The curse of antitrust facing bilateral monopoly: Is regulation hopeless?

Andrey Shastitko\textsuperscript{a,b}, Claude Ménard\textsuperscript{c}, Natalia Pavlova\textsuperscript{a,b,}\textsuperscript{*}

\textsuperscript{a} Russian Presidential Academy of National Economy and Public Administration, Moscow, Russia
\textsuperscript{b} Lomonosov Moscow State University, Moscow, Russia
\textsuperscript{c} Centre d’Economie de la Sorbonne, Université Paris 1 Panthéon-Sorbonne, Paris, France

Abstract

This paper is about the challenges that antitrust authorities face when dealing with bilateral monopolies. The curse of antitrust refers to traps threatening the efficient applicability of antitrust policies in these situations. Standard theories diverge about the attainability of equilibrium under bilateral monopolies but share skepticism about its efficiency if it ever exists. We suggest a different approach, based on transaction cost theory. First, since bilateral monopolies often develop in the upper segment of value chains, misalignment between parties may generate negative externalities. Second, if parties reach an agreement, the impact of the governance mechanism implemented must be assessed beyond the usual parameters of prices and quantities. Indeed, the risk of negative externalities in the absence of appropriate governance increases dramatically when “critical transactions” are at stake. With vertical integration prohibited, second-best alternatives in which antitrust authorities leave room for innovative hybrid governance may allow internalizing externalities while avoiding high switching costs.

Keywords: antitrust policy, bilateral monopoly, hybrid governance, switching costs, critical transactions, bargaining power.

JEL classification: D4, K21.

1. Introduction

The application of antitrust legislation on markets characterized by monopoly or monopsony is relatively routine and predictable. However, a market structure in which two standalone but interdependent firms, reciprocally acting as buyer and seller, engage in monopolistic activities is clearly non-standard, although not unusual. The specificity of this situation should be stressed because

\textsuperscript{*} E-mail address: nspavlova@econ.msu.ru

© 2018 Non-profit partnership “Voprosy Ekonomiki”. This is an open access article distributed under the terms of the Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0).
agreements between parties, notwithstanding the parameters of legally binding contracts, are characterized by the formalization of long-term, non-standard, complex interactions that involve highly specific assets. In this context, intervention by a third party can be regarded only as an extreme measure, with the challenging issue raised by the need to maintain the continuity of transactions between partners.

Indeed, if a firm dominates the market as a seller but has its power counterbalanced by a single buyer, what is the point for antitrust authorities to intervene? Conditional to the absence of collusion, does not the resulting interaction lead to competitive conditions? If it is so, implementing the general competition laws should suffice.

However, this reasoning relies on implicit assumptions that can support misleading analysis as well as wrong normative conclusions and regulation. This paper deals with this issue. It intends to show the possible negative effects of applying standard antitrust legislation to bilateral monopolies when significant switching costs are at stake.¹

In Section 2, we briefly review the literature on bilateral monopoly, with specific attention to issues of antitrust regulation when interdependent firms face significant switching costs due to the specificity of assets involved in the relationship and the critical nature of the resulting transactions. In this context, attention is paid to the choice of governance mechanisms and the possibility of making errors in that choice. Section 3 develops the underlying model, stretching the standard approach to shed light on this issue of adequate governance mechanisms when critical transactions shape the relationship in a bilateral monopoly, generating risk of negative externalities. Section 4 illustrates through three examples from the Russian antitrust practices in three different industries within which interdependent relationships resulted from the privatization process of the 1990s. Emphasis is on the challenges due to contractual arrangements built between firms that can face high switching costs if they cannot secure their critical transactions. Section 5 discusses, with some details, the possible role of competition authorities in that context. We conclude with some lessons regarding the need for innovative policies in the context of bilateral monopolies.

2. Some indications from the literature

Notwithstanding some important contributions, discussions about the foundation and effects of antitrust policies regarding bilateral monopolies remain relatively sparse in the economic literature.

2.1. Initial syntheses

A key step in that respect is the survey by Machlup and Taber (1960), focusing on the conditions of equilibrium in bilateral monopolies from Cournot to the 1950s. In their paper, they identify three possible situations in the case of markets for intermediate goods.

¹ See Shastitko and Pavlova (2017a) for an initial development in this direction.
1. Both price and quantity of an intermediate product are unique in equilibrium, with the quantity being less than what it would be under vertical integration so that the consumer price in the final market is higher.

2. Both quantity and price are not uniquely determined but may vary within a certain range. Again, the quantity at equilibrium will remain lower than with vertical integration.

3. Quantity at equilibrium corresponds to that of vertical integration. In this configuration, the price in the final market is unique and not exceeding the one that occurs under vertical integration. However, uncertainty remains with respect to the price in the market of intermediate goods, a price determined through negotiations. According to the model’s assumptions, the resulting equilibrium will affect only the distribution of profit between parties in the intermediate market, with no impact on the final market.

According to Machlup and Taber (1960, pp. 110–111), the third option is likely to be the most relevant, the two others resulting from the myopic view of economists who look at bilateral monopoly as a complication of monopoly or monopsony models rather than as a situation commanding its own model. However, the authors also suggest that the first two approaches are quite acceptable to describe certain markets (Machlup and Taber, 1960, p. 112). This fluctuation in the analysis points at the underlying problem, which is about the adequate arrangements when it comes to organizing transactions among firms engaged in long-term relationships involving highly specific assets and facing high switching costs. We call such transactions “critical” in that they condition the continuity of economic activities under review.

The Machlup and Taber overview prevailed for quite some time. Pushing further the analysis, some contributions explored the conditions and/or mechanism under which a unique equilibrium could be ensured. Foldes (1964) took into account the intertemporal preferences of parties and threats of delays in exchange. Blair and Kaserman (1987) considered the possibility of a solution through a contract with a pricing formula that would meet specific criteria. Dobbs and Hill (1993) focused on take-or-pay contracts as a mechanism to reduce the uncertainty zone. It is in this context of uncertain solutions that some authors (Campbell, 2007; Baker et al., 2008; Blair and DePasquale, 2011) explored under what conditions antitrust authorities could authorize vertical mergers, without a general agreement on the issue.

