-guiding ideal to act in favor of the community, and democracy is better served if common good is actively pursued.

The use of the concept of public interest as a heuristic tool is present in the United Kingdom and in Spain and Germany, but in Spain and Germany, it is a legal obligation to be followed by the State, whereas in the United Kingdom it guides a limited set of situations.

This simple overview of different uses of public interest in the law shows that institutions are always inspired to serve the community and public officials are bound to pursue the public interest. Economists, at least some of them, reject any possibility that a public official – or anyone – may act in any way but for his or her own interest; hence the public interest might be an excuse for advancement by protecting the interests of groups. Law and economics clash for being on the one hand idealist (law) whereas on the other hand pessimistic (economics), but the power of law guarantees that public officials will be restrained on their self-interest. The law carries penalties that cannot be defined as idealistic, such a jail terms or hefty fines.

Consistent with the Public Choice Theory and the Chicago Theory of Regulation that dispute the existence of a benevolent government always searching for maximum social welfare, in the last decades the promotion of integrity and the prevention of corruption in the public space have increased. Although studies and policies on

corruption do not focus on public interest, they tackle self-interest of public officials. The matter has taken a prominent position on the agenda of international organizations with the adoption of the United Nations Convention against Corruption (2003). Indeed, important efforts to oblige States to implement regulations to curb the phenomenon preceded the UN Convention, as the OECD Convention on Combating Bribery of Foreign Public Officials in International Business Transactions (1999) or the Organization of American States Inter-American Convention against Corruption (1996). National legislation has gathered pace (e.g., UK's Bribery Act 2010, US Foreign Corrupt Practices Act).

Under the light of integrity legislation, corruption emerges when a public official obtains a private gain or status from office. Indeed, corruption is defined as "the abuse of public office for private gain" (World Bank 1997). A more comprehensive definition is given by Robert Klitgaard (1988) "behaviour 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 behaviour."

Behind the international effort to curtail corruption is the belief – and evidence – that corruption distorts prices. Inefficient markets with covert and upward redistribution of wealth within a society increase costs and reduce investment and produce a decline in society's moral and ethics (Senior 2006). Corruption can be petty or grand, and even there are some authors that affirm that in some societies, corruption helps failed States to run properly.

Using Robert Klitgaard's (1988) corruption equation (corruption equals monopoly power plus discretion by officials minus accountability), if public officials are rational, then public interest – as opposed to personal interest – should be pursued by them if the expected gain by the corruption is less than the penalty of being caught and prosecuted. This conclusion leaves the possibility that public interest can be a motivation for public officials when corruption is not present.

Cross-References

- ▶ Administrative Corruption
- ▶ Altruism
- ▶ Constitutional Political Economy
- ▶ Cost–Benefit Analysis
- ▶ Efficiency
- ▶ Governance
- ▶ Market Failure: Analysis
- ▶ Political Corruption
- ▶ Public Choice: The Virginia School
- ▶ Rent Seeking
- ▶ Rule of Law

Acknowledgment The author would like to acknowledge the helpful research assistance of Ms Francisca Henriquez Prieto.

References

Arrow K (1951) Social choice and individual values, 2nd edn. Wiley, New York

Becker GS (1983) A theory of competition among pressure groups for political influence. Q J Econ 98(3):371. https://doi.org/10.2307/1886017

Brown C, Jackson P (1998) Public sector economics, 4th edn 1990. Blackwell Publishers, Oxford

Buchanan J, Tullock G (1962) The calculus of consent: logical foundations of constitutional democracy. University of Michigan Press, Ann Arbor

Coleman J (1980) Efficiency, Utility, and Wealth Maximization. Yale Law School Legal Scholarship Repository. Retrieved from http://digitalcommons.law.yale.edu/fss_papers/4202

Craig P (1991) Constitutions, property and regulation. Public Law 538–554

Downs A (1957) An economic theory of democracy. Harper, New York

Feintuck M (2010) Regulatory rationales beyond the economic: in search of the public interest. In: Baldwin R, Cave M, Lodge M (eds) The Oxford handbook of regulation. Oxford University Press, Oxford

Garner B (ed) (2004) Public interest. In: Black's law dictionary, 8th edn. Thompson West, St. Paul

Hägg PG (1997) Theories on the economics of regulation: a survey of the literature from a European perspective. Eur J Law Econ 4(4):337–370. https://doi.org/10.1023/A:1008632529703

Hale M (Sir) (1787) A treatise, in Three Parts. Pars Prima, De jure maris et brachiorum ejusdem. Pars Secunda, De portibus maris. Pars Tertia, Concerning the custom of goods imported and exported. Considerations touching the amendment or alteration of laws. A

discourse concerning the Courts of King's Bench and Common Pleas. In: Hargrave A (ed) Collection of tracts, vol I
Hantke-Domas M (2003) The public interest theory of regulation: non-existence or misinterpretation? Eur J Law Econ 15(2):165–194. https://doi.org/10.1023/A:1021814416688
Harnay S, Marciano A (2011) Seeking rents through class actions and legislative lobbying: a comparison. Eur J Law Econ 32:293–304
Public Interest in UK Courts (2011) Retrieved 30 July 2014, from http://publicinterest.info/
Klitgaard R (1988) Controlling corruption. University of California Press, Berkeley
Laffont J-J, Tirole J (1993) A theory of incentives in procurement and regulation. MIT Press, Cambridge, MA
Lewin L (1991) Self-interest and public interest in western politics (trans: Lavery D). Oxford University Press, Oxford
Lewis C (2006) In pursuit of the public interest. Public Adm Rev 66(5):694–701
Middleton R (1996) Government versus the market. Edward Elgar, Cheltenham
Mueller D (1989) Public choice. Cambridge University Press, Cambridge
Parejo L (2003) Derecho Administrativo. Ariel Derecho, Barcelona
Peltzman S (1976) Toward a more general theory of regulation. J Law Econ 19(2):211–240
Posner R (1974) Theories of economic regulation. Bell J Econ Manag Sci 5(2):335–358
Rawls J (1971) A theory of justice. Harvard University Press, Cambridge, MA
Schmidt-Assmann E (2003) La Teoría General del Derecho Administrativo como Sistema: Objeto y fundamentos de la construcción sistemática. (trans: Bacigalupo M, Barnés J, García Luengo J, García Macho R, Rodriguez de Santiago JM, Rodriguez Ruiz B, . . . Huergo A). Marcial Pons, Madrid
Senior I (2006) Corruption – the world's big C: cases, causes, consequences, cures. The Institute of Economic Affairs, London
Sola JV (2004) Constitución y Economía. LexisNexis Abeledo-Perrot, Buenos Aires
Stigler GJ (1971) The theory of economic regulation. Bell J Econ Manag Sci 2(1):3. https://doi.org/10.2307/3003160
Takeovers: the public interest test – Commons Library Standard Note. (n.d.). Retrieved 30 July 2014, from http://www.parliament.uk/business/publications/research/briefing-papers/SN05374/takeovers-the-public-interest-test
Tanzi V, Schuknecht L (2000) Public spending in the twentieth century: a global perspective. Cambridge University Press, Cambridge
Weisbrod BA, Benjamin CB (1978) APPENDIX: Some concepts of the public interest and public interest law. In: Weisbrod BA, Handler JF, Komesar NK (eds) Public interest law an economic and institutional analysis. University of California Press, Berkeley
World Bank (1997) Annual Meetings, WorldBank Group Issue Brief Corruption and Good Governance. Retrieved from www.worldbank.org/html/extdr/am97/br__corr.htm

Public Investments: Broadband

Nicola Matteucci
Department of Economics and Social Sciences, Marche Polytechnic University, Ancona, Italy

Abstract

After a long phase of liberalizations, privatizations, and deregulation, in telecommunications the policy pendulum has moved in the opposite direction, and the State is back in many countries. By mid-2000s, the deployment of first-generation broadband communications had reached high levels of diffusion and, thanks to market-friendly regulation, market competition has generally led consumer prices to decline. However, marked dynamics did not ensure universal service (especially in rural areas), nor the continuous upgrade of the networks, with the migration to higher capacity NGAN – two requisites increasingly demanded by downstream market developments (e-Services, cloud computing, IoT). This entry reconstructs the main market, Government and (cohesion) policy failures occurring in the sector, and the responses so far experimented in the EU, in an international perspective. In particular, it illustrates its State aid control system and analyzes the current trade-off that the new industrial policy is posing to the policy-maker in a liberalized sector.

