

CHAPTER II

DEVELOPMENTS IN COMPETITION THEORY

1. INTRODUCTORY REMARKS

Over the past decades, there have been fads and fashions in competition theory but there has also been substantial development. Three developments are particularly noteworthy: (1) the challenge posed by contestability theory to the structure-conduct-performance approach, (2) the application of game theory to issues of competition and (3) the rise of the New Institutional Economics with its emphasis on the relevance of organisational structures and their consequences for competitive behaviour. In this chapter, an attempt will be made at spelling out possible implications of these three developments for competition policy. This theoretical framework will later be used to evaluate the EU merger policy as it is currently practiced.

Before describing these three more recent developments in sections 3, 4, and 5 of this chapter, the so-called “structure-conduct-performance” paradigm will shortly be described and critically evaluated. This has been done many a time and we do not claim to add any new or original insights here. Yet, it is important to clarify some of the basic assumptions of this approach because it has been of overwhelming importance. Although it has been heavily criticised on various grounds, its core ideas still loom large in many competition authorities. After shortly presenting some of the insights of the so-called “Chicago school” in section two, we turn to the three developments just alluded to.

2. THE HARVARD APPROACH

The Harvard Approach, although often vigorously attacked, is still the most influential approach guiding competition policy all over the world. Notwithstanding its sometimes-outdated appearance, some observers (e.g., Lopez 2001) claim that its adherents are not only alive but also well.

The approach can be traced back to John M. Clark’s (1940) programmatic paper in the *American Economic Review*. It belongs to the old tradition of industrial organisation (which has been superseded by the so-called New Industrial Organisation) in which general hypotheses were developed on the basis of single observations that were then generalised and empirically tested. This approach is also called “workable competition”, which indicates that some deviation from the model of perfect or pure competition is tolerated. Until today, the model of perfect competition has remained *the* textbook model that occupies many first chapters of many textbooks (see, e.g., the introductory chapter to Tirole 1988). The outcome of that model

is a state of the world that cannot be improved upon without making at least one actor worse off. In that sense, it establishes a frame of reference with which the realised states of the world can be compared. The assumptions of the model are, however, wildly unrealistic and could never even be approximated in reality. As a matter of fact, it has been shown that marginal improvement towards the ideal can make things even worse (the so-called Second-Best Theorem; Lipsey/Lancaster 1956). The Harvard approach is an attempt to get to grips with the critique concerning the model of perfect competition, yet preventing to throw out the baby with the bath water, namely not to give up some standard of reference that could be used in order to evaluate realised states of the world.

2.1. Main Points

Representatives of the Harvard approach claim that there exists a causal chain from the *structure* of a market through the *conduct* of the participants in the market to the *performance* of the market. With regard to market structure, two factors are taken to be of overwhelming importance: (1) the number of suppliers, and (2) the concentration of their market shares. Other factors that have at times been recognised abound: the degree of product differentiation, the degree of market transparency, the kind of production technology used (are economies of scale relevant?), the relevance of barriers to entry, the market phase (introduction, expansion, saturation, decline), the degree of vertical integration, elasticities of price or income, etc. Notwithstanding the large number of factors that could influence market structure, the two factors playing the key-role always remained the number of suppliers and their market shares.

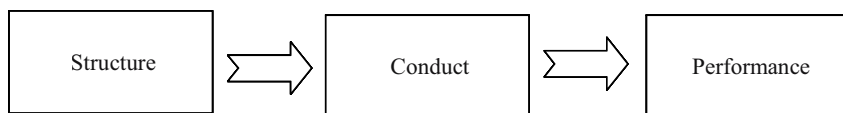


Figure 3: The Structure-Conduct-Performance Paradigm

Market conduct is operationalised by the way market participants make use of possible strategies. Here the price-setting behaviour turned out to be of central importance. Other factors sometimes subsumed under the conduct of firms include the propensity to act competitively but also the propensity to enter into anti-competitive agreements with competitors. Market performance is measured by looking at prices, qualities, quantities, but also technological progress, and, quite importantly, profit rates.

Drawing on the theoretical model of perfect competition, the “optimal” sequence of structure-conduct-performance would translate into polypolistic markets (structure) in which firms have small market shares which leads to their charging prices that are equivalent to marginal cost (conduct), which results in a profit rate that just

covers the costs of factor inputs (performance). But the representatives of the Harvard approach were unsatisfied with the concept of perfect competition. This was exactly the reason why Clark (1940) proclaimed the alternative of “workable” competition. According to representatives of workable competition, perfect competition models are fine in order to bring about static efficiency, i.e., the situation in which welfare cannot be improved by reallocation of any factors. Static efficiency is threatened by market structures with a low number of competing firms that have, in turn, high market shares. But the concept of perfect competition is a static concept. There is no space for innovation and technological progress, in short: for dynamic efficiency. Innovation presupposes the capacity to invest into research and development. According to the representatives of workable competition, a certain degree of concentration of firms is needed in order for them to be able to finance research and development.

Concerning “optimal” market structures, representatives of the Harvard approach thus find themselves in a dilemma between static and dynamic efficiency. They usually opt for a “middle of the road” approach: allow a market structure with a moderate degree of concentration but ensure by way of an active merger policy that it will remain moderate because high levels of concentration would enable firms to reap monopoly profits by setting the price above marginal costs thus violating static efficiency. To ensure moderate levels of concentration, mergers were often prohibited. If there is indeed a clear and unequivocal causal link between structure, conduct, and performance, then, in order to ascertain the workability of a market it is, at least in principle, sufficient to test for market structure *or* market conduct *or* market performance. This means that should market performance be unsatisfactory, this would be sufficient for proclaiming market structure to be unsatisfactory – and possibly demand political intervention in order to make it “workable.”

For a long time, the primary occupation of representatives of this approach consisted in assembling industry data and estimate regressions of the type

$$\Pi_i = \alpha + \beta S_i$$

in which Π_i stands for the profitability of an industry i which is supposed to be determined by the structure S of that industry. For a long time, empirical evidence was supposed to be in favour of this simple three-step process. Shepherd (1972), e.g., presented evidence showing that between 1960 and 1969, a significant positive correlation between the market share of a firm and its profitability existed. Many of the representatives of this approach took evidence of this as sufficient for a competition policy of not allowing mergers to be carried out but at times also of busting existing companies (such as the famous AT&T case). The Harvard approach thus seemed to be the perfect justification for interventionist competition policies.

2.2. Policy Implications

Believing in an unidirectional causal link from structure to conduct to performance and having a clear-cut reference of what the performance of a market should look like is the basis for far-reaching interventions into market structures as well as firm conduct. The Harvard approach would call for interventions in case the structure of a market is not likely to reach its mix of static and dynamic efficiency goals. Intervention into market structures can thus cut both ways: if the number of firms is deemed to be too small to reach the proclaimed performance indicators, divestiture might be the policy called for. The number of firms, can, however, also appear to be too large, e.g., if heavy investments into research and development seem to be necessary in order to speed up technological progress. In that case, representatives of the Harvard approach would call for merger-enhancing policies. Such policies have even played a certain role on the level of the European Union.