Blair et al. (1989) investigated the treatment of bilateral monopolies in textbooks and found that the third approach prevailed. They agreed that this was likely the correct solution and traced it back to Bowley (1928), who argued that the optimum came from parties maximizing profits, which would determine a quantity corresponding to the maximum public well-being on the intermediate market, while price would be established through negotiations between the monopoly and the monopsony. It should already be noted that such negotiations define a governance mechanism that requires the existence of rules accepted by the parties, an important issue in the “Coasian” perspective about ways to internalize externalities.

---

2 In Russia, Galperin et al. (2004), amongst others, proposed a bilateral monopoly model based on the second approach.
2.2. Recent developments

Notwithstanding these valuable contributions, interest in bilateral monopolies faded away, with rare recent contributions shifting the attention to relatively peripheral issues with respect to the theoretical core, e.g. the privatization of state-owned bilateral monopolies (Bose and Gupta, 2013) or social responsibility in bilateral monopolies (Brand and Grothe, 2015; Goering, 2014).

This lack of interest for bilateral monopoly is quite amazing if we consider the significance of such market arrangements in the economy and the possible externalities they can generate, the magnitude of which depends on the outcome of negotiations to allocate sectoral profits between parties. The outcome itself is a tributary of the modalities embedded in a contract or of the adjustment mechanism implemented to adapt to the changing dynamics between interdependent firms.

In that respect, discussions coming out of the Coase theorem, regarding how to internalize externalities and the perspectives it opens, if rightly interpreted, on the various modalities that can organize negotiations and exchange under bilateral monopolies, are relevant here. In the now standard interpretation of the Coase theorem, which assumes zero transaction costs, bilateral monopolies are compatible with competition, as illustrated by the famous example of the relationship between farmers and ranchers (Coase, 1960, 1988; see also Bowley, 1928; Shastitko, 2010).

However, the resulting equilibrium does not presume what will be the allocation of the overall gains for parties involved (the “income” effects). This is where the possibility of externalities (positive or negative) enters into the picture. If transaction costs are positive, due to redistribution effects and/or the presence of externalities, the resulting deviation from the Pareto optimal volume of transactions means that costs and benefits are not fully reflected in the price system or in other conditions of the contractual relations. This point is especially important when the bilateral relationship involves rigid interdependence between the components of the technological chain, thus determining the existence of critical transactions, that is: transactions that condition the continuity of the economic activity at stake and that command the implementation of carefully designed governance mechanism (Künneke et al., 2010).

2.3. A different perspective

The perspective thus opened, which is in line with the analytical developments of transaction cost economics (Williamson, 1996; Menard and Shirley, 2008), raises the issue of an effective application of competition policies to the case of bilateral monopoly. Indeed this type of market arrangement can be interpreted in the context of what Williamson (1985, ch. 2) identified as the fundamental transformation, according to which relationships that are competitive ex ante transform ex post into interdependent relationships between a handful of parties.

---

3 It must be noted that Coase (1991) strongly opposed this overstated interpretation that focuses on the “blackboard” assumption of zero transaction costs (see his Nobel lecture).

4 This paper does not discuss the disputable questions of antitrust policies in general, as raised, for example, by the Austrian school of economics.
due to significant switching costs. Joskow (2002) vividly illustrated the essence of this antitrust curse: implementing antitrust legislation to promote competition under these conditions may actually become a hindrance to contractual relations between the buyer and the seller, generating transaction costs that have negative impact on welfare and may even challenge the very existence of otherwise beneficial transactions. Such intervention actually provides a typical example of Type-I errors in law enforcement (illustrations are provided in Joskow, 2002; Manne and Wright, 2010; Shastitko, 2012b, 2013; Avdasheva and Kryuchkova, 2015).

The problem, thus pointed out, also provides a basis to discuss issues related to erroneous choices of governance mechanisms for such critical transactions, choices which can prevent selecting organizational arrangements that would allow cost-minimizing solutions (Williamson, 1985, ch. 3 & 14; Shastitko, 2016; Shastitko and Menard, 2017).

This issue is particularly challenging when transactions organized within a bilateral monopoly are interrelated with other transactions. One possible manifestation of this interrelation is the double markup and the resulting negative vertical externality under the so-called monopoly pyramid, as when there is one seller and one buyer on the intermediate market and one seller on the final market, with the intermediate product seller benefiting from bargaining power. The model also presupposes constant returns to scale and competition amongst buyers of the final good (Rozanova and Avdasheva, 1998; Avdasheva, 2000; Avdasheva and Dzagurova, 2010; Shastitko, 2010). Negative vertical externality translates into a price on the final market that is higher than what it would be with an integrated firm, while the volume of transactions is reduced, so that there are significant net losses in well-being.

However, negative externalities coming out of the double markup do not exhaust the problems raised by bilateral monopolies. The multiple equilibria associated with these organizational structures, at least when they are not regulated, also involve a technological dimension. Indeed, if the production processes that develop under bilateral monopoly are characterized by continuity and/or by rigid technological interdependencies, critical transaction features and critical technical functions which they must satisfy, as observed in network infrastructures (Künneke et al., 2010), may generate important switching costs in other industries as well.

In that respect, differences between monopoly power and bargaining power, the subject of an increasingly active discussion amongst economists and lawyers specializing in antitrust policies (Shastitko and Pavlova, 2017a; Lianos and Lombardi, 2016) provides elements to discuss alternative governance mechanisms as well as regulatory involvement of antitrust authorities in organizing transactions between two interdependent entities.

The challenge in that respect (as in so many other situations) is how to balance the principle of non-interference in contractual relationships amongst firms with the simultaneous protection of competition and freedom of individual market players, on the one hand, and the protection of the interests of these same players when they are involved in a dispute, on the other hand. The first aspect refers essentially to the presumption of the inexpediency of third-party intervention in contractual relationships: regulatory actions should, accordingly, be restricted to the general rules regulating competition. However, when there is high resource specificity and continuity of interaction involved, so that opposing interests of
the parties might become a challenge and negative externalities might result, specific interferences may be required. Interference can take two forms: (1) settling disputes through pre-court procedures and mechanisms designed to solve private conflicts (e.g. arbitration); or (2) addressing antitrust authorities endowed with quasi-court functions, but exclusively as enforcers using tools and mechanisms (including tactics to “force the parties to make peace”) that can secure the stability of the contractual relationships when parties must face changing circumstances while maintaining continuity through repeated transactions.