Synonyms

State intervention in the market

Liberalization, Privatization, and Market-Friendly Regulation

Since the 1980s (in the USA) and then the 1990s in Europe, there was a mounting consensus toward market liberalization and privatization of public companies (see entries on ▶ State-Owned Enterprises) – especially those active in network industries and utilities. Telecommunications were one of the sectors most heavily interested, starting with the famous AT&T breakup in the USA. According to the received wisdom, the privatization of this sector should have brought the largest societal benefits – at least in terms of retail prices (Newbery 2004); a main motivation was that technological developments would have reduced the competitive bottlenecks that had been justifying public monopolies and heavy old-style regulation (such as rate of return); then, market liberalization would have spurred pro-competitive dynamics, be able to progressively deregulate the concerned sectors, and move from *ex ante* (regulation) to *ex post* (antitrust) norms. Such a received wisdom does inform the current European Regulatory Framework for electronic communications (ex Directive 2002/21/EC and its further amendments) (see entry on ▶ Telecommunications – if any has been planned for the European Regulatory Framework).

As a matter of fact, in parallel with the paradigm change unfolding at the political and institutional levels, a technological revolution was also occurring, characterized by the advent of digital technologies for signal coding, transmission, and reception, that would have been deeply transforming the economics of network communication industries – particularly that of television and telecommunications. In particular, thanks to digital technologies and growing equipment standardization, the legacy communication networks (*in primis*, copper, and cable TV ones) have been transformed in a multipurpose infrastructure and able to carry a converging array of distinct services: voice telephony, video content, and data. As a consequence of this technological breakthrough, telecommunication networks have progressively converged with traditional audiovisual media, and new hybrid communication platforms and services have been introduced, such as IPTV (Internet protocol TV) (Matteucci 2016). Concerning the market consequences of the technological and institutional transformation, it is conventionally believed that the outcomes differ by market (services and countries), although presenting some broad regularities. In wireless services (*in primis*, mobile phones, where digital technology caused a relevant relaxation of the existing spectrum constraints), liberalization and privatization were unquestionably pro-competitive and elicited a unprecedented surge of new investments by entrants in both developed and developing countries, with the best results achieved where spectrum management was more effective. In fixed telephony services, the impact of the institutional transformation remained less clear-cut, due to the natural monopoly features still persisting along the distribution network (local loop), the varying effectiveness of the regulatory and policy mix, and the multifaceted implications of digital technologies. For example, Bacchiocchi et al. (2011) find that for EU-15 countries in voice services, regulation played a larger impact in driving down prices than privatization. On a similar vein, Florio (2004), after studying with cost-benefit analysis tools the famous British Telecom privatization in the UK, reconsiders the received wisdom and concludes that effective regulation may be much more important for the overall welfare than ownership regimes (public versus private) per se and that the latter had a small effect on the long-term dynamics of consumer prices and productivity – at least in the UK. Indeed, this evidence recalls what standard microeconomic theory predicts that an unregulated monopolist (or dominant firm) sets profit-maximizing prices, higher than competitive ones: hence, regulation may be the crucial aspect to look at.

In today's Internet economy, where consumer's usage is increasingly multipurpose and hybrid and requests higher bandwidth capacity, broadband services stand as the main reference product for the telecommunication industry and the policy-makers, replacing the older emphasis on voice services – despite the fact that no formal universal service provision currently covers the

first, differently from the second, according to the existing European Regulatory Framework. In first-generation broadband (conventionally classified as those services offering from basic to premium data transmission rates, respectively, from 2 to 20 Mbps in downloading, and epitomized by ADSL technologies), we have now accumulated a sufficient evidence on the relation between institutional setting and industry performance, especially for EU countries. Here, liberalization together with wholesale access regulation brought about a higher degree of static competition that generally resulted in lower consumer prices. In turn, a favorable retail price dynamics – prevalent in non-concentrated markets – translated into higher take-up: for example, Distaso et al. (2006) found that, in EU-14, lower local loop unbundling prices (henceforth, LLU, the most important type of wholesale access regulation) encouraged the take-up of broadband subscriptions, even though, on the side of the generation of new investment, the detected effects were generally negative (Cambini and Jiang 2009). In the EU, the widespread adoption of the "ladder of investment" regulatory model (henceforth LOI, setting access levels and terms in a way to progressively incentivize entrants to move from service-based to infrastructural-based competition) is believed by its proponents to have contributed to rapidly create competitive retail markets while hopefully stimulating gradual investment strategies by new entrants aimed at building alternative networks (Caves 2014). However, among the encountered drawbacks, wholesale access regulation and LOI could not cater for providing adequate stimuli to rapid network rollout, differently from what achieved with infrastructural competition, where the latter has been available (mostly between DSL copper and cable networks) (Briglauer et al. 2014).

Another fundamental stylized fact in the industry has been that, after liberalizations and privatizations, competition (both service and infrastructural) and private investment for rolling out digital networks first and foremost unfolded in urban areas, while less populated (rural) and socioeconomically disadvantaged areas across the EU were left behind, because of the higher network deployment costs and of the insufficient revenues characterizing these territories: these conditions typically configure cases of market failures (a situation when an output or activity yields net social returns higher than private ones, so that its market production is suboptimal) and/or problems of social and territorial cohesion (the existence of large GDP differences and resource unbalances between citizens and between regions). In this case, regulation cannot help so much, since even the facilities of incumbent operators (those having significant market power, henceforth SMP) are frequently unavailable or insufficient in these laggard areas, and no universal service provision is in place for broadband services in the EU.

As a consequence, rural and marginal areas accumulated a substantial gap – both concerning the timing and the quality of the communication services offered and subscribed. In particular, in 2013 (the deadline for reaching the first infrastructural objective of the Digital Agenda for Europe, henceforth DAE – EC 2010), the broadband coverage of the EU homes was nearly completed when considered in nominal terms and including all the potential networks (wireless and wired, fixed and mobile) of the first generation; however, the coverage was sensibly lower, on average, if only the more performing fixed technologies (xDSL, cable, and the fixed wireless WiMax) are taken into account: nominally, the latter was at 97.2%, but in real terms, it was often substantially inferior (depending on country and legacy network characteristics). However, the gap of rural areas was still relevant: in fact, in the same year, the nominal fixed coverage of the EU rural areas was lagging at 89.8% of the population (79.9% in 2012) (data from Digital Agenda Scoreboard 2014). Then, if one considers the NGAN achievements (next-generation access networks, including at that time broadband technologies such as VDSL, Cable DOCSIS 3.0, and FTTP), figures were much more disappointing, with fast broadband (ensuring transmission rate >30 Mbps in downloading) covering on average only 62% of the EU homes, but again dropping to only 18.1% in rural areas (with the coverage mostly

attributable to VDSL deployments, involving the incremental upgrade of legacy copper-based networks with segments of fiber optic).