An alternative way of intervening into the market process is to monitor firm conduct. One could, e.g., tolerate a rather high market share but closely control conduct, e.g., by publishing maximum prices, etc. A variant of this approach is implemented in Art. 82 of the EEC Treaty, which prohibits any abuse of dominant market positions. Approaches to monitor and sanction firm conduct have often been evaluated as unsuccessful and have been used less and less over an extended time period.

2.3. Critique

Both the approach and its policy implications have been heavily criticised. Some of the more important points are simply highlighted here.

- Assuming that there is a causal chain, it would supposedly not be unidirectional: performance would have repercussions on the number of market participants, hence on market structure.

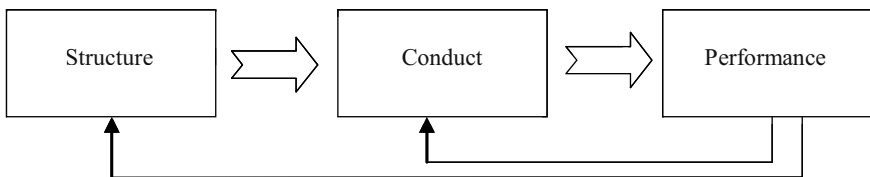


Figure 4: Structure-Conduct-Performance Paradigm under repercussions

- Correlations are not necessarily causal chains. High profitability could, e.g., also be the consequence of lower production costs. If low production costs enable a firm to capture a large market share, one would indeed expect a high correlation between market share and profitability.

- There is a very small number of actions that can be said to always be conducive or always be detrimental to competition: price cuts have often been said to be part of a predatory strategy, yet they can be a sign of fierce competition that forces firms to lower their prices to reflect marginal cost. Observing conduct will, in other words, seldom be sufficient to establish that competition is not workable in a particular market.
- The number of criteria for the market performance test is very high. This means that criteria need to be prioritised, be attached certain weights, etc.
- Representatives of a procedural approach toward competition would claim that “competition is a discovery procedure” (Hayek 1978) and that its results are therefore systematically unpredictable. Neither the optimum number of firms in a market nor their optimal conduct can be known *ex ante* but have to be discovered by the competitive process. The interventionist stance promoted by this approach would therefore be inimical to the very basis of a market economy.
- The possibility of widespread interventionism is detrimental to predictability, the criterion identified as crucial for a good competition policy in chapter I. This becomes very apparent in the following quote (Kantzenbach/Kottman/Krüger 1995, 25): “A structure oriented competition policy often reduces the level of planning security.” Kantzenbach is the most influential adherent of the Harvard approach in Germany.

Especially in its heyday, the 1960s, the paradigm was highly influential. When its followers observed some action that was not immediately explainable, they supposed the action was undertaken in pursuit of monopolising. That firms were trying to economise – and thus to increase welfare – did often not even come to the minds of those who believed in the paradigm. In the early 70s, the later Nobel laureate Ronald Coase (1972, 67) wrote: “If an economist finds something – a business practice of one sort or another – that he does not understand, he looks for a monopoly explanation. And as in this field we are very ignorant, the number of ununderstandable practices tends to very large, and the reliance on a monopoly explanation frequent.” Barriers to entry were another obsession of the representatives of this approach. We will deal with them in some detail in the third section of this chapter. Before, we turn to the so-called Chicago approach which can only be understood as an answer to Harvard: Whereas followers of the Harvard approach suspected monopolizing practices just about everywhere, followers of the Chicago approach turned the whole story squarely onto its head: they explained just about every behaviour with underlying efficiencies.

3. THE CHICAGO APPROACH

What was to become the Chicago School of Antitrust started out as a critique against the concept of workable competition that was summarised in the last section.

3.1. Main Points

Instead of aiming for a variety of – partially incompatible – goals as “Harvard” had done, “Chicago” radically reduced complexity by focusing on one single goal, efficiency. Market structure did not play any role – as long as outcomes were efficient. Monopoly was, in fact, radically re-interpreted: if a firm has been able to establish a monopoly position, this was taken as an indicator that it must be the most efficient firm in the market (the so-called “survival of the fittest” or “survivor test”). Reasons for firms being able to establish monopoly positions could, e.g., lie in their achieving economies of scale or cost savings as an effect of learning by doing. According to Chicago, it would be foolish to prohibit firms from achieving efficiencies because these mean cost-savings and, in the end, higher consumer surplus.

Chicago economists distinguish between three kinds of competitive constraints, namely (1) natural, (2) artificial and (3) state-created ones. Natural constraints to competition are not created by men, they just exist as such (e.g., if there is just one river that can be used for shipping or just one deposit of bauxite). Even if they have an influence on competitive results, trying to fight them would be pointless because they are not wilful creations of men. The artificial creation of constraints to competition by competitors is deemed to be “foolish and self-defeating behaviour”. Since erecting such constraints is not in the rational self-interest of firms, their appearance is supposed to be highly unlikely. Hence, competition authorities should not be pre-occupied with them.

State-created competitive constraints are, however, a different story. These include a host of regulations. A good example is tariff and non-tariff trade barriers to trade that protect domestic firms from having to compete on an equal footing with foreign firms. Such trade barriers will result in a loss of consumer surplus, and thus efficiency. Some of these constraints are very dangerous: take import caps as an example. For a long time, Italian regulations prohibited the import of more than 3,500 Japanese-made cars into Italy on an annual basis. No matter how good, cheap or efficient Japanese cars were, Japanese carmakers were completely barred access to the Italian markets beyond the threshold of 3,500 cars. Such policies can obviously entail heavy costs for consumers. Ultimately, they will also hurt the producers as they will be (partially) exempt from competition. In the long run, their competitiveness will decrease and their business prospects worsen.

In 1968, Oliver E. Williamson published the so-called trade-off model in which potential costs of a horizontal merger are weighted against its potential benefits. At the time, Williamson was often considered to belong to the Chicago School. In the meantime, he has, of course, been central in the development of Transaction Cost Economics to which we turn later in this chapter. The trade-off model is, however, still an important and central model and thus deserves to be shortly presented here.

This model entails a worst-case scenario: suppose that a merger changes outcomes from the perfectly competitive case (hence price equals marginal cost) to the monopoly case (in which marginal returns equal marginal costs). In that case, the welfare losses of the merger can be depicted as the triangle ADH in figure 5. These are the costs, and the possibility of this triangle is the main reason why many mergers have not been passed. But Williamson does not stop there. He stresses that there are potential benefits that should be taken into consideration, namely benefits based on lower cost curves. There are, of course, a variety of reasons why costs could be lower subsequent to a merger, lower input prices being the most obvious one. The pre-merger cost curve is depicted by MC_1 , the post-merger cost curve by MC_2 .