3. Revisiting bilateral monopoly’s model

How do these considerations connect to the standard representation of bilateral monopoly and how far can we go in stretching this model to include the issues thus raised? To explore these questions, we start with a short reminder of the classical representation of bilateral monopoly that we then extend to explore policy issues when: (1) transaction costs are significant; (2) the distribution of bargaining power is asymmetric; (3) and/or negative externalities need to be taken into account. We also point out the numerous limitations such modeling faces when it comes to formulating sound normative competition policies.

3.1. Bilateral monopoly with zero transaction costs

The fictitious assumption of zero transaction costs (Coase, 1988; 1991) allows pointing out the Pareto-optimal allocation of resources regardless of their initial distribution and of possible asymmetries in bargaining power. Blair et al. (1989) provide an illustration of bilateral monopoly operating under such conditions, with a volume of transaction corresponding to a competitive one while prices might be below or above the one that would be reached under purely competitive conditions (Fig. 1).

![Fig. 1. Equilibrium in the standard bilateral monopoly model.](image-url)

*Source: Blair et al. (1989, p. 832).*
Let us assume that there is only one seller and one buyer operating in the market for intermediate goods \( x \), with the buyer being also a monopolist on the next processing stage that, for convenience, we consider as the final market. Quantity and price are determined through negotiations (at zero costs) between the two parties. Following Blair et al. (1989), we denote:

- \( C(x) \), the aggregate production costs for the good \( x \);
- \( Y \), an input required to produce the final good and supplied at the constant price \( p_y \), due to the monopolistic position of the supplier with respect to its buyer;
- \( Q \), the quantity of the final good \( x \) produced, conditional to the provision of \( y \);
- \( p_x \), the price of good \( x \);
- \( P = P(Q) \), the inverse demand function for the final product.

Under vertical integration, the profit of the integrated firm would be:

\[
\pi^I = P[Q(x,y)] Q(x,y) - C(x) - p_y y. \tag{1}
\]

At the equilibrium, the production of \( x \) and usage of \( y \) would be determined by the following maximizing conditions, under which the marginal product of each intermediate resource is equal to the marginal cost of its production/purchase:

\[
\left( P + Q \frac{dP}{dQ} \right) \frac{\partial Q}{\partial x} = \frac{dC}{dx}, \tag{2}
\]

\[
\left( P + Q \frac{dP}{dQ} \right) \frac{\partial Q}{\partial y} = p_y. \tag{3}
\]

At this point, the monopolist who is also a monopsonist reaches the maximum aggregate profit.

But what happens if the firms are not integrated and must determine the equilibrium parameters through negotiations? Sticking to a simplified representation, three situations are possible: advantages in bargaining power are on the supplier side or on the buyer side or there are no advantages for either party. Note that, in our view, an advantage in negotiation for one party does not necessarily presume competitive behavior by the second party, a fundamental difference from the interpretation adopted in Gould and Ferguson (1980) or Galperin et al. (2004), amongst others. Moreover, we do not consider, at this point, advantages that could come from the intervention of the antitrust authority to prohibit anticompetitive behavior.\(^5\)

Under these specifications, the monopolist’s profit (“Upstream”, noted \( u \)) will be:

\[
\pi^u = p_x x - C(x), \tag{4}
\]

while the profit of the monopsonist, simultaneously a monopolist in the market for final goods (“Downstream”, noted \( d \)) is:

\[
\pi^d = P[Q(x,y)] Q(x,y) - p_x x - p_y y. \tag{5}
\]

\(^5\) In this paper, we do not consider the sources of bargaining power advantages (for that aspect, see Shastitko and Pavlova, 2017b). As for the potential role of antitrust authorities in the context discussed here, see Shastitko (2017).
Let us now consider the three stylized situations regarding the position of the parties in the bargaining process.

**Case 1. Bargaining advantage for the monopolist.**

In this situation, the monopolist profit is maximized, while participation to the exchange on the side of the monopsonist is conditional to the existence of a profit equal or superior to zero. As the monopolist dominates the negotiation, the monopsonist profit will actually be zero. Under the usual maximization conditions, the resulting price for intermediate goods is:

\[ p_x = \frac{\{P[Q(x,y)] Q(x,y) - p_y y\}}{x}. \] (6)

The monopolist profit is:

\[ \pi^u = P[Q(x,y)] Q(x,y) - C(x) - p_y y = \pi^I. \] (7)

Thus, the equilibrium quantity on the market for intermediate goods is achieved at a level that maximizes the aggregate profit which is entirely captured by the monopolist at the equilibrium price \( p_3^u \) in Fig. 1.

**Case 2. Bargaining advantage for the monopsonist.**

Symmetrically, the monopsonist profit is maximized subject to the condition that the monopolist profit is equal or superior to zero. The price on the market for intermediate goods then becomes:

\[ p_x = \frac{C(x)}{x}. \] (8)

The monopsonist profit is:

\[ \pi^d = P[Q(x,y)] Q(x,y) - C(x) - p_y y = \pi^I. \] (9)

No surprise if, in this situation symmetrical to the previous one, the equilibrium quantity on the intermediate goods market ensures the maximum aggregate profit, which is now fully captured by the monopsonist \( p_3^d \) in Fig. 1.

**Case 3. No bargaining advantage for either party.**

In the absence of any bargaining advantage, the result of the negotiation is determined by a point on the contract curve in the space \((x, p_x)\) which in turn is determined via the intersection points of the isoprofit curves for the monopolist and monopsonist. In Blair et al. (1989, pp. 837–840), the contract curve is a vertical line stemming from \(x_3\), which corresponds to the volume that maximizes the aggregate profit. The boundaries of the contract curve are determined by the possible actions the firms can undertake in the event of failed negotiations, the condition of non-negative profit for each party limiting the contract curve within the range delineated by the prices \(p_3^u\) and \(p_3^d\) (Fig. 1).