For these reasons, in the current phase of the sector, where a timely rollout of new access infrastructure assumes a paramount societal importance (especially for the wired ultrafast networks, which however involve the highest sunk costs for cabling the terminal link), the policy pendulum is shifting from old-type wholesale access regulation (LLU and LOI) to new models (incentive based and geographically differentiated) and, increasingly, to active industrial policies. The latter, involving an original role for investment with public money, is especially needed if one wants to foster the deployment of wired ultrafast networks to comply with the stated targets and deadlines of the DAE that envisage an ubiquitous and fast-increasing connectivity capacity of the territories as a way to enhance the market competitiveness and socioeconomic growth of the EU society.

The State Comes Back in Telecommunications

Industrial policies aimed at solving the digital divide started to be experienced with first-generation broadband, both in Europe and elsewhere, before igniting a hot economic policy debate with second generation (NGAN). Since 2003, a certain number of small-scale (regional or municipal) State aid (henceforth, SA) measures supporting network deployments had been authorized by the EU Commission and mainly financed with European structural and investment funds or national finances: these measures gradually unfolded across countries as soon as the market diffusion path of broadband services (supply-side coverage and demand-side subscriptions) was progressing and encountering the first difficulties and failures, in relation to the different country sector's maturity (the first measure approved was the Cumbria project, in the UK, EC 2003). These policies acquired momentum in coincidence with the US subprime mortgage financial crisis and the following worldwide macroeconomic turbulences. In the EU, the first noticeable initiative came out in 2008: the European Economy Recovery Act (EC 2008) loudly announced the introduction of specific measures and funds (roughly EUR 1 billion) earmarked for solving the digital divide affecting rural areas, focusing on the supply-side (investment in infrastructure). Similar and sometimes more ambitious initiatives were adopted by other industrialized countries, along a clear "interventionist" industrial policy agenda. In the USA, in 2009, the extensive budget of the American Recovery and Reinvestment Act included a sum of $7.2 billion for grants and loans targeting broadband unserved and underserved areas (LaRose et al. 2014): this Act had a clear protectionist flavor, arriving to mandate also a sort of "Buy American" condition for funding, that provoked strong opposition among NAFTA partners (Canada). Other developed countries, including Australia and New Zealand – famous for their usually orthodox neoliberal policy agendas – introduced similarly ambitious plans of public intervention; then, the countries experiencing the highest public support were Japan and South Korea, two recognized worldwide ICT (Information and Communication Technology) leaders. Finally, active industrial policies for broadband were also enacted in BRICS and developing countries – with mixed characteristics and results (for Latin America, Galperin et al. 2013).

In the EU, the awareness of the policy-maker about the real extent of the infrastructural digital divide grew slowly, in parallel with the improvements of the surveying methodologies and available statistics. As a matter of facts, several original initiatives of country territorial mapping began to show that the network coverage gaps were numerous, spotty and very granular, and far from being limited to marginal rural areas – as an uncritical "market failure" rationale had let to assume. On the contrary, in some member states like Italy, historically characterized by dispersed models of urbanization and business location, coverage gaps and service disruptions happened to be very frequent also in strategic socioeconomic areas (export-oriented industrial districts and leading touristic locations), besides being present at the

edges of metropolitan and other urban areas (Matteucci 2014).

At the European level, the draft of the 2009 Broadband Guidelines (EC 2009; updated by the 2013 version, EC 2013) and the DAE (EC 2010) marked a radical step in elaborating a less rhetoric and more pragmatic approach, with respect to the earlier volitive but rather unrealistic agendas for Information and Communication Technologies (henceforth ICT) and Research and Development (R&D) support – from the Lisbon strategy in 2000 to the i2010 action plan in 2005. In particular, the Broadband Guidelines and DAE first paid a special attention to the worrying sluggishness showed by the NGAN market transition in Europe, with respect to the reference areas (US and, above all, the Asian leaders). As a consequence, in this period, some EU member states started to frame systemic and more ambitious nationwide broadband plans, designed to solve the digital divide with mixed policies, financed by a variety of funds and instruments (including public-private partnerships). Finally, in 2015, acknowledging the continuing relevance of these policy priorities, the Commission confirmed the orientation of its policy issuing another "Investment Plan for Europe"(also called "Juncker Plan," EC 2014a), aimed to stimulate the EU economy with an initial budget of EUR 315 billion devoted to new investment – including NGAN networks; then, in late 2016, the Commission President announced that the European Fund for Strategic Investments (EFSI) (to be employed for the Investment plan) would be doubled by 2022.

The more interventionist policy phase that began at the end of the 2000s was motivated by a series of reasons – both structural and short term. A main one reappraises the traditional macroeconomic benefits of countercyclical fiscal policies (typically advocated by neo-Keynesians), after years of massive consensus on the neoliberal rhetoric and the EU "austerity" economic policy orientation, increasingly put under pressure after the 2008's financial crisis (Schmidt and Thatcher 2013). Then, more micro-founded explanations acknowledge the specific positive impact of broadband investment on driving innovation, GDP growth, productivity, and competitiveness (Analysys Mason, Tech4i2 Ltd 2013) – an impact that is proved to be sizable, yielding estimates for the demand multiplier within the range 1.2–1.5. The micro-founded dynamics at work are several. Besides increasing directly the GDP through public expenditures, broadband networks, as many ICT, are "general purpose technologies," thereby possessing key valuable characteristics enhancing their socioeconomic impact. First, broadband infrastructure is particularly strategic since it channels and exchange increasing amounts of codified information and knowledge that, being public goods and involving positive externalities, are a primary source of innovation and economic growth. Second, broadband investment both spurs large complementary expenditures in other goods and services in the downstream economy. Third, broadband services are highly pervasive, and they enable further innovation and productivity gains in the using sectors, also due to the powerful network externalities and coordination dynamics (of the type "chicken egg") unfolding with downstream ICT and not-ICT sectors.

At the same time, investment in broadband networks creates its own demand and market (as often most types of infrastructure do, in a "Say's law" vein); for this, infrastructural investment in a free market economy may well experience coordination failures and holdup behavior by risk-adverse network operators. As a matter of fact, in the post-liberalization phase, the generalized drop in the investment rate experienced in the sector can be interpreted as driven by a rational strategy of privatized fixed network incumbents (now subject to the stock markets dividend constraint): they first aim to avoid cannibalization between NGAN and existing services and, second, opt for incremental and less risky investment paths (e.g., using cheaper FTTC architectures, instead of FTTH ones) to migrate to the new NGAN paradigm. Hence, in network sectors severe market, failures can occur, leading to service under provision and delays contrary to the public interest, however expressed. As a main example, in the current process of digitalization of the public administration (Seri et al. 2014), no forced migration (sunset dates) to digital services like e-Government, e-Health, or e-Participation can be reasonably mandated by

public authorities until all the population had been serviced with adequate broadband access to the Internet. Similarly, promising market developments such as cloud computing and the Internet of Things (IoT) require ubiquitous and constantly upgraded broadband capacity, but this frontier can be severely retarded – or even compromised – by the extent of the old and new types of infrastructural digital divide.