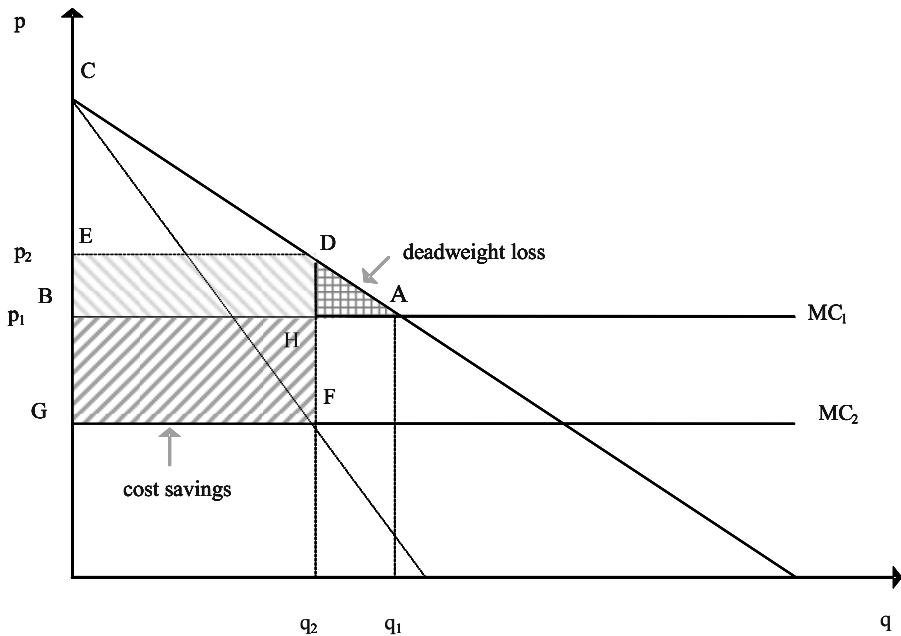


Figure 5: The Williamson Trade off

The gains of the merger in terms of saved resources are the difference in cost curves (i.e., $MC_1 - MC_2$) times the quantity produced and they are depicted as the rectangle which is called “cost savings”. In this case, cost-savings are expected to outweigh the deadweight loss. From an efficiency point of view, the merger should thus pass although the merged entity might decide to increase prices.

This is, however, only correct if the decisive criterion used to decide merger cases is overall welfare, which can be decomposed into consumer rent and producer rent. The pre-merger consumer rent is depicted by the triangle abc , the producer rent is zero. Post-merger, things have changed: The consumer rent has been reduced to

cde, while there now exists a positive producer rent depicted by the rectangle defg. The triangle adh is nobody's gain, that is why it is called deadweight loss. We thus observe a redistribution of rents from the consumers to the producers. Most members of the Chicago School now argue that producer rent is also part of overall welfare and does not constitute a problem as such. Others argue that the decisive criterion should not be overall welfare, but only consumer rents. They are thus ready to forego efficiencies in production.

It is one of the important achievements of the Chicago approach to have pointed to the detrimental effects of state-mandated constraints with regard to the functioning of competition. An important task of competition policy would thus be the undoing of state-created competitive constraints that inhibit the realisation of efficiencies.

3.2. Policy Implications

The implications of the Chicago approach for competition policy are straightforward: market structure should not be an intermediary goal of competition policy. High degrees of concentration are the result of attempting to achieve efficiency and should therefore not be suspect as such. Merger control should be handled restrictively and should only set in past very high market shares. Divestiture should not be pursued since it leads to the reduction of consumer surplus by making the realisation of cost savings impossible.

The detrimental role that representatives of the Chicago approach attributed to state-mandated constraints to competition was just spelled out. The ensuing policy implication is easy to name: the state was called upon to reduce the amount of regulations that effectively worked as constraints to competition. This could be tariff and non-tariff barriers with regard to border-crossing trade. But it could also be health- or safety-standards that often result in effectively protecting a limited number of domestic producers from more vigorous competition and thereby keeping consumer surplus at unnecessary low levels.

Representatives of Chicago did not, however, opt for a complete absence of competition policy as is sometimes argued. They argued, e.g., that horizontal agreements such as collective fixing of market shares or price fixing agreements should be prohibited. Predictability, a concept of crucial importance here, is highly valued by the lawyers and economists who belong to the Chicago school of antitrust. In order to make competition policy as predictable as possible, they opted for predominantly using *per se rules* as opposed to the *rule of reason*.

These two concepts have played a major role in antitrust policy. Due to their direct implications for predictability, they will be dealt with here in some detail. It should be stressed that their use is not confined to adherents of the Chicago approach. On *per se rules*, US Supreme Court Justice Thurgood Marshall had the following to say: "Per se rules always contain a degree of arbitrariness. They are justified on the assumption that the gains from imposition of the rule will far outweigh the losses and that significant administrative advantages will result. In other words, the potential competitive harm plus the administrative costs of determining in what particular situations the practice may be harmful must far outweigh the benefits that

may result. If the potential benefits in the aggregate are outweighed to this degree, then they are simply not worth identifying in individual cases.” (cited after Bork 1978, 18)

Per se Rule vs. Rule of Reason

According to per se-rules, certain actions are prohibited as such, i.e., regardless of the consequences that they would supposedly bring about in a specific case. According to the rule of reason, a competition authority or a judge has to decide whether the behavioural effects in a specific case have detrimental effects on a given goal such as welfare or efficiency. The decision between these two types of rules is not an either-or decision because per se elements can be combined with rule of reason elements: in many jurisdictions, cartels are prohibited per se, but can be allowed given that certain offsetting effects are expected to materialise. With regard to mergers, one can often observe the opposite approach: below certain specified criteria, mergers are allowed per se, past those criteria, the rule of reason sets in and the competition authorities will have to evaluate predicted welfare effects of a specific merger.

Per se-rules are certainly conducive to predictability. Competition authorities and judges will not have to make complicated welfare evaluations. Therefore, this saves on decision-making costs. On the other hand, there might be cases in which the competition authorities and the judges could be almost certain that the negative effect that is generally connected with a certain behaviour and that has led to the passing of a per se-rule in the first place will not materialise in a specific case under consideration. In some cases, per se-rules thus force competition authorities and judges to ignore knowledge that they really have. Hayek (1964) has shown that this can be rational if the aggregate sum of the advantages of per se-rules outweigh the aggregate costs of having to ignore better knowledge that one might dispose of in a minority of cases.

The rule of reason, on the other hand, is based on a cost-benefit assessment of individual cases. Its application thus presupposes widely available knowledge concerning the use of quantitative techniques. Especially, where competition authorities lack economics-trained staff and where judges do not receive economic training on a regular basis, use of the rule of reason seems problematic. Yet, if efficiency arguments as introduced in the last section are to play a role in merger policy, use of the rule of reason is indispensable.

3.3. Critique

The Chicago approach has been heavily criticised on various grounds. Some of the points are just mentioned here:

- The models are still built on the concept of perfect competition. More recent developments are not sufficiently taken into account;

- Some of the assumptions are not likely to lead to correct predictions (rationality, absence of barriers to entry);
- The models were wrong in some important details; Chicago, e.g., argued that monopoly positions could not be expanded upstream or downstream by way of vertical integration. In the meantime, it has been shown that this can be done (e.g., Riordan 1998)

The Chicago approach can only be understood as an answer to Harvard. What is fascinating about the two approaches is that both of them were inspired by traditional price theory and welfare economics. Comparing the two approaches shows that a very similar theoretical body can be used to develop completely diverging models and to draw radically different policy conclusions. Whereas Harvard always looked for a monopoly motivation of some action, Chicago tried to justify almost every action by first asking whether it could not enhance efficiencies. What was obviously lacking were more fine-grained approaches with a more elaborate theoretical basis and a lesser degree of ideological prejudices.