Points \( A (x_1, p_1) \) and \( B (x_1, p_1) \) in Fig. 1 correspond to the usual solutions for models of pure monopoly and pure monopsony, respectively. However, since each
of these solutions assumes competition on the other side of the market, we do not consider these solutions as relevant for the analysis of a bilateral monopoly.

3.2. Bilateral monopoly with positive transaction costs and externalities.

Indeed, the previous model reflects only part of the picture, which may be enhanced with the double markup model, leading to a negative externality along the vertical relationship due to the deviation from the Pareto optimal volume of transaction on the market for intermediate goods. It must be remembered that the double markup model assumes asymmetry in the bargaining power on that market in favor of the seller, while production is characterized by continuous returns to scale. There is also another important prerequisite, which is the existence of market power on the adjacent market (usually down the technological chain), hence the other name also used to coin the double markup model, the monopoly pyramid.

Using the same variables as in the previous section, we additionally assume, for sake of simplicity, that the production of one unit of the final goods requires only one unit of the intermediate good and that its price is equal to the average production cost of the final good. Then, equations for profits of the firms on both sides of the market are:

$$\pi_u = p_x x - C(x),$$
$$\pi_d = P(x)x - p_x x. \quad (10)$$

If the inverse function of demand for the final good is linear ($D_Q = C_T$ in Fig. 1, with $C_T$ the cost of transforming one unit of $x$ into one unit of $Q$), with $P(x) = a - bx$, then the inverse function of demand for the intermediate product ($D_x$) is given by the formula $p_x(x) = a - 2bx$.

With a monopoly on the market for intermediate goods, the quantity $x$ at equilibrium is:

$$x^* = \frac{a - dC}{dx} = \frac{4a - dC}{4b}, \quad (12)$$

which is also equal to $Q^*$, the equilibrium on the market for final goods. Under these conditions, the profits of the upstream and downstream parties become equal to:

$$\pi_u = \left(\frac{a - dC}{dx}\right)^2 \cdot 8b, \quad (13)$$
$$\pi_d = \left(\frac{a - dC}{dx}\right)^2 \cdot 16b, \quad (14)$$

while the profit of an integrated firm would be:

$$\pi_I = \left(\frac{a - dC}{dx}\right)^2 \cdot 4b. \quad (15)$$
This profit may also be interpreted as resulting from a monopolized market for final goods while competition exists on the market for intermediate goods. In both cases, the profit of an integrated firm is higher than the aggregate profit of the parties in the double markup situation.

Thus, the result from a double markup situation is one of lower sales and lower gains on the market for final goods even when it is assumed that there is a monopoly on that market and competition on the market for intermediate goods. This difference in the aggregate gain can be understood as costs not reflected in the terms of the agreement between the buyer and the seller on the market for intermediate goods. An alternative scenario would be that the buyer holds the bargaining power. We then fall back to the standard model of multiple equilibria with respect to both prices and quantities transacted.

However, a more fundamental question raised by the lower performance of the bilateral monopoly operating under the conditions described above in comparison to what it would be for an integrated firm is the following: what prevents the parties from forming an agreement, if they are legally prohibited to merge, so as to capture at least part of the net losses? Or what prevents the parties from improving an existing binding agreement if there is one? In other words, how can the distribution of gains from voluntary exchanges in a situation of positive transaction costs affect the coordination characteristics of the exchange (that is: the equilibrium conditions)? Seventy years ago, Samuelson gave a brief answer to this question: “...for many types of bilateral monopolies, the ultimate equilibrium may be achieved beyond the contract curve” (Samuelson, 1947, p. 238). However, the explanation for this deviation, if we assume rational actors, remains unanswered.

3.3. Bilateral monopoly with positive transaction costs and structural alternatives for internalizing externalities.

To explore this issue, an important condition must be met: there exists at least one structural alternative for internalizing externalities that outperforms the status quo. In the case of the double markup model, this means the existence of an equilibrium with quantities lower than those traded under zero-transaction-costs but higher than if the externality/ies were not internalized.

Note that there is already one such option investigated in the literature: vertical integration. To go further, we make an additional assumption: transaction costs for internalizing externalities through vertical integration are significant. Moreover, there is the goal of preserving competition in adjacent markets. In this context, we shall argue that a hybrid governance mechanism is the best way to deal with these externalities while limiting switching costs if this mechanism combines simultaneously: (1) the possibility of joint planning and adjustment to changing circumstances; (2) the preservation of parties’ residual rights, so that unified control is excluded.

In order to take on board the role of switching costs as a key factor in the interaction of the firms and their capacity to reach a Pareto-improving solution, specific information about the specific circumstances of this interaction must be available. This requirement is very demanding and generates methodological complexities that are challenges for antitrust authorities. The issue may end
up in a “prisoner’s dilemma” type of situation in which the dominant strategy of the two firms might be to address the antitrust authority for arbitration, notwithstanding the costs involved (particularly the costs of collecting and processing the information through committees and the costs of lawyers and experts (Shastitko, 2017).

Nevertheless, there might be a possibility for anti-trust authorities to “regulate” the contractual relationships within a bilateral monopoly that would allow Pareto-improvement, conditional to: (1) requirements for disclosing information and ensuring compliance by the parties; (2) the choice of collateral taking into account the “ugly princess” problem;⁶ (3) the choice of a guarantor, able to verify information for agreements resulting from a compromise; and/or (4) the choice of a governance mechanism that matches the characteristics determining the criticality of the transaction at stake. In this context, whether the antitrust authority refuses to consider the case between parties to the bilateral monopoly or adopts a pro-active position can be regarded as the choice between competing governance mechanisms with respect to dispute resolution. However, stating which solution is preferable without considering the specific context, while sticking to the ultimate goal of preserving competition, leads to uncertainty regarding law enforcement. Two questions then arise. (1) Are the resulting transaction costs too high? (2) Does this uncertainty affect incentives of parties to build contractual relations based on coordination rather than engaging in arm’s length relations over distribution that would make it unpredictable which party (if any) would benefit? When costs of legal uncertainty are high and reduce incentives to form coordination agreements, choosing one solution (remaining neutral vs. being pro-active along specific guideline) as the “standard” to enforce the law and solve disputes might be superior to leaving pure discretionary power in the hands of the regulator as it would at least reduce legal uncertainty.