Indeed, for the EU to compromise on reducing the digital divide while not interfering with the outcome of the post-liberalization phase has proved much more difficult than elsewhere, due to the constitutional status of the general prohibition of SA, ex Art. 107(1) TFEU, and its stringent exceptions. At the end, the target of completing the coverage of broadband, although not submitted to a formal provision of universal service under the current Regulatory Framework, was codified as two distinct infrastructural targets of the DAE (EC 2010): to complete the coverage of the EU population (1) with first-generation services by 2013 and (2) with second-generation ones (NGAN) by 2020. These policy targets can be achieved through generous market incentives (various types of subsidies) and active policy-making; the latter, for some member states, initiated a new challenging season of direct public ownership of the built infrastructure.

EU State Aid Control for Broadband

The current version (2013) of the Broadband Guidelines (henceforth Guidelines) represents the *summa* of the relevant EU policy-making, detailing all the admissible forms of public intervention in the sector ("vertical" SA). Acknowledging that liberalization and privatization profoundly reshaped the telecom market, the Commission first illustrates (EC 2013: Artt. 9–54) the main principles and instruments that any measure involving public money has to follow, to comply with EU competition norms. Basically, there are three main instruments for public financial intervention (see Table 1), according to

Public Investments: Broadband, Table 1 Public support for broadband networks: instruments available in the EU

Instrument	EU founding norm (s)	Goal(s)	Main criterion of scrutiny	Example case
State aid	Art. 107-9 TFEU	Prohibition of any unlawful form of SA, subject to predefined exceptions	Use of State resources, economic selective advantage potentially distorting competition and affecting internal market trade	See the List of the Commission's decisions (source)
Market economy investor principle	Art. 345 TFEU	Equal treatment between private and public regimes of property ownership and entrepreneurship	Public financial support to an undertaking paid under normal market conditions	Case C 53/06-NL *Citynet Amsterdam-Investment* (*FttH*) *network*
Service of general economic interest	Article 106 (2) TFEU, the "Altmark case" jurisprudence, and "SGEI package"	Guaranteed supply of a SGEI	Imposition of public service obligations in return of specific compensation	Case N 331/08-F *THD Hauts-de-Seine*
Other administrative and regulatory measures	Regulatory Framework, SA case law, Directive 2014/61/CE	Achieve better market transparency and coordination and investment costs savings	To provide more information, transparency, open access to and efficient construction of network infrastructure	National legislation and regulation implementing the Directive 2014/61/CE

Source: our compilation from various sources, including EC (2013, 2014), and the List of the Commission's decisions (http://ec.europa.eu/competition/sectors/telecommunications/broadband_decisions.pdf)

the Treaty on the Functioning of the European Union (TFEU, EU 2012) and the associated hard and soft law norms: (1) the SA norms, (2) the application of the Market Economy Investor Principle (henceforth, MEIP), and (3) the imposition of the special regime of Service of General Economic Interest (SGEI). Further supporting initiatives may concern instruments not falling within the meaning of SA, such as administrative and regulatory measures aimed at facilitating market transparency and coordination. A main example is the "Cost Reduction" Directive (n. 2014/61/CE, EU 2014), due to be implemented by the beginning of 2016, that foresees four main pillars of normative harmonization and promotion of best practices: (1) mandating reusage of existing physical infrastructure to avoid inefficient duplication, on behalf of market entrants; (2) enforcing coordination and transparency of planned civil works, catering for efficient construction; (3) enabling faster, simpler, and more transparent concession of administrative permits; and (4) modernizing norms applicable to new buildings, to be equipped with NGAN-ready facilities.

MEIP and the SGEI are the residual cases: while MEIP concerns public intervention carried out "at market terms" (in commercially profitable areas), SGEI de facto regards not profitable areas ("white areas"; see *infra*). Both have been used rather infrequently, due to the application complexities and the potential institutional inconsistencies and normative loopholes involved. Basically, the MEIP case follows the (now somehow neglected) principle of equal treatment between private of public property ownership and entrepreneurship and involves the direct or indirect attribution of public capital (or other tool) to an undertaking under circumstances which mirror normal market conditions (replicating the case of a standard private investor), thereby not qualifying as SA under Art 107 TFEU. The SGEI case, in turn, requires to demonstrate that existing private operators cannot offer in the near-future adequate broadband coverage in a universal way, thereby excluding a portion of the population. Under this prerequisite, an undertaking may be entrusted with the operation of an appropriately defined SGEI, receiving in return a public service compensation. However, the accompanying conditions are quite demanding, since they are aimed at excluding any possible market interference with existing operators: for example, the motivation of inadequate digital connectivity for businesses in the targeted area would not suffice nor would be possible to entrust the SGEI broadband undertaking with retail operations. In other terms, SGEI feasibility in this sector needs to be limited to the network construction and the supply of wholesale access services through the public offer of the built passive infrastructure, neutrally open (EC 2013, Art. 21–25), whenever such an offer has been demonstrated as being unavailable from private operators. It follows that the practical feasibility of setting a SGEI nowadays is severely limited in any sector that underwent a massive liberalization process, like in the case of telecom and broadband; equally, as it is now, it is very difficult to use the SGEI instrument to relaunch the paradigm of the public company, where this has been privatized.

In essence, the SA norms (see entry ▶ State Aids and Subsidies) are the default system of public intervention – also for broadband networks. Art. 107(1) TFEU states a general prohibition of any unlawful form of SA, subject to predefined exceptions (ex Art. 107(2) and (3)), aimed at safeguarding the normal functioning of the internal market. The latter goal is also targeted by the sector liberalization process, and in this sense, the Guidelines match the general SA norms with the principles of the Regulatory Framework, setting binding requisites for efficient and not-distorting state intervention. Accordingly, any measure financed with public money must be preceded by a detailed mapping of the territory, aimed at distinguishing three types of areas: "black," where infrastructural competition between at least two network operators is present; "gray," where the service is provided by a unique network; and "white," where no service is available, now or in the foreseeable future (3 years). Potentially, public intervention is admissible in white areas, while it becomes more difficult to authorize in gray ones and normally is not admitted in black areas (except for ultrafast NGA networks (offering at least 100 Mbps speeds), where the 2013 Guidelines

opened a possibility upon proving the performance "step change" with respect to existing fast NGAN, EC 2013; Artt. 82–85). Mapping of the territory must be conducted at the national level through public consultations with telecom operators, and, once the measure is authorized, any public fund needs to be granted on the basis of open, transparent, and competitive procedures, where bidding operators cannot be discriminated on the basis of the solutions employed ("technological neutrality" principle). Then, once built, the publicly owned or the private infrastructure that got the subsidy (both arrangements are possible) is subject to a reinforced wholesale access regime – independently of the SMP qualification of the winner – and network operations must be monitored for a certain period, to avoid overcompensation ("claw back" clause on the extra profits).

Over time, the Commission's practice on broadband has evolved and adopted a "more stringent economic approach," aimed at enhancing its market-friendliness (in line with the 2005 State Aid Action Plan and continuing with the 2012 State Aid Modernization Initiative). This also implied a growing complexity of the relevant soft law that the Commission has somehow tried to streamline by pushing forward the convergence with other "horizontal" SA norms, for example, the recent inclusion of the easier cases (involving only broadband "white" areas) in the General Block Exemption Regulation or the solution of the previous loopholes contained in the 2007–2013 Regional Aid Guidelines, which had provided a more generous authorization pathway.