4. CONTESTABILITY THEORY

We have just seen that – under certain conditions – the representatives of the Chicago approach would not be intrigued by high market shares of a limited number of competitors. Quite to the contrary, they would argue that because these competitors are more efficient, they have been able to grow at the expense of their less efficient competitors. We now turn to the theory of contestable markets that was developed in the early 1980ies and whose representatives also claim that – under certain conditions – market structure is not a good predictor of the performance to be expected in a particular market. Formulated differently: although a market might be described by a narrow oligopoly or even a monopoly, this does not necessarily have to stand in the way of allocative efficiency. Contestability theory can thus also be read as a critique of the Harvard approach.

4.1. Main Points

William Baumol and his various co-authors (e.g., 1982) reach their central insight by integrating potential competition into their analysis. If some potential entrant can credibly threaten to enter a market in a “hit-and-run” manner, this will lead to the erosion of all monopoly profit by the incumbent. “Hit-and-run” means that a new competitor can enter into a market but will be able to leave it before the incumbent has a chance to react, e.g., by retaliation. Since this is a very important development of competition policy, we want to have a closer look at the underlying model.

Baumol et al. (1982) illustrate the relevance of potential competition by drawing on the case of natural monopoly. Economists talk of natural monopolies if fixed costs are so important that the entire relevant demand of a certain good can be supplied cost efficiently by a single supplier. Markets in which fixed costs play an important role are markets in which a net, e.g. a telephone or a railway net, is necessary

before a single unit of goods can be supplied. Once a supplier has at its disposal a net, the average cost of every additional unit decreases over the entire relevant demand range. To transport only one container on a newly erected railway net will be very expensive indeed, every additional container transported will reduce average costs because the fixed costs are spread over a larger number of units.

Natural monopolies pose a problem for politics: on the one hand, monopolies are generally considered to be undesirable, on the other, a higher number of nets would lead to an increase of average costs since the fixed costs would have to be borne more than once.⁹ The answer to this predicament has traditionally been to introduce regulation and have some agency control the behaviour of the natural monopolist. In natural monopolies, application of the price-setting rule “price should equal marginal costs” would lead to losses of the supplier since he would not be able to recoup his fixed costs. Therefore, a host of other price-setting rules have been discussed, “price should equal average cost” being one of them. What is more interesting is that numerous agencies are run on the basis of this theory (in the U.S., e.g., the Federal Aviation Administration and the Federal Communication Commission).

The achievement of Baumol et al. (1982) now lies in challenging conventional wisdom concerning the necessity to regulate natural monopolies. Suppose an incumbent has hitherto set price equal to average costs and now tries to increase his price above that level. According to Baumol et al. (1982), this would induce somebody else to enter into this market with a lower price who could still make profits. Just for completeness: the incumbent cannot make itself better off by setting a price below average costs because that will not allow him to recoup his fixed costs. We can thus conclude that

- there will only be one firm on that market;
- the firm does not make any profits;
- price equals average costs (see also Tirole 1988/99, 678-80).

Notice that this result was achieved without widespread regulation. It is secured simply by the threat of a potential entrant entering the market.

The approach has often been criticised by pointing at the entry conditions that have to be present for contestability in order to achieve its beneficial results. Baumol and Willig (1986) have stressed the following fact: “Contestability does not imply the ability to start and stop production without cost— only the ability to sell without vulnerability to incumbent’s responses for a time long enough to render all costs economically reversible.” It is thus not the absence of sunk costs as such but the ability of the incumbent to react to the entry before a potential entrant has earned the costs that had to be sunk. An incumbent might be able to change prices rather quickly. In certain instances this is, however, not sufficient to make customers lost to the entrant come back immediately, e.g., if the entrant has agreed on contracts long enough to enable him to cover all the sunk costs incurred in order to enter into the market. A case that is often named in order to prove the real-world relevance of contestability theory is the airline industry. If entrants can lease aircraft on rather short terms, their sunk costs might be sufficiently low to make entry worthwhile.

The conditions under which hit and run entry can pay off are the following: (1) there is a pool of potential entrants; (2) entrants do not have a cost disadvantage; (3) sunk costs are low; and (4) contracts are long *or* incumbents are slow to react.

4.2. Policy Implications

The insights gained by adherents of contestability theory indicate that there was great potential to deregulate telecommunications, the airline industry, but also railways and public utilities. With regard to competition policy, the insights mean that even monopolies with high fixed costs could be accepted because from a welfare-economic point of view, the equilibrium described above is the best achievable one given production technology (price equals marginal costs could only be achieved if the state was to pay subsidies to the natural monopolies).

A more general conclusion from contestability theory is that market structure can be entirely irrelevant for the outcomes to be expected on a particular market. Anti-trust agencies should thus inquire into the possibilities of potential entrants to constrain the price-setting behaviour of incumbents. Narrow oligopolies would not be suspect as such but only if high barriers to entry prevent potential competitors from controlling effective competitors in their behaviour.

4.3. Critique

This theory has been met with scepticism on various grounds. Since we have already dealt with some of it in the discussion above, the major points are just sketched here without lengthy explanations.

- There are few markets in which price inflexibility would be high enough to make hit-and-run entries worthwhile.
- There are few markets in which sunk costs, i.e., irreversible investment, are sufficiently low to make hit-and-run entry worthwhile.

The models of competition policy never even faintly resemble messy economic reality. But we believe that an important message concerning competition policy follows from the models of contestability theory, namely that market structure can be negligible as long as entry barriers into a market do not constitute a serious threshold. This will be taken up again in greater detail in chapter IV.

5. THE CONTRIBUTION OF GAME THEORY: THE NEW INDUSTRIAL ORGANISATION

The “old” approach of Industrial Organisation, of which the Harvard approach described above is one articulation, was primarily interested in empirical research. Its representatives were interested in describing firm behaviour by estimating equations that were based on micro data as described above. More theoretically inclined

economists tended to describe the “old” industrial organisation as basically atheoretical in nature: science was here misunderstood as consisting of measurement, sound theoretical foundations why certain equations were estimated in the first place were often lacking. Much of this has changed with the advent of game theory in industrial organisation. This is true to such a degree that one can talk of a “new industrial organisation” that is much more theoretically inclined than its predecessor.