Indeed, an important aspect policy-wise is the actual possibility for the antitrust authority to ensure that parties to contractual relations in a situation of bilateral monopoly have incentives to satisfy the requirements mentioned above (information disclosure, exchange of collateral, guarantor) so as to reach an efficient governance mechanism. Solutions will then be found directly by the parties, without third-party interference regarding their specific conditions. In other words, it is likely that if the conditions mentioned are met, there will be economies on transaction costs, thanks to the creation by parties themselves of the necessary mechanism for redistributing the aggregate gain without significant damage to the Pareto-optimal quantities.

4. Three illustrative cases

The Russian antitrust authority has been repeatedly confronted with cases of bilateral monopolies along the lines described above. It is so, largely because of the structural changes of the Russian economy in the 1990s, characterized

⁶ In transaction costs theory (Williamson, 1983), the “ugly princess” refers to the search for the optimal hostage to mitigate the risk of opportunistic behavior on the part of the contractor. The expression comes from the example of a king who would have to give one of his equally beloved daughters as a hostage to a perpetrator. The best option for him is then to give up the “ugly” one since the perpetrator is less likely to hold her.
by the de-integration and redistribution of technologically interconnected assets amongst what became ultimately independent rights holders. In other terms, interdependent assets previously integrated within a firm were redistributed amongst different, independent firms created in the transition process. The resulting problem, typically the creation of bilateral monopolies, can hardly be considered unique to Russia. However, the diversity and scale at which it happened in a short period in this country have generated repeated disputes that have mobilized regulatory authorities.

The problem obviously came out of the rapid, systemic transformation from a centrally planned economy to a market economy. Very early on, Blanchard and Kremer (1997) pointed out that this transformation, with its focus on privatization, led to disorganization (or an aggravation of the disorganization that already existed before this systemic shock) largely due to significant switching costs. Consequences were later qualified as the Pikalev syndrome (Shastitko, 2012a), which identified how conflicting relationships amongst interdependent firms create acute social problems tentatively resolved through antitrust legislation, not because it is best suited to this task, but because it could provide short term answers.

Using the theoretical lenses provided by our discussion of the previous section, we now turn to three recent cases that ended in nearly divergent decisions by the antitrust authority and that illustrate the puzzle it faces. More precisely, our analysis refers to the concept of critical transactions and the mechanisms of governance they command (Künneke et al., 2010). Although this concept was initially introduced to understand the specific structure of network infrastructures, we consider it is relevant for understanding the problem that the antitrust authority faced in these cases, that is: the distortions introduced in the privatization process through the establishment of bilateral monopolies without the governance mechanisms that could secure the degree of control and the capacity to adjust that transactions at stake would require (Künneke et al., 2010, p. 502). One additional problem, in the Russian context, is that reforming the industry, for example through a redistribution of assets that would restore unified governance over such critical transactions, was not an option for the antitrust authority. Since there is a high probability that the existing distribution of assets will remain the same, a given restriction imposed by the institutional environment, the issue becomes that of exploring second-best solutions, their comparative advantages, drawbacks and associated risks.

4.1. MMK—RUSAL

The first case deserves attention because of the non-standard approach taken by the Russian Federal Antimonopoly Service (FAS from now on) in its discussion of a bilateral monopoly problem in relation to the need to protect competition.

RUSAL is the dominant buyer of B1 and V melted coal tar pitch for electrodes. This pitch is a raw material for anodes used in the production of aluminium. On the other hand, Magnitogorsk Iron and Steel Works (MMK), is a major player in the pitch market, tar pitch being a by-product of its large coke production. In 2015, the companies failed to reach an agreement concerning the main clause of the contract—the price, and MMK went to FAS, accusing RUSAL of imposing
onerous contractual terms thanks to its market power. Having examined the case, FAS confirmed the abuse of its dominant position by RUSAL. However, FAS also found indications that the inability of MMK to promptly sell pitch to another buyer led to an emergency suspension of its pitch production. Simultaneously, RUSAL, facing a dead-end in its negotiations with MMK, switched to foreign producers of pitch, an alternative that was far from cheap! Taking into account the critical impact resulting from the disrupted transactions between MMK and RUSAL, that is: the possibility of shutting down coke production in Russia (at least that was the argument of MMK) and of a potentially major environmental impact (tar pitch is a very toxic product and MMK did not have storage capacities large enough to dispose of its production), FAS pressured the firms to reach an agreement, and the contract was finally signed along the terms proposed by RUSAL.

What is particularly interesting here is the underlying logic that led to this agreement. The technological interdependence of two dominant firms (the seller and the buyer) determined the existence of critical transactions between them. Failure to reach an agreement on supply terms through “ordinary” negotiations led to highly unfavorable consequences for both and could have threatened the overall production of aluminium and metal coke in Russia. Confronted by this situation, FAS warned the two firms, urging them to reach an agreement and using the threat of direct interference to pressure them to negotiate. The firms understood the warning and reached an agreement that ended the antitrust case.

The information provided by the parties while the case was under review showed that there would have been significant switching costs for both entities if they had to deal with alternative partners and pointed out that the specific assets involved—the resource and its associated technology—determined the critical nature of the transactions at stake, which required monitoring through mechanisms of governance that were absent or improperly designed. In pressuring parties to directly negotiate an agreement, FAS encouraged the companies to engage into a governance mechanism—direct negotiation between parties in a bilateral structure—that does not correspond to the standard competition policy approach but that secured the survival of the interdependent activities of the parties. Clearly, this was motivated by consideration of the negative externalities that a breach of transactions would have introduced.