Since 2003, a large number of SA cases have been analyzed by the EU Commission: as of July 2016, 148 decisions (mostly positive) have been finalized, with a growing presence of country-wide measures (national plans) for NGAN deployment, involving higher budgets; the latter fact is attributable to the larger deployment costs involved in second-generation broadband – mostly, civil engineering works for digging and laying passive infrastructure along the access network (such as ducts, wires and antennas), arriving to account for 70–80% of the total investment costs. For the largest EU member states (the "big 5"), the representative budget of the most recently authorized national NGAN plans ranges from EUR 2 to 4 billion. On overall, the received wisdom predicates that the net balance of the Commission's activity is positive: for example, Chirico and Gaál (2014) notice that only a tiny minority of its numerous decisions was appealed.

As a matter of fact, the net balance of the SA activity is multifaceted and much more complex to gauge – even sticking to strict economic considerations. First, economic approaches used in antitrust theory and practice remain prevalently static, so that they miss to consider alternatives of public intervention that may score higher when examined in a truly dynamic setting. Sometimes, even some SA pro-competitive principles (such as that of technological neutrality) can yield unintended effects: while the first version of the Guidelines was inspired by a truly "driving" conception of SA policy (Gómez-Barroso and Feijóo 2012), there are good reasons to believe that the current de facto policy assimilation between wireline and wireless NGAN (adopted since EC 2013) can give incentive to less ambitious rollout plans and networks. Second, EU competition law continues to present cases of institutional conflict and normative loophole with other domains – *in primis* with regional and urban policies – and their possibly diverging objectives of social and territorial cohesion (Colomb and Santinha 2014). Third, while the standard Commission's scrutiny is based on *ex ante* evaluation, there is a real scarcity of evidence and literature on *ex post* evaluation of the implemented measures, which in turn is aggravated by the enduring lack of statistical evidence surrounding the sector's operations and the heterogeneous member states' institutional performance – especially at the disaggregated territorial level. Indeed, some broadband SA measures (e.g., EC 2014b) have been encountering severe implementation retards and even problems of institutional capacity that led to a perverse mix of market and Government failures, delaying a prompt achievement of the intended goals (especially, from a cohesion point of view).

Conclusion and Future Directions

To sum up, taking into account the worldwide trend toward higher State intervention on the market and active industrial policy, and their stringent trade consequences (including cases of "beggar-thy-neighbor" policies), it is difficult to predict whether the delicate policy equilibrium so far achieved in SA practice for liberalized network sectors in EU can be firmly maintained in the future – especially in the post-Brexit era. In particular, SA policy could be affected by much stronger pressures and conflicts from member states than before. Moving to a larger picture, in the most pessimistic scenario, due to the constitutional nature of the SA norms, the potential imbalances arising from its application could even arrive to damage the social and institutional cohesion of the EU.

References

Analysys Mason, Tech4i2 Ltd (2013) The socio-economic impact of bandwidth. Report for the European Commission, London. Available at: http://ec.europa.eu/digital-agenda/en/news/study-socio-economic-impact-bandwidth-smart-20100033

Bacchiocchi E, Florio M, Gambaro M (2011) Telecom reforms in the EU: prices and consumers' satisfaction. Telecommun Policy 35(4):382–396

Briglauer W, Frübing S, Vogelsang I (2014) The impact of alternative public policies on the deployment of new communications infrastructure – a survey. Rev Netw Econ 13(3):227–270

Cambini C, Jiang Y (2009) Broadband investment and regulation: a literature review. Telecommun Policy 33(10–11):559–574

Cave M (2014) The ladder of investment in Europe, in retrospect and prospect. Telecommun Policy 38(8–9): 674–683

Chirico F, Gaál N (2014) A decade of State aid control in the field of broadband. Eur State Aid Law Q 1:28–38

Colomb C, Santinha G (2014) European Union competition policy and the European territorial cohesion agenda: an impossible reconciliation? State aid rules and public service liberalization through the european spatial planning lens. Eur Plan Stud 22(3):459–480

Digital Agenda Scoreboard (2014). Available at: https://ec.europa.eu/digital-single-market/en/digital-scoreboard

Distaso W, Lupi P, Manenti FM (2006) Platform competition and broadband uptake: theory and empirical evidence from the European Union. Inf Econ Policy 18(1):87–106

European Commission (EC) (2003) Cumbria broadband – project access – advancing communication for cumbria and enabling sustainable services, Case N282/2003 – United Kingdom, 10.12.2003, C(2003) 4480fin, Brussels

European Commission (2008) A European economic recovery plan, communication from the Commission to the European Council, 26.11.2008, COM(2008)800 final, Brussels

European Commission (2009) Community guidelines for the application of State aid rules in relation to rapid deployment of broadband networks, Communication 2009/C 235/04, 30.9.2009, Brussels

European Commission (2010) A digital agenda for Europe, COM(2010)245, 19 May, Brussels

European Commission (2013) EU guidelines for the application of State aid rules in relation to the rapid deployment of broadband networks, Communication 2013/C 25/01, 26.1.2013, Brussels

European Commission (2014a) An investment plan for Europe, Communication from the Commission, COM/2014/0903 final, Brussels

European Commission (2014b) Prolongation of the National Broadband Plan Italy, Case SA.38025 (2014/NN) – Italy, 11.12.2014, C(2014) 9725 final, Brussels

European Union (2012) Consolidated version of the Treaty on the Functioning of the European Union, 2012/C 326/01, 26.10.2012, Brussels

European Union (2014) Directive 2014/61/EU of the European Parliament and of the Council of 15 May 2014 on measures to reduce the cost of deploying high-speed electronic communications networks, 23.5.2014, Brussels

Florio M (2004) The great divestiture. Evaluating the welfare impact of the British privatisations 1979–1997. The Mit Press, Cambridge, MA

Galperin H, Mariscal J, Viecens FM (2013) One goal, different strategies: an analysis of national broadband plans in Latin America. Info 15(3):25–38

Gómez-Barroso J, Feijóo C (2012) Volition versus feasibility: state aid when aid is looked upon favourably: the broadband example. Eur J Law Econ 34(2): 347–364

LaRose R, Bauer JM, DeMaagd K, Chew HE, Ma W, Jung Y (2014) Public broadband investment priorities in the United States: an analysis of the broadband technology opportunities program. Gov Inf Q 31(1): 53–64

Matteucci N (2014) L'investimento nelle reti NGA a larga banda: la 'questione settentrionale' (The investment in NGA broadband networks: the 'Northern question'). Econ Polit Ind J Ind Bus Econ 41(4):9–25

Matteucci N (2016) Standards, IPR and digital TV convergence: theories and empirical evidence. In: Lugmayr A, Dal Zotto C (eds) Media convergence handbook, vol I, 1st edn. Springer, Berlin

Newbery D M. (2004) Privatising network industries. Cesifo working paper n. 1132, Feb

Seri P, Bianchi A, Matteucci N (2014) Diffusion and usage of public e-Services in Europe: an assessment of country level indicators and drivers. Telecommun Policy 38(5–6):496–513
Schmidt VA, Thatcher M (eds) (2013) Resilient liberalism in Europe's political economy. Cambridge University Press, Cambridge

Public-Private Partnerships

Luciano Greco
Department of Economics and Management and CRIEP, University of Padua, Padua, Italy

Definition

Public-private partnerships (PPPs) are organizational forms involving public and private institutions and aiming at the provision of assets, goods, or services that, to a large extent, are relevant in terms of public interest. Alternatively, the same tasks (e.g., construction, operation, maintenance, financing) can be pursued by full-fledged public organizations. The nature of tasks and involved institutions (e.g., for-profit firms, nongovernmental organizations (NGOs), governmental agencies) and the constraints shaping interactions among partners (e.g., contract or information incompleteness) determine different effects as regards the distribution of risks and payoffs among partners and allocative efficiency.