5.1. Game Components

Before discussing some advantages – and correspondingly some disadvantages – that the widespread use of game theory in industrial organisation entails, a very short description of the basic components of games might be in order. Game theory helps to analyse situations in which strategic uncertainty is present. Strategic uncertainty is always present if the outcome of an action does not only depend on my own action but at least on the action of one more actor. Strategic uncertainty is distinguished from parametric uncertainty in which the outcome depends on some move of nature, e.g., whether it rains or snows. A game is regularly made up of six components:

- (1) The players. A distinction is often made between two- and more actor games.
- (2) The rules. They describe the options of the various players. An important distinction with regard to competition issues is whether players are assumed to move simultaneously or sequentially. It depends on the structure of the game whether it is an advantage or a disadvantage to be the first mover.
- (3) The strategies. A strategy is a complete description of all possible options that could open to the player during the entire course of a game.
- (4) The information set. Assuming complete information means that the players fully know the rules of the game, the strategies available to all actors, but also the payoffs that result from various strategy combinations. Perfect information, in turn, is present if an actor perfectly knows all the previous moves of the players he interacts with.
- (5) The payoff-function. It contains the utility values that all players attach to all possible outcomes of a game.
- (6) The outcome. Here, the concept of (Nash-)equilibrium is of special importance. Nash equilibrium is a situation in which, given that all other players have chosen their moves and will stick to them, no player has an incentive to deviate unilaterally because there is no possibility to make himself better off by such a move.

5.2. Advantages of Using Game Theory in Competition Theory

Game theory assumes players to be individual utility maximisers that act rationally in their pursuit to maximise individual utility. The Prisoners’ Dilemma famously shows that individual rationality does not automatically translate into collective rationality. What is best for oneself is not necessarily best for the group. Individual rationality does not necessarily lead to collectively optimal results. Formulated dif-

ferently: there are situations in which Adam Smith's invisible hand simply does not work. One example are cartel agreements: although all participants to a cartel could make themselves better off by fulfilling the terms of the agreement, individual rationality will often lead to some cartel members reneging on the agreement and thus let the entire cartel agreement bust.

Empirically, the overwhelming majority of all markets have oligopolistic structures. It is well known and economists have long explicitly recognised that in oligopolies, strategic interactions among the members play an important role ("oligopolistic interdependency"). Quantities sold, prices and profits realised depend not only on my actions but also on what my competitors do. Strategic uncertainty is thus present and game theory is an excellent tool to analyse interaction situations involving strategic uncertainty.

Additionally, game theory carries with it the potential to bring to an end the perennial conflict between outcome-oriented and process-oriented approaches to competition. The Harvard approach would be the paradigmatic example of an outcome-oriented competition approach: some performance characteristics are declared as normatively desirable, if these characteristics are not fulfilled empirically, some interventionist act is called for. Representatives of process-oriented approaches, in turn, believe that the outcomes of competitive processes are systematically unpredictable. They, therefore, refrain from stating criteria that competition should bring about but rather focus on the rules according to which the competitive process should be organised. Being able to make normative statements about how the process should be organised (what antitrust rules would make the process welfare-enhancing) presupposes knowledge concerning the working of the process. Game theory has the potential to help us understand some processes better. At the same time, it also carries the potential to understand interrelationships between process and outcome better. If these processes are better understood, this might eventually enhance our capacity to pass more adequate competition rules.

Game theory might also help to question the outcome-oriented view of competition policy. An eminent scholar of the new industrial organisation, Louis Phlips (1995, 12), observes: "Pervasive to the entire argument is the idea that antitrust authorities are not social planners. A social planner wants price equal to marginal cost, plus optimal taxes and subsidies. Antitrust authorities want the best possible market structure given technology and tastes, and, given this market structure, as much competition as is compatible with it and with entrepreneurial freedom. But that is precisely, it seems to me, what is described by a perfect competitive Nash equilibrium." Phlips here seems to argue that a decision needs to be made between the concept of antitrust authorities as social planners and a concept that sees their function in strengthening and maintaining as much competition as possible under the concrete circumstances. He seems to argue against the social planning concept which is built on the model of perfect competition and that plays such a dominant role in the structure-conduct-performance paradigm. Instead, he is an advocate of the Nash-equilibrium, which he interprets as a description of how much competition is possible given the relevant circumstances.

This is an interesting position because it implies that an either-or decision needs to be made. Many adherents of the new industrial organisation do, however, supposedly not share this position. Instead, the outcomes postulated by welfare economics would still be hailed as the theoretical ideal. Game theory can be interpreted as a theory informing actors what would be in their (utility-maximising) interest given that they were rational. Assuming that they are rational, it can be used to predict what actors will do under various circumstances. It can thus also be interpreted as a positive theory. The either-or view advocated by Philips is therefore not convincing: one can still believe in the fundamental theorems of welfare economics and simultaneously analyse what the results of certain interactions are given specific circumstances. If the predictions deviate too much from the ideal striven for, then many policy-oriented game theorists would be ready to propose changing the circumstances. This could, e.g., mean to change competition rules, to increase sanctions, etc. All these changes would be aimed at bringing reality closer to a theoretical ideal. Game theory does thus not fundamentally alter the policy stance of industrial organisation. Proponents of a game-theory based industrial organisation could still be advocates of far-reaching interventions.

In the literature, a number of additional advantages for the use of game theory in industrial organisation are named:

- the introduction of sequential decision-making processes (Güth 1992);
- the explicit recognition of incomplete information (Güth 1992).

These advantages should, however, not lead one to conclude that the heavy use of game theory in industrial organisation is warmly welcomed everywhere.

5.3. Critique Concerning the Use of Game Theory in Competition Theory

There has been a good deal of criticism concerning this tool. The old industrial organisation was a discipline with a primary interest in empirical results that lacked a sound theoretical basis. The new industrial organisation is a discipline with elaborate and elegant models whose application to real-world problems is, however, often very problematic or even outright impossible. Louis Philips (1995, 11) an advocate of the use of game theory in competition policy is very frank in admitting it: "... I know that much work remains to be done on practical questions, such as how a given industry can be identified as being in a Nash equilibrium, how it gets into such an equilibrium, how it gets out of it, and how it moves from one such equilibrium to another one."

Game theoretic models will only help us in coming up with good predictions if its central assumptions are not too far off the mark. The rationality assumptions regularly used in game theory have been met with scepticism. Werner Güth (1992, 272), e.g., believes that the rationality hypothesis is the central weakness of the use of game theory in industrial organisation. If rational behaviour as assumed by the theory cannot generally be taken for granted then game-theoretic predictions will be incorrect even if the model itself is adequately specified (1992, 272).

The summer of 2000 offers a good real-life test of the applicability of game-theoretical models: In Germany, third generation mobile phone (UMTS) licences were auctioned off in a highly regulated process. Some game theorists analysed the rules and made far-reaching predictions concerning the outcome of the auction and were proven wrong (see, e.g., Moldovanu and Jehiel 2001). One could now reply that the participants in the auction were no experts in game theory and the result therefore diverged from the one expected. Yet, all companies participating in the auction heavily relied on experts – even including Nobel laureates. This shows that the predictions derived from game-theoretic models do not seem to be very reliable.

As soon as games are played repeatedly, a very large number of outcomes become possible (this insight is called folk theorem by game theorists because it was common knowledge long before it was formally proven). For the predictive power of game theory, this is a serious problem: one central criterion for judging the quality of theories is the number of outcomes that it predicts will *not* occur. If a large number of equilibria are possible in repeated games, this is thus a serious problem for the predictive quality of game theory. Attempts to deal with this problem, such as equilibrium selection theories as advocated by Selten (1975) have been only moderately successful.¹⁰

In any theory, the outcomes are driven by the assumptions imputed into a theory. This is, of course, also true for game theory. But with regard to game-theoretic models, the sensitivity of the outcomes to minor modifications in the assumptions seems to be very far-reaching. Formulated the other way around: game-theoretic models are not robust.