4.2. Sayanskhimplant—Rosneft

At the end of 2016, the FAS website published information noting that, in 2015, there was a contract between Rosneft, a supplier of polyethylene, and Sayanskhimplant, a PVC producer. Polyethylene is the main raw material used by Sayanskhimplant. This input was supplied through a pipeline connecting the two facilities and, for Sayanskhimplant, there was no viable alternative for its provision. Sayanskhimplant complained to the Irkutsk FAS Administration alleging that, at the end of 2015, Rosneft had voluntarily reduced its supply of ethylene to

---

7 The problem comes from the difficulties of transporting liquid coal pitch over great distances by land and by sea (see https://www.vedomosti.ru/business/articles/2015/10/29/614767-fas-zapodozrila-rusal-inezhenii-tsen-nasire).
push prices upward. As a result, Sayanskhimplant had to stop its activity, because the quantity delivered was not large enough to secure the minimum required technically, so that there was a threat of full stoppage of PVC production.

The Irkutsk FAS branch initiated prosecution against Rosneft. The case was then transferred to the central FAS for consideration about whether the Law on Competition Protection was violated. Based on an analysis of the ethylene market, FAS established that a bilateral monopoly had been formed for this market, in which the market power of the seller was counterbalanced by the market power of the buyer. Following this investigation, FAS decided that the conflict was, in essence, a purely private economic dispute since the supply of ethylene was based on a voluntary contract. Hence, notwithstanding the admitted bilateral monopoly conditions on this market, the Sayanskhimplant—Rosneft case was considered a dispute regarding strictly the breach of obligations under the contract, falling under civil law. Following this decision and to avoid the disruption and transaction costs of going to courts, the firms negotiated and, in December 2016, an amicable agreement between the parties was approved by the Arbitration Court of Irkutsk Region.

It is noticeable that, in this case, there was no assessment of the respective bargaining power of the parties, nor any consideration for the possibility of negative externalities along the technological chain coming out of this market structure. FAS simply discharged the case by qualifying it as a private dispute, which motivated the firms to find a solution that would avoid the risky process of addressing the legal system.

4.3. Uralkali—Solikamsk Magnesium Plant

A third case is that of carnallite. The enriched carnallite market in the Russian Federation has a structure close to that of a bilateral monopoly, with Uralkali the only supplier and Solikamsk Magnesium Plant the largest out of three buyers. Enriched carnallite is a key resource for the latter. Its supply is delivered in accordance with Uralkali’s marketing policy, which claims to provide access to enriched carnallite at a non-discriminatory price for all buyers.

In 2014, FAS Russia received instructions from the Russian Federation Government to check if the price established by Uralkali for the delivery of carnallite to Solikamsk Magnesium Plant was “reasonable”. FAS initiated an investigation, which considered, amongst other factors, if it had been appropriate to include an investment component in the price for enriched carnallite in order to cover the costs for Uralkali to expand and upgrade its production facilities. One central argument to this solution was about the “cooperative” nature of the investment, since its benefits would largely accrue to the buyer, allowing Solikamsk Magnesium Plant to better satisfy a growing demand, whereas for Uralkali, the realization of this project would not necessarily lead to lower costs. Upon considering the case, FAS Russia concluded that the price established in accordance with Uralkali’s marketing policy did not exceed the costs and profits necessary to produce and sell this product. Hence, the case was closed.

An important characteristic of this case is therefore that FAS played the role of a mechanism to investigate and possibly settle a dispute at the initiative of the Russian Government. Based on the information available, the case differs
from that of MMK vs. RUSAL in two aspects. First, there was no threat to interrupt transactions between the parties. Although enriched carnallite is a critical input for Solikamsk Magnesium Plant, examination of the “reasonable” character of the price was carried out at the initiative of the Russian Government. While the investigation by FAS was going on, no contract was signed. An agreement was reached and signed at the beginning of 2016 for six months and at a price incorporating the extra-investment costs of the supplier.\footnote{https://www.kommersant.ru/doc/2983087} Second, the possibility that the relationship between the two parties defined a bilateral monopoly and its possible consideration in the decision of the regulatory authorities was apparently never raised as a significant issue.

3.4. Some lessons from these cases

One basic principle that should guide the development of an effective legal system is to provide incentives for economic actors to use resources to create value rather than rents. Important criteria in that respect are consistency between rules and their uncontroversial enforcement. The three examples we reviewed raise the question of whether regulation enforcers have legal grounds solid enough to make, almost simultaneously, decisions that are orthogonal to each other while avoiding being purely discretionary.

Our hypothesis is that if antitrust authorities have legal grounds to intervene, there are specific characteristics they should take into consideration in enforcing the law. Of particular importance in that respect is the understanding of the potential consequences and side effects of their decision on the relationship between buyers and sellers as well as on third parties, so that negative externalities need to be taken into account. A lesson from the examples, briefly reviewed above, is that cautious decision-making based on a clear assessment of the impact of decisions can be crucial when dealing with bilateral monopolies.

In the case of MMK vs. RUSAL, a critical transaction was at stake and the impact of a disruption in the provision of pitch was estimated unfavorably to MMK. The bilateral relationship between the companies was discussed in the context of simultaneously protecting competition. Antitrust authorities used the threat of direct intervention to push parties to rely on a “private” mechanism of governance, a direct negotiation to reach a contract, notwithstanding the collusive dimension it could involve.

The case of Rosneft vs. Sayanskhimplast (and of Uralkali vs. Solikamsk Magnesium Plant in that respect) differs in its conditions. Ethylene, which is produced by Rosneft through a subsidiary (Angarsk Polymer Plant, APP), is not a strategic product for Rosneft while continuous disruption in its supply could challenge the very existence of Sayanskhimplast. This put one party at risk of collapsing, as it became apparent when a major disruption due to malfunctioning equipment at APP (in 2016) pushed shareholders of Sayanskhimplast to the verge of selling the company. This would have changed the market structure, particularly with Rosneft as a potential buyer, which would restore the vertical integration that existed before privatization, challenging the claim that FAS supports competition. The mechanism used by FAS to solve the conflict substantially dif-
ferred from the MMK vs. RUSAL case: by classifying the case as falling under civil law, FAS opened the door to settlement through the courts. To avoid this option, the uncertainty it creates and the transaction costs involved, parties secured their transactions through arbitration, thus consolidating the bilateral monopoly, a second-best alternative with respect to standard competition rules motivated by the need to internalize externalities amongst nodes in the technologically interdependent supply chain.