Economics and Institutions

Several concepts of PPPs have been introduced by practitioners and academics. A comprehensive perspective about PPPs can be found in the management literature. Kivleniece and Quélin (2012) define *"public-private ties [. . .] as any long-term collaborative relationships between one or more private actors and public bodies that combine public sector management or oversight with a private partner's resources and competences for a direct provision of a public good or service."* In such a rather extensive definition, we may include a quite large array of institutional arrangements and activities, such as procurement of public infrastructures, collaborations to face societal challenges (e.g., poverty or disease eradication in developing countries), or research-driven joint ventures (Perkmann and Schildt 2015).

In line with the purpose of this essay, we focus on a narrower concept of PPPs as *"an agreement by which the government contracts a private company [or a consortium of firms] to build or improve infrastructure works and to subsequently maintain [and operate] them for an extended period of time [. . .] in exchange for a stream of revenues during the life of the contract"* (Engel et al. 2014, p. 2). The motivation is twofold. First, the theoretical (particularly, contract theory) literature on PPPs has primarily focused on the optimal design of incentive schemes and on the assessment of costs and benefits of *outsourcing* of important phases of public investment projects. However, the bulk of findings of these contributions can be extended to non-infrastructure PPPs. Second, the most common and recurrent use of PPP concept among practitioners is related to public infrastructures (several initiatives of public institutions in different countries are going to change this: e.g., the EU Commission has recently introduced the concept of *research public-private partnerships* which are joint ventures between public and private institutions to pursue common interest tasks related to large investments in research).

In the public debate, PPPs are often confused with privatizations. In both cases, the government may transfer public assets and activities to private institutions (possibly, for-profit firms) who are then in charge of running them. Nevertheless, two differences between such organizational forms are crucial. First, privatized assets and activities are not (anymore) part of government-specific tasks, while PPP infrastructures and services are (still) part of them. For example, a prison or a barrack continues to be essential facilities of law-and-order or defense governmental policies, even though these are managed by a PPP agreement (instead of a governmental agency). On the contrary, a privatized firm or building serves almost exclusively private purposes (e.g., profit

maximization). In many cases, government continues to play a supervisory role on privatized firms that, on a normative point of view, is not very different with respect to that on other firms in the same sector (this is quite clear in the case of competitive markets, but the same argument holds also for regulated sectors, where governmental authorities play just a market-failure-correction role). Second, privatizations are intended to be forever (unless major political changes foster re*nationalization*), while PPPs feature (long but) finished terms, whereby assets and activities are transferred back to government at the end of contracts.

The previous argument has important policy implications. Consider that the government outsources to private institutions activities featuring a strong public interest (e.g., construction and operation of a hospital in countries where healthcare is publicly provided). Independent of the specific organizational setting, such an arrangement is a PPP. Indeed, outsourcing does not divest government of its role, nor it diminishes its responsibilities as regards final outcomes. For example, in the case of healthcare PPPs, any pitfall in the functioning of a hospital implies legal, economic, and political responsibilities for the governmental authority in charge of healthcare provision.

Practitioners' View

PPPs – as above defined – are a well-established practice to construct and operate public infrastructures in several countries (Bezançon 2005; Engel et al. 2014). Practitioners from public institutions and private business have pointed out three channels by which PPPs may improve social welfare, as compared to traditional, government-led frameworks:

1. *Efficiency gains* may derive from enhanced management of tasks – e.g., design, construction, and operation – and risks (many contributions in economics and management literatures emphasize innovation as an autonomous source of dynamic efficiency gains; for the purpose of this essay, we classify it among efficiency gains from enhanced task management).
2. The government may be able to attract and channel private financing into initiatives of public interest, thus relieving the public finance (i.e., *financial leverage*).
3. Government infrastructure policies are affected by perverse political incentives (e.g., investments in "white elephants," procrastination of maintenance expenditures, expenditure political cycles) that may be corrected by the involvement of private partners (i.e., *market discipline*).

The international experience has shown that such benefits are not systematically delivered by PPPs (Engel et al. 2014; Saussier 2015). Also, in the last 30 years, these organizational forms have been characterized by growing financial and institutional complexity. For these reasons, many countries have introduced public oversight mechanisms. A central task of the latter is the *ex ante assessment* of the balance between costs and benefits (i.e., the so-called value for money) of PPPs, as compared to traditional public organization (Burger and Hawkesworth 2011).

In the last 20 years, different strands of academic literature have investigated the conditions under which the abovementioned benefits are likely to materialize.

Contract Theory Literature

As argued, complexity has become a distinguishing feature of PPPs. For example, the most prominent form of such arrangements, the *project financing* technique, puts a *project company* – also called *special purpose vehicle* or *consortium* – at the center of *"a web of contracts aimed at distributing tasks, risks, costs, and revenues among all (private) partners"* (Greco 2015). This explains why, in the last decades, contract theory contributions have been quite useful to elicit the main drivers of potential benefits of PPPs.

The standard analytical framework to assess the role and the features of PPPs is based on a *sequential moral hazard* problem (Hart 2003; Martimort and Pouyet 2008; Iossa and Martimort 2015). In this essay, we abstract from issues related to the *selection* of firms, and we focus

only on the post-contractual problems that are typical of long-term interactions like PPPs. In the simplest case, the public project cycle can be represented by two sequential tasks: in the first phase, a public infrastructure (e.g., a highway) has to be built; in the second phase, the infrastructure has to be operated and maintained to provide services (e.g., freight and passenger transportation). The social value of the public infrastructure is determined by the investment effort, which is implemented in the building phase, and by the management effort, which is implemented in the operation phase.

As a matter of fact, governments need external resources and competences to build and, in many cases, to manage public infrastructures. In this context, the main policy issue is whether it is better to outsource the two tasks to different agents (i.e., a building firm, in the first phase, and an operating-and-maintenance firm, in the second phase) or to a single agent (i.e., a consortium of firms that are specialized in different tasks). In the literature, the first setting – based on sequential contracting – is often called *unbundling* of tasks, as opposed to the alternative *bundling* of tasks. The latter represents, in a very stylized fashion, the basic contractual structure of PPPs.

A Basic Model

To grasp the essential issues of the choice between bundling and unbundling – which explains efficiency gains from PPPs – let us consider a very simplistic model. The quality of public infrastructures is almost always multidimensional; however, for our purposes, we consider the monetary value of the capacity of the infrastructure to satisfy end users and, possibly, to produce positive externalities on the whole society. Thus, the investment in infrastructure (say, highway) quality can be either high – i.e., $I = 1$ – or low, i.e., $I = 0$, and generates a gross social benefit $B = \tilde{B} + \mathrm{bI}$, where $\tilde{B}$ is the exogenous (possibly, random) benefit generated by a minimum-quality highway and $b > 0$ is the marginal benefit of increasing the quality of the highway.