Some, but not all, game-theoretic models seem to assume a curious asymmetry concerning the information at the disposal of many actors: actors might have incomplete, imperfect or asymmetric information but it is sometimes assumed that the (scientific) observer is not constrained by such problems. Now, if a player is able to fool those he is interacting with it is hard to see why he should not be able to fool scientific observers watching him.

Game theory thus has lots of advantages as well as disadvantages. One problem with game theory that has not been mentioned so far is the assumption of a firm as “given”. Nobody asks for the rationale of its existence because it simply exists. We now turn to a theory in which firms are not assumed to be exogenously given any more but are endogenous to the competitive process. This is the new institutional economics.

6. THE CONTRIBUTION OF THE NEW INSTITUTIONAL ECONOMICS: TRANSACTION COST ECONOMICS

The New Institutional Economics is a success story. At least five Nobel laureates can be counted as belonging to it (Kenneth Arrow, Ronald Coase, Friedrich Hayek, Douglass North, and Herbert Simon). Its competition theory branch, transaction cost economics, has made considerable impact on U.S. antitrust policy. It has, e.g., led to a fundamental modification of the evaluation of vertical as well as geographical restraints which were shown to be welfare enhancing under specific circumstances

(Williamson 1985, chapter 14 and *passim*). In Europe, however, its effects on competition policy have been rather marginal. We therefore decided to present this approach in a little more detail than the other approaches dealt with in this chapter.

6.1. Transactions and Transaction Costs

The representatives of Transaction Cost Economics believe that transactions are fundamental for the economic process. They are interested in the analysis of conditions under which welfare-enhancing transactions take place. As already spelt out in the chapter on predictability, the costs that have to be incurred to execute a transaction are a crucial factor for the number of transactions to be expected. Transaction costs are thus a basic concept used by representatives of the New Institutional Economics. Transaction costs are

- the costs of searching exchange partners with whom to transact and to get information on the qualities of the goods that they offer as well as information concerning their reliability;
- the costs of reaching agreement, i.e., bargaining and decision-making costs;
- the costs of monitoring whether the exchange partner has delivered as promised;
- the costs that have to be incurred to get the terms of the original contract implemented, e.g., fees for lawyers and court costs.

In economics, transaction costs have long been neglected. It was thus implicitly assumed that the costs of transacting were zero. This amounts to assuming that all actors were fully rational and had at their disposal complete knowledge concerning every conceivable state of the world. This is a highly unrealistic assumption and the representatives of transaction cost economics can claim credit for having outlined the consequences of assuming transaction costs to be positive.

Another traditional assumption in economics was to model the firm as a production function, i.e., as a technological relationship between inputs and outputs. In this approach, the firm really was a black box, because the process by which inputs were transformed into outputs was completely ignored. But different organisational structures have different consequences on the incentives of those working inside the firm and also on those transacting with the firm. It is, again, the merit of the representatives of transaction cost economics to have pointed to the crucial importance of organisational structures. Conceptualising organisation structures as devices to economise on transaction costs can lead to a fresh look at some business practices that had hitherto been judged as monopolising behaviour which can be interpreted differently now. This should, quite obviously, have far-reaching consequences for competition policy.

6.2. *Assumptions of Transaction Cost Economics*

Transaction cost economists start from assumptions that are different from some of the more established approaches. True, taken separately, most of these assumptions have been around for a long time. It is, however, the merit of transaction cost economists to have synthesised them into a coherent theory. Three assumptions are of particular importance: (1) bounded rationality; (2) opportunistic behaviour, and (3) asset specificity. We shortly want to deal with every assumption in turn.

- (1) *Bounded rationality*; this assumption means that actors are not assumed to be completely rational anymore but only limitedly so. Starting from boundedly rational individuals would make the assumption of actors who try to maximise utility in every instance a shaky one. Nobel-laureate Herbert Simon (1955) therefore proposed to assume that actors behave in a “satisficing” manner. Actors form expectations concerning the level of utility they hope to secure. As long as they factually secure that level, they do not have any incentives to change their behaviour. Only if the utility level aspired for is not reached anymore do they start to search for modified behaviour with the aim of reaching their old level of utility again. One consequence of bounded rationality is that contracts will not cover every possible contingency that could arise. They will, in other words, remain incomplete. This means that situations can arise that are not fully anticipated in contracts and the contracts do thus not specify the consequences of these situations completely, either. In such situations, general structures that can be used to settle conflicts are needed.
- (2) *Opportunistic behaviour*; this assumption means that actors who can make themselves better off to the detriment of others should generally be expected to do so. If no institutional safeguards are available making opportunistic behaviour unattractive, then many potentially welfare-enhancing transactions will not take place.
- (3) *Asset specificity*; this assumption means that some assets can only be used for very specific purposes. Opportunistic behaviour in combination with asset specificity can, e.g., become relevant if a good with specific characteristics needs to be produced before it can be sold. Ex ante, the buyer has every incentive to point at his ability and willingness to pay. Once the good is produced – and the second-best way to use this specific product is worth a lot less than the first-best – the buyer can be expected to ask for a reduction in price.

Economic institutions now serve the purpose of reducing transaction costs. Depending on the relevance of the assumptions just spelt out and on a number of factors to be spelt out in a minute, different “governance structures” (Williamson) are optimal in order to cope with the specific circumstances of the situation. Anything from a classic spot market transaction to a hierarchical firm can be a governance structure. It is important to note that governance structures can be conceptualised as a continuum with the two forms just mentioned as their most extreme points. In between, a

large number of hybrid forms such as long-term contracts, franchising agreements, etc., are conceivable.

Before the introduction of transaction cost into economics, the existence of firms could not be convincingly explained. If one assumes market transactions to be executable without any costs, it is unclear why any transactions should not be executed via the market. Hierarchies, firms, however, are a way to ensure transactions not through voluntary consent but through command. Setting up and running organisational structures is certainly connected with positive costs and in the absence of transaction costs it is unclear why they should be incurred. As soon as transaction costs are introduced, this whole picture changes dramatically. A shorthand for defining transaction costs is to call them the “costs of using markets” (Coase 1937). As soon as transaction costs and organisation costs are taken into consideration, predictions concerning the (optimal) size of firms can be generated: a firm will grow until the marginal revenue of integrating yet another activity is equivalent to the marginal costs that have to be incurred in order to integrate that activity. Formulated differently: the expansion of a firm will stop as soon as the transaction cost savings from integration are less than the additional organisation costs to be incurred.

The central hypothesis of transaction cost economics is that the specific characteristics of the relevant transactions determine the optimal governance structure. Williamson analyses the effects of three characteristics, namely of (1) asset specificity, (2) the degree of uncertainty, and (3) the frequency with which the relevant transactions are expected to take place. These three can be considered as independent variables, the optimal governance structure is the dependent variable to be explained with the independent ones. Simply put: we would expect governance structures to be more integrated, the more specific the assets used in some business relationship, the more important the role of uncertainty, and the more frequently transactions are expected to occur.