It should also be noted that these different approaches, due to consideration of secondary effects, may raise competing views amongst institutions (e.g. antitrust authority vs. courts), which can increase uncertainties amongst companies locked in bilateral relations. Also, there is another consequence that competition authorities must consider and this concerns the varying criticality of a transaction between parties to a bilateral monopoly, potentially creating asymmetry in their bargaining power, as illustrated by the Rosneft vs. Sayanskhimplast case.

We now turn to the role of antitrust authorities in such circumstances and to some possible consequences for Russian competition policy.

5. The role of antitrust authority

Based on the above discussion of the theoretical background to the regulation of bilateral monopolies and on the illustrative cases from Russia, we now submit a more normative approach to what can or could be the role of antitrust authorities. After a short summary of the ongoing legislation in Russia regarding bilateral monopoly, we take a look at alternative mechanisms which could be considered by the antitrust authority.

5.1. Bilateral monopoly in Russian legislation

The ongoing legislation in Russia does not include special provision for bilateral monopoly, nor does it mention it specifically at all. This has been interpreted by the antitrust authority as bilateral monopoly falling under its responsibility and submitted to the standard antitrust approach. This interpretation raises at least three series of important questions.

(1) Can the monopolist and the monopsonist be simultaneously complainants and defendants on the same case?

This problem clearly emerged in the MMK—RUSAL case, with the antitrust inquiry starting with a complaint from MMK, accusing RUSAL of abuse of its dominant position as buyer and followed by an adverse complaint by RUSAL against MMK, accused of abusing of its dominant position as seller! So, FAS Russia faced a situation in which each side qualified as defendant accused of anti-competitive conduct. Then the question becomes: should the two cases be examined separately, as two distinct affairs, or jointly?

From an economic perspective, the answer is quite obvious: analyzing the accusations separately would certainly lead to the omission of important characteristics of the case. As Blair et al. (1989) pointed out, treating a bilateral monopoly as if it were a monopoly or a monopsony in isolation would lead to misunderstanding what the optimal equilibrium would look like as well as what is the mechanism to reach that goal. However, from a legal perspective, the answer might not
be that obvious, especially if the technical difficulties involved in the analysis are taken into consideration.

In the case of MMK—RUSAL, according to Russian competition law, the infringements allowed the authority to issue a preliminary warning that compliance with certain conditions by the two firms involved would prevent initiating a case. Since they did comply and found a private agreement, the initial case of MMK against RUSAL was closed and the counter case of RUSAL against MMK was not even formally opened. However, this cannot always be so. First, according to the law, not all infringements qualify for a preliminary warning. Second, if conditions for a preliminary warning are not fulfilled, a case must be opened which can actually lead to the two cases mirroring each other. Third, if an agreement is not reached between parties while the investigation is ongoing, there is a possibility that both sides would be found guilty, which can set a precedent that would dissuade potential complainants to turn to the antitrust authority to protect their rights.

(2) Can market boundaries be defined separately on the supply and demand side?

In antitrust policies, “relevant markets”, which delineate both the product at stake and geographic boundaries, are defined by the possibility and feasibility of substitution—usually from the point of view of consumers. However, when it comes to abuse of monopsony power, it seems logical to rather define the market through the eyes of the producers and their capacity to switch to other clients. In other words, market definition can differ when looked at from the supply side or from the demand side. So what happens when considering bilateral monopoly? The question might seem irrelevant for “pure” bilateral monopoly, with just one buyer and one seller, so that their “market” is clearly delineated. But what can be said when two large firms have significant market power, each dominating its side of the market, and show high interdependencies while still operating with other suppliers and/or buyers, so that they have a (limited) possibility to switch partners, as in our illustrative cases?9

The European Commission, in paragraphs 20–23 of its “Commission notice on the definition of relevant market for the purposes of Community competition law” (EC, 1997) pointed out that supply-side substitution can be taken into account in the definition of the relevant market when sufficiently effective and immediate. An “old” Russian version of a similar document, about market definition and related instruments of analysis,10 also included provisions about the possible use of supply-side substitution to define a market. However, a more recent version11 does not explicitly state that possibility. This is unfortunate since many so-called bilateral monopoly situations likely require additional analysis on both sides of the market. Taking into account supply-side substitution (or demand-side substitution if the inquiry concerns a possible abuse of power by the buyer) might be crucial in defining the “relevant market” and can be essential in the decision to

---

9 Switching options may also be limited by their costs. For example, RUSAL bought coke from China, but at a price several times higher.

10 Order of the FAS Russia of April 25, 2006 No. 108 “On approval of the Procedure for analyzing and assessing the state of the competitive environment on the market.”

11 Order of the FAS Russia of April 28, 2010 No. 220 “On approval of the Procedure for the analysis of the state of competition in the commodity market.”
investigate a bilateral monopoly situation as a single antitrust case as opposed to considering two different cases simultaneously.

3) Can the antitrust authority shift the balance of market power or bargaining power?

Lodging a complaint with the antitrust authority can significantly influence the balance of expected costs and benefits for the defendant as well as the complainant, moving the bargaining power in favor of the later. However, if the initial defendant lodges a countervailing complaint, the balance of bargaining power can shift back. The resulting “market” equilibrium can significantly differ from the initial position, depending on the weight given to the arguments of the opponents by the antitrust authority. Hence, this authority can exacerbate the existing advantage of one side or becomes itself a countervailing power, redistributing the bargaining advantage in favor of the previously disadvantaged side. The impact on the parties, as well as on the social well-being, will depend on how the current situation is assessed by the authority as well as by the form and direction of its intervention (if it intervenes).

5.2. Alternative modes of intervention

From our discussion, three alternative positions can be identified for the antitrust authority.

A. No intervention. This position corresponds to the “classic” point of view in competition policies, which states that antitrust should stay out of the relationship developed in the context of bilateral monopolies due to the high risk of errors of the first type in law enforcement.

B. Intervention based on standard antitrust methods, namely:

B.1. Investigating possible market power abuse on each side separately. As already pointed out in sections 3 and 4, this approach delivers outcomes that are often far from socially optimal.