During the management phase, the operation and maintenance costs are $C = \tilde{C} - \delta I$, where $\tilde{C}$ is the exogenous (possibly, random) cost component and δ is the marginal impact (or externality) of the first-phase investment on operation-and-maintenance costs. In most contract theory models on PPPs, the latter plays a crucial role: $\delta > 0$ implies that the first-phase investment is also beneficial in terms of second-phase infrastructure management; such an impact is negative when $\delta < 0$. A couple of examples may clarify this point: investments in technologies for touchless tolling improve the service quality of a new (or newly refurbished) highway but also reduce operation costs (i.e., labor costs), hence $\delta > 0$; conversely, investments to increase the number of lanes of a highway reduce congestion and, thus, increase the service quality but also maintenance costs during the operation phase, i.e., $\delta < 0$. For the sake of simplicity, we abstract from investments and managerial efforts that are typically implemented in the second phase to improve the social value of services provided by the infrastructure.

The *first best optimal outcome* is obtained by choosing $I = 1$ if and only if $b + \delta > 1$ – this result derives by the maximization of the social welfare: $B - I - C = \tilde{B} - \tilde{C} + (b + \delta - 1)I$. In the following, we assume that this condition is met; hence, investing in quality is always optimal from the point of view of the whole society. A benevolent government could attain the first best optimal outcome in a world where contracting with firms to outsource tasks is not affected by any significant informational or contractual imperfection. More technically, information is symmetric – i.e., investments are fully observable – and contracts are complete, i.e., investments can be verified by a court without any cost. In such a "wonderful world," the choice between bundling and unbundling is immaterial. Any contractual scheme can be used to reach the first best. On the contrary, the choice between PPPs and sequential contracting becomes relevant when outsourcing is affected by substantial informational or contractual imperfections.

Asymmetric Information and Incentive Design

If the government cannot observe the first-phase investment of the builder (i.e., a building firm, under unbundling, or a consortium, under

bundling), neither it can infer it by the observation of B or C (because of the noise created by the exogenous random components $\tilde{B}$ and $\tilde{C}$), a classic moral hazard problem arises. In the absence of any incentive mechanism, the builder has no interest to invest in infrastructure quality. However, the government can design contracts to provide (monetary) incentives to the builder.

Under unbundled contracting, the government contracts with a building firm, in the first place, and then with another firm for operation-and-maintenance tasks. In our simplified setting, there is no moral hazard problem for operation and management. Thus, the second-stage contract is optimally a cost-plus one, $T_o = C$. In turn, the operator earns a zero profit and does not face any operation risk. As regards the building phase, the government can design an incentive payment scheme that is based on observable variables, $T_b(B) = T + tB$, where T is the fixed-price component of the payment to the builder and t is the marginal reward for each unit of extra benefits generated by the infrastructure. The government (correctly) anticipates the reaction of the builder, that is, to invest in quality if and only if $tb > 1 -$ i.e., the building firm maximizes its profit: $T - I = T + t\tilde{B} + (\mathrm{tb} - 1)I$. Thus, the government can control the decision of the builder by regulating the power of incentives: for $t > \underline{t}_u = \frac{1}{b}$, the builder invests.

Under bundled contracting, the government outsources both tasks to a single consortium of firms. The payment function is now $T_p(B, C) = T + t(B - C)$. Also in this case, the government anticipates that the consortium invests in high quality if and only if $t(b+\delta) > 1 -$ i.e., the consortium maximizes its profit: $T - I - C = T + t\left(\tilde{B} - \tilde{C}\right) + [t(b + \delta) - 1 + \delta]I$. Again, the government can control the decision to invest of the consortium by a suitable power of incentives: for $t > \underline{t}_p = \frac{1-\delta}{b+\delta}$, the consortium invests.

Let us remark that the minimum power of incentives to induce investment in infrastructure quality is different with respect to what we have in the unbundling case. In particular, *the power of incentives has to be stronger (respectively, weaker) under unbundling than under bundling if* $\delta > 0$ *(respectively,* $\delta < 0$*)*. The previous statement follows by the comparison of $\underline{t}_u$ and $\underline{t}_p$. This difference is irrelevant as far as increasing the power of incentives, i.e., t, is costless, which is the case when moral hazard is the only constraint that affects contracting between government and firms. In turn, both contractual forms afford the first best optimal outcome, and again the choice between bundling and unbundling is immaterial.

In the real world, facing higher-power incentives usually involves additional costs for the agents. For example, *risk-averse* firms require a risk premium (i.e., larger expected returns) to accept riskier contracts (i.e., larger t). In turn, to provide appropriate incentives to firms, the government has also to accept to pay larger expected transfers. A similar effect is determined by *limited liability* (or wealth) constraints faced by firms, which introduce lower bounds to government payments. By the previous arguments, we have:

Proposition 1 *Under asymmetric information and costly incentives, the government prefers bundling over unbundling (or the reverse) if and only if* $\delta > 0$ *(or* $\delta < 0$*).*

The basic idea is that the internalization of the externality between first-phase investment and second-phase operation costs reinforces (or weakens) contractual incentives if the externality is positive (or negative). A first policy implication of this literature is that *the enhanced management of synergic tasks is a crucial driver of efficiency gains from PPPs, because of information and financial frictions in the economy.*

Incomplete Contracts, Incentives, and Property Rights

In many real-world situations, observable variables – such as the quality of a hospital or a prison – cannot be assessed with a sufficient degree of precision to be verified by a court at negligible costs (Hart 2003). These quite common situations are characterized by incomplete contracts.

Incomplete contracting has dramatic impacts on the functioning of (alternative) contractual

schemes between government and firms. In particular, a *holdup* problem arises: in the terms of our model, the government is unable to commit to reward investments on the basis of observed (but unverifiable) infrastructure quality; firms rationally anticipate this outcome and do not invest.

Assuming that the ownership of the public infrastructure belongs to the government, the outcome of unbundling and bundling is equivalent to what would happen in the asymmetric information model (that we considered in the last section), when the only contract that the government can sign is such that $t = 0$. Thus, we have:

Proposition 2 *Under incomplete contracts and government ownership, the government prefers bundling over unbundling if and only if* $\delta > 1$*; otherwise, the two contractual schemes determine the same outcome.*

Contract incompleteness destroys the capacity of government to design (monetary) incentives for firms. In this situation, PPPs may improve social welfare if the intrinsic incentive underlying bundling is sufficiently strong (see Proposition 2).

In this framework, another important driver of incentives is the *distribution of rights* between government and partners (Bennett and Iossa 2006). To clarify this point, let us consider a specific case. If the public infrastructure is a building for (government) offices, it may clearly have alternative uses such that the quality of the infrastructure can be (fully) rewarded on the private market (e.g., a building for private offices). In this case, awarding ownership over such infrastructure to the builder reinforces its incentives to invest in quality. Indeed, once the infrastructure is built and its (high) quality observable, if the government does not pay to reward such quality, the private firm owning the infrastructure can renegotiate the terms of the contract with the government and sell the building on the private market. Anticipating such a situation, the builder invests in quality.

Four remarks are in order. First, the result with private ownership is the opposite with respect to the case of government ownership where, as above discussed, the builder has no incentive to invest. Thus, the distribution of rights is a (rough) tool to regulate the power of builder's incentives.