This is a static description concerning the optimal size of the firm. Changes in transaction as well as in organisation costs can be one factor leading to changes in optimal firm size. Reductions in organisation costs would, *ceteris paribus*, increase optimal firm size, reductions in transaction costs would, again *ceteris paribus*, decrease optimal firm size. In the early years of transaction cost economics, there was indeed the simple dichotomy between “Markets and Hierarchies” (Williamson 1975). In the meantime, representatives of this approach tend to think of these two forms of organisation as the extreme points of a continuous line, which allows for a multitude of so-called “hybrid” contractual agreements. They allow to explain the rationale of franchising, joint ventures, long-term contracts, etc., which were traditionally met with much scepticism by competition authorities.

With regard to the new industrial organisation, we observed that its representatives have produced exciting theories but that the empirical tests were somewhat lagging behind. This judgment cannot be made with regard to transaction cost economics. Although measuring concepts such as asset specificity or uncertainty with any degree of reliability seems no mean feat, it has been done successfully. Most empirical studies measuring asset specificity have relied upon Williamson’s (1985, 95f.) proposal to distinguish four kinds of it, namely (1) site specificity (costs of

geographical relocation are great), (2) physical asset specificity (relationship-specific equipment), (3) human asset specificity (learning-by-doing, especially in teams comprising various stages of the production process), and (4) dedicated assets (investments that are incurred due to one specific transaction with one specific customer). When estimating the effects of asset specificity on governance structures, one thus needs ways to measure one (or more) of these four kinds. Physical proximity of contracting firms has been used as a proxy for site specificity (e.g., by Joskow 1985, 1987, 1990 and Spiller 1985) and R&D expenditure as a proxy for physical asset specificity. With regard to both human asset specificity and dedicated assets, survey data have been used.

Instead of describing the empirical evidence in any detail here, we refer the reader to the survey of empirical studies by Shelanski and Klein (1999) and quote an early paper by Joskow (1991, 81) who observes that the empirical literature testing hypotheses based on transaction cost economics “is in much better shape than much of the empirical work in industrial organisation generally.”

Nevertheless, one empirical study dealing with alternative explanations for vertical mergers is too much to the point not to be cited. Spiller (1985) compared the predictive qualities of transaction cost economics with those of the market power paradigm of the Harvard approach. The latter predicts that the benefits of a merger increase in the degree of (supplier) market concentration, while transaction cost economics predicts that they increase in the degree of asset specificity. Gains from mergers are here operationalised according to the unexpected gains in the firm’s stock market prices at the announcement of the merger. Spiller finds that gains from mergers are smaller the greater the distance between the merging firms, i.e., the lower site specificity, whereas the industry concentration has no significant effect. This can be interpreted as evidence that the power of transaction cost economics in predicting mergers is higher than that of the more traditional structure-conduct-performance paradigm.

6.3. Policy Implications

In academia, the New Institutional Economics is a highly successful research programme. This can also be proved by looking at citation records. But in European competition circles, transaction cost economics as an important part of the New Institutional Economics has probably not received the attention it deserves. Policy implications resulting from the programme have not entered the pages of many textbooks. Therefore, we propose to describe the methods used by representatives of the programme to get to policy implications here.

Many approaches in competition theory have traditionally drawn on some ideal state of the world, perfect competition being the most famous such state. Real-world results were then compared with theoretically derived states of the world. Needless to say, reality often appeared as utterly bad compared to the theoretical ideal. For many economists, the next step was then a small one: demand that the state intervene to make the actors behave in such a way that would at least approximate the theoretically derived ideals. Basically, this notion should have been discredited ever

since the concept of Second Best was published (by Lipsey and Lancaster in 1956). It is shown that attempts to emulate the prerequisites for the ideal world can lead to outcomes that are worse yet. To take an example from competition theory: an oligopoly is supposed to be an untransparent market structure, theory assumes a transparent market structure. At times, it has then been demanded that official price offices be founded that collect and publish prices by all oligopolists in the market. This could, however, lead to worse outcomes since parallel behaviour by oligopolists would be facilitated.

Transaction cost economists only compare realised states of the world with other realised states of the world. At best, they only take realizable states of the world into consideration. This means that no abstract ideal is painted any more but that one asks for marginal improvements that take the current situation that one finds itself in as the starting point. This approach is today called “comparative institutional analysis” and the idea was first coined by Ronald Coase (1964). When representatives of the welfare-economic approach demand state interventions as corrections to market failure, they often commit a logical mistake: after having identified some “market failure” (as compared to a theoretically derived ideal) they demand state interventions and assume that the state functions perfectly. This is, of course, a dishonest procedure: if market failure is taken into account, then government or constitutional failure should also be taken into account. The state and its representatives do not function without cost. Williamson (1996, 195) tries to take this into account and proposes the concept of “remediableness”. If one proposes a new policy, one better take the costs of getting from the current status quo to the proposed policy explicitly into account. Getting there might be costly (necessary investment but also political opposition are just two possible cost components). The proposed policy only constitutes an improvement if the returns from the new policy are higher after the costs of getting there have already been reduced from the expected benefits. This leads Williamson to redefine the notion of efficiency (1996, 195): „An outcome for which no feasible superior alternative can be described and implemented with net gains is presumed to be efficient.“

Taking both the relevance of transaction costs as well as the modified definition of efficiency into account, a number of policy implications can be derived:

There are conditions under which vertical integration can enhance efficiency. This will be the case when transactions cost savings outweigh additional organisation costs. Under such circumstances, the prohibition of mergers would be detrimental to overall efficiency. They should thus be allowed.

There are conditions under which other forms of governance such as long-term contracts or exclusive dealing contracts can enhance efficiency. If a certain service quality can only be upheld given some exclusive contracts, this might be a case in point. A similar point can be made with regard to geographical restraints. The policy implications are obvious: investigate whether the conditions are fulfilled. If so, do not prohibit the specific restraints because that would decrease overall efficiency.

Furthermore, conglomerate concentration may be inexplicable from the point of view of technology, but may very well be explicable by looking at the firm as a governance structure. If conglomerate concentration can be reconceptualised as a result

of economising on transaction costs, it should not be punished anymore because that would decrease efficiency.

These policy implications were derived supposing that governance structures are a result of firms' attempts to economise on transaction costs. Above, some emphasis was put on the method used by representatives of transaction cost economists, namely comparative institutional analysis. Taking this method seriously can have far-reaching policy implications too. According to it, the existence of barriers to entry as such is not sufficient for demanding intervention by competition authorities as long as it cannot be proven that there is a better structure that can be implemented at reasonable cost. Williamson (1996, 282) writes: „... while the mere existence of entry barriers was previously thought both objectionable and unlawful, this non-comparative approach has been supplanted by one in which (as an enforcement matter) the relevant test is not whether entry impediments exist but whether a remedy can be effected with net social gains. As a result, arguments regarding the mere existence of entry barriers no longer carry the day.”