B.2. Investigating possible infringements on both sides simultaneously. As we have seen, this solution raises difficult methodological and procedural problems that require an innovative approach by the antitrust authority.

C. Endorse the role of a special mediator. In this perspective, the antitrust authority uses its position and influence to urge (and possibly “nudge”) parties to reach a mutually beneficial agreement.

This last option seems the most complicated, in that it involves informality, the potential role as a mediator for the antitrust authority not being embedded in the existing legislation. However, it presents the potential benefit of building convergence between the initially opposed points of view, avoiding the information gap that would threaten the capacity for the antitrust authority to calculate the optimal solution and establish how it could be reached, while it still permits, through the negotiation between parties, to minimize the externalities that would come out of a breach in the relationship and to reduce the negative impact it would have on other related nodes along the value chain.

However, this approach raises another question: what are the advantages of the antitrust authority performing this task rather than, say, a court or a traditional mediator? Two advantages can be pointed out. First, the antitrust authority has a comparative informational advantage due to its knowledge of concentrated
markets and forms of infringement. Second, it benefits from being able to use the threat of antitrust sanctions if parties do not reach an agreement.

If we look back to our illustrative cases, this is essentially what happened in the MMK—RUSAL case, in which the threat of sanctions gave parties an incentive to change their strategies and reach an agreement, thus avoiding the costs of antitrust sanctions while simultaneously partially internalizing the social costs of disrupting critical transactions. In a different way, this is also what happened in the Sayanskhimimplast—Rosneft conflict. In this case, arbitration helped upgrading the initial contract through an amicable agreement reached with FAS acting as a third party. This last approach prevents the antitrust authority to act directly as a mediator while making use of its informational advantage.

Notwithstanding the relative success of these modes of intervention in these cases, critical transactions having been secure through mutually beneficial agreements, it is not clear how systematic such practices can be and what legislative changes their formalization would require. Also, it is not clear either what the social costs and benefits are or what they would be if an agreement could not be obtained. Providing incentives for parties to reach an agreement is somewhat similar to “nudging” in behavioral economics (Thaler and Sunstein, 2008), with the buyer and the seller initially acting contrary to the socially beneficial alternative and being “nudged” in the optimal direction. However our cases differ, with parties being firms and not individuals prone to cognitive biases, and with the threat of sanctions being the tool to “nudge” them.

The modalities of interference to “govern” the bilateral monopoly that we have analyzed can also be viewed through a different lens, as a way to improve micro-institutions (in these cases firms) along the Kaldor—Hicks—Zerbe criterion (Zerbe et al., 2006). Indeed, the approach we have explored is not only about institutional changes, modifying the role of antitrust authority, but also about the alternative governance mechanisms this authority can use to incentivize parties to innovate in order to avoid or reduce the negative externalities that would come out of their failure to reach an agreement and secure their critical transactions.

6. Conclusions

The bilateral monopoly situation shows characteristics that make the application of antitrust legislation more constrained than in the cases of monopoly or monopsony. In particular, parties operating in that structure face switching costs which, according to transaction costs theory, require governance mechanisms adapted to the interdependence and interactions of these parties so as to mitigate contractual hazards and negative externalities.

The interference of a third party to play this role requires detailed knowledge of the specific circumstances under which the bilateral monopoly operates. In that respect, this demanding requirement provides grounds for the standard approach to regulation, which urges no interference from antitrust authorities in contractual relations under bilateral monopoly beyond the general rules regulating competition. On the other hand, a closer look at many bilateral monopolies suggests that the absence of regulatory intervention, be it by antitrust authorities or by courts, can lead to major negative externalities when technological and/or related socio-
economic factors provide the rationale for this market structure. More precisely, the existence of critical transactions linking the two sides may generate negative externalities for upward suppliers on the technological chain as well as for product consumers, but also for actors not directly involved in this production, for example when a lack of agreement may cause environmental risks or penalize employees in other sectors of activity.

The gravity of such externalities for the contracting parties and beyond largely depends on whether the unregulated transaction is critical and what its characteristics are. One difficulty is that criticality might be perceived very differently by parties to the transaction, leading to conflicts that would not make sense under the unified ownership of an integrated firm. In a bilateral monopoly, parties may be tempted to take advantage of externalities to gain competitive advantages in their bargaining power. Such strategies may lead to a redistribution of power favorable to one party, thus generating important coordination problems, often expressed through conflicts.

Thus, while it makes sense to preserve the general presumption regarding governmental intervention in contractual relations between private parties, including the case of bilateral monopoly, as inappropriate, it remains necessary to identify situations in which the “benign” neglect of sub-optimal governance mechanisms between parties leads to negative coordination effects and externalities spreading largely “beyond” these parties. When situations fall into this category, grounds may exist for regulatory interventions by a third party (e.g. antitrust authorities, courts, independent arbitrators). However, the comparative advantages of preserving the status quo, which generates negative externalities, or implementing a mode of governance of the relationship, viewed as more effective and preventing vertical integration but limiting competition, must be assessed as rigorously as possible.

The interventions of the Russian antitrust authority in recent cases illustrate the difficulties of the exercise. The regulator’s strategy seems to have been innovative, although not uniform, providing incentives for the two sides of the bilateral monopolies at stake to negotiate mutually beneficial agreements, using threat of significantly higher transaction costs, mainly related to sanctions or courts intervention, if there were delays or even failure in reaching such agreements. On the one hand, the solutions adopted allowed to internalize the social cost of the externalities for a prolonged or even failed bargaining process. On the other hand, there are costs involved so that the question remains of whether this approach is scalable and, if so, what form would fit better with the usual role of antitrust authority and its formal embedded nature in competition legislation. These issues remain open to future research.

Acknowledgements

Special thanks to Nelli Galimkhanova (FAS Russia) and Kirill Dozmarov (Art de Lex) for useful insights, particularly for section 5. The authors also benefited from comments from participants to the seminar on “Antitrust and bilateral monopoly”, organized by the Competition and Industrial Policy chair at the Moscow State University on January 25th, 2018.
References