Second, the main driver of the power of incentives (under private ownership) is the market value of alternative uses of the public infrastructure. In some cases (as in the previous example), the market value is likely to fully reward investments in quality. However, in the generality of cases, public infrastructures are not perfectly fungible for private purposes. In turn, only a fraction of investments in quality can be rewarded, which dampens incentives to invest. In some cases, no reward can be found on the private market for investments in public infrastructure quality (e.g., electricity networks). As the market value of alternative uses of the public infrastructure goes down, private ownership loses the capacity to convey incentives to invest in quality.

Third, a result similar to Proposition 2 can be obtained in any case: the incentive that is embedded in bundled contracting simply adds to the (possible) incentive that is determined by private ownership over public infrastructure.

Fourth, the contract theory literature has highlighted that the optimal design of property rights should also take into account the nature of outsourced assets and activities that, in turn, drives the interests of government and private partners in final outcomes (Besley and Ghatak 2001; Schmitz 2015).

The policy implication of these arguments is that *the distribution of rights compounds the design of monetary incentives to determine the results of alternative contractual schemes, in frameworks affected by contract incompleteness and asymmetric information.*

Extensions and Future Directions

The contract theory literature has confirmed that PPPs may deliver relevant efficiency gains whenever some form of externality exists between sequential tasks that constitute a public project cycle. Such an externality has not necessarily a technological nature, but it may also arise by financial constraints affecting private partners.

The previous arguments highlighted another crucial point, i.e., the proper distribution of risks among partners. Relying on a simple asymmetric information model, we pointed out that contracts that transfer risk to firms are useful to enforce stronger incentives but also involve additional costs for the government aiming at inducing firms to bear more risks (Martimort and Pouyet 2008; Iossa and Martimort 2015). In turn, it is crucial to assess which risks private contractors are able to handle and at which cost. Also, contractual designs should be conceived to improve appropriate risk sharing between government and firms (Engel et al. 1997).

This literature has also provided clear indications about another alleged benefit of PPPs identified by practitioners. An important result is that the *financial leverage* motive to undertake PPPs is clearly wrong (Engel et al. 2013). To put it in a very simple way, *the private financing of public infrastructures is a sort of public debt*. The reason why we observe, in the real world, a strong correlation between fiscal restraints and PPP investments is often argued to be linked to political incentives – e.g., in many countries public finance regulations may create incentives to undertake PPPs in order to bypass fiscal rules (Maskin and Tirole 2008). But such motives are not proven by compelling empirical evidence (Buso et al. 2017), and this point still needs deeper investigation.

The international experience shows that PPPs tend to work quite well in some sectors (e.g., highways, ports), while they tend to fail in others (Engel et al. 2014). On top of this, performance across countries appears to be mixed: the same kind of project may work properly in some countries (e.g., hospitals in the UK) and fail in others. From these considerations, two main questions arise. What mechanisms could explain the observed heterogeneity of PPP performance across countries and in time? Related to the previous question: how PPP performance should be measured?

Contributions to different strands of economics and management literature are trying to delve into the first issue. The institutional setting certainly plays a role to explain heterogeneity of PPP performance in time and across countries. However, recent contributions have highlighted how, within the same institutional setting, organizational schemes based on sound business models are important keys toward PPPs' success (Villani et al. 2017). Other contributions have pointed out the role of agency problems and transaction costs within the consortium of private firms (Hoppe et al. 2013; Greco 2015). Further issues that still need improved understanding are related to renegotiations, which tend to reduce efficiency gains from PPPs (Engel et al. 2014), and are related to the lack of flexibility that, over long time spans and in particular sectors, may affect specifically PPPs (Martimort and Straub 2016).

As highlighted in the previous discussion, practitioners' debate has conveyed the idea that PPPs may be beneficial because of the market discipline they are able to introduce in public infrastructure planning. Following this argument, the mere involvement of private firms is a signal that a public infrastructure is viable from the economic and financial point of view. Clearly, this is a wrong idea: government assessment needs to precede the decision to undertake PPPs (Engel et al. 2014). However, this failed argument bring us to the very important issue of *how to measure the ex post performance of PPPs*. This issue is crucial both to assess theoretical predictions of the economics and management literature and to run policy evaluation studies. Since panel data econometrics is hardly useful in this case, because of the large heterogeneity across PPP contracts – even in the same sector, case studies or alternative econometric techniques (e.g., methods based on synthetic counterfactual) are required. This is surely an avenue of future empirical research that is worth to pursue.

Cross-References

- ▶ Externalities
- ▶ Franchise
- ▶ Governance
- ▶ Government
- ▶ Incomplete Contracts
- ▶ Institutional Economics
- ▶ Organization

- ▶ Political Economy
- ▶ Privatization
- ▶ Public Goods
- ▶ State-Owned Enterprises

References

Bennett J, Iossa E (2006) Building and managing facilities for public services. J Public Econ 90(10):2143–2160

Besley T, Ghatak M (2001) Government versus private ownership of public goods. Q J Econ 116(4): 1343–1372

Bezançon X (2005) Histoire du Droit Concessionnaire en France. Entrep Hist 38(1):24–54

Burger P, Hawkesworth I (2011) How to attain value for money: comparing PPP and traditional infrastructure public procurement. OECD J Budg 2011(1):1–56

Buso M, Marty F, Tran PT (2017) Public-private partnerships from budget constraints: looking for debt hiding? Int J Ind Organ 51(C):56–84

Engel E, Fischer RD, Galetovic A (1997) Highway franchising: pitfalls and opportunities. Am Econ Rev 87(2):68–72

Engel E, Fischer RD, Galetovic A (2013) The basic public finance of public-private partnerships. J Eur Econ Assoc 11(1):83–111

Engel E, Fischer RD, Galetovic A (2014) The economics of public-private partnerships. A basic guide. Cambridge University Press, New York

Greco L (2015) Imperfect bundling in public-private partnerships. J Public Econ Theory 17(1):136–146

Hart O (2003) Incomplete contracts and public ownership: remarks, and an application to public-private partnerships. Econ J 113:C69–C76

Hoppe EI, Kusterer DJ, Schmitz PW (2013) Public-private partnerships versus traditional procurement: an experimental investigation. J Econ Behav Organ 89(C):145–166

Iossa E, Martimort D (2015) The simple microeconomics of public-private partnerships. J Public Econ Theory 17(1):4–48

Kivleniece I, Quélin BV (2012) Creating and capturing value in public-private ties: a private actor's perspective. Acad Manag Rev 37(2):272–299

Martimort D, Pouyet J (2008) To build or not to build: normative and positive theories of public-private partnerships. Int J Ind Organ 26(2):393–411

Martimort D, Straub S (2016) How to design infrastructure contracts in a warming world: a critical appraisal of public-private partnerships. Int Econ Rev 57(1):61–88

Maskin E, Tirole J (2008) Public-private partnerships and government spending limits. Int J Ind Organ 26(2):412–420

Perkmann M, Schildt H (2015) Open data partnerships between firms and universities: the role of boundary organizations. Res Policy 44:1133–1143

Saussier S (2015) Économie des Partenariats Public-Privé. Développements Théoriques et Empiriques. De Boeck

Schmitz PW (2015) Government versus private ownership of public goods: the role of bargaining frictions. J Public Econ 132(C):23–31

Villani E, Greco L, Phillips N (2017) Understanding value creation in public-private partnerships: a comparative case study. J Manag Stud 54(6):876–905