7. IN LIEU OF A SUMMARY: CONSENSUS AND DISSENSUS BETWEEN THE VARIOUS APPROACHES

Looking at the relationship of the various theoretical developments, there are complementarities as well as incompatibilities. Although modern Industrial Economics has come a long way, it seems still firmly rooted in the structure-conduct-performance paradigm. Not only have the main questions remained the same; the basic conjectures have also largely remained unchanged. What has changed is the toolkit: whereas the Harvard paradigm used to be primarily inductive, modern Industrial Economics has turned deductive, being based on game theory. Moreover, modern Industrial Economics does not solely focus on structural factors anymore, but tries to incorporate the behavioural incentives of the relevant actors.

Transaction Cost Economics also takes the behavioural incentives of the actors explicitly into account. Many of its representatives also draw heavily on game theory and there is thus substantial overlap with modern Industrial Economics. This is, e.g., documented in the textbook by Tirole (1988) who is one of the leading representatives of The New Industrial Organisation; for a number of chapters, he draws heavily on Joskow's lectures delivered at MIT. Joskow is, of course, one of the leading representatives of Transaction Cost Economics.

Yet, there are a number of incompatibilities between NIO and TCE. The most important one seems to be the underlying standard of reference: the NIO remains within traditional neoclassic thinking: define an abstract welfare standard, compare reality with it, if reality diverges substantially, devise some policy in order to get reality closer to the theoretical standard. Transaction Cost Economics believes that such an approach is of little help. Based on the notion of Comparative Institutional Analysis, it redefines efficiency in a way that takes the specific constraints explicitly into account. Traditional welfare economics has identified a host of so-called “market failures”. TCE explicitly recognises that it is not only the market that can fail but also government and bureaucracies. Taking these failures into account, one can of-

ten not improve the current situation. If that is the case, it is called “efficient”, even though it does not fulfil the tight criteria named by more traditional approaches. Though representatives of both approaches talk of “efficiency”, they mean completely different things.

Representatives of TCE start from the assumption that the borders of a firm are the result of transaction cost minimising strategies. Contracts and agreements that have generally been subsumed under behaviour “in restraint of competition” need to be re-evaluated: these can be horizontal, vertical, and conglomerate ones. It was representatives of TCE who were able to show that these are often entered into with the goal of saving on transaction costs and that they were thus not necessarily restraining competition. Market structure in the traditional sense does thus not play a decisive role in TCE anymore. In recent decades, new forms of cooperation between firms have emerged.

Structural approaches towards competition policy still seem to be the dominant ones. One possible reason is that it is still very difficult to measure transaction costs as well as other central notions of TCE such as asset specificity, uncertainty, and frequency. In order to gain further ground, representatives of TCE should thus think hard about hands-on approaches how to deal with these issues. After having delineated the relevant market, structure-oriented economists need to do some simple number games to come up with concentration ratios and the like.

So far, only three approaches – namely the traditional structure-conduct-performance paradigm, the New Industrial Organisation, and Transaction Cost Economics – have been mentioned. As fully-fledged approaches, they seem indeed to be dominating discussions on competition policy. But what about the other two approaches and their relationship to these three more important ones?

“Chicago” did not only develop as a critique to Harvard but also to antitrust policy US style. Its representatives had the impression that there were many inconsistencies in antitrust policy as practiced in the US during the 60s and 70s. This is the reason why Bork (1978) gave his book the subtitle “A policy at war with itself.” Many of the shortcomings pointed at by representatives of “Chicago” have been corrected in the meantime: state-mandated barriers were not only recognised as a serious impediment to competition but were dismantled to a large degree during the privatisation and deregulation policies observed in many countries in the 80s and 90s (see also Chapter III for more on this).

The representatives of contestability theory did not carry out an attack against Harvard as sweeping as Chicago did. Yet, on theoretical grounds, they were often taken more seriously than the Chicago boys as they argued out of the same paradigm and came to the conclusion that under certain, carefully specified conditions, structure did not matter at all for the results (“performance”) to be expected in a market. Although contestability theory has been criticised because these conditions seem very seldom – if ever – to apply in reality, it has also made an important effect on competition theory: according to it, the effectiveness of potential competition crucially depends on the significance of barriers to entry. Part of the message is almost identical to that of Chicago, although the representatives of contestability come from a different theoretical background: (competition) policy should focus on reducing

state-mandated barriers to entry, as this will increase the likelihood of beneficial results of the competitive process.

In order to make the various points of consensus and dissensus among the various approaches even more concrete, we will discuss the ways in which they deal with one issue that all approaches that pretend to be applicable to policy issues need to deal with somehow: the recognition that mistakes can occur and how one deals with that possibility. Two types of mistakes can be distinguished:

- Type I errors: Efficiency-increasing and thus welfare increasing mergers are wrongly prohibited.
- Type II errors: Mergers that are not efficiency-enhancing and thus not welfare-increasing are wrongly allowed.

This classification of possible errors is based on welfare economic grounds. From a welfare economic point of view, the gains of any merger can be expressed in terms of increased productive efficiency. But mergers can also cause allocative inefficiencies if they enable a firm to be powerful.¹¹ The competition authority thus needs to make a trade-off between gains in productive efficiency and losses in allocative efficiency. It can only commit two mistakes. It can either exaggerate the expected allocative inefficiencies (and turn notified mergers down although they should be passed) or it can overestimate the gains in productive efficiency (and clear the merger although it should be prohibited). These two types of errors thus reflect the welfare economic approach towards mergers.

For competition policy this does, however, not mean that only those mergers should be allowed that explicitly generate efficiencies. In competition policy, all mergers should be passed as long as they do not overly restrain competition. Problems only arise if a merger threatens to overly restrain competition. Only in that case should efficiency considerations play an explicit role. From a welfare economic point of view, one would then ask whether allocative inefficiency can be expected to be overcompensated by gains in productive efficiency.

Any competition authority faces the dilemma of having to trade off the two types of errors against each other. If the authority decides to take a tougher stance on mergers, thus letting fewer mergers pass, it reduces the probability of committing type II errors but simultaneously increases the probability of committing type I errors. The inverse relationship also holds: if a competition authority decides to take a more relaxed stance on mergers, thus letting more mergers pass, it reduces the likelihood of committing type I errors but simultaneously increases the probability of committing type II errors. The choice is thus a genuine dilemma.

Table 3: *The Trade-off Between Type I and Type II Errors*

		Efficiency	
		Enhancing	Reducing
Decision	Prevented	<i>Type I error</i>	Correct Decision
	Allowed	Correct Decision	<i>Type II error</i>

It is, of course, tempting to think of the “optimal” decision concerning the trade-off that would supposedly consist of minimizing the overall costs expected. The costs caused by type I errors consist of the unrealised efficiency gains that would have resulted had the mergers that were in fact forbidden been implemented. But these are not the entire costs: Every decision by a competition authority contains signals concerning likely future decisions: if it takes a tough stance on a particular merger, it can be expected to take a similarly tough stance on similar mergers. This will most likely lead some potentially welfare-increasing mergers to never be seriously pursued because every merger prohibited is connected with huge costs for the notifying parties. These could be called the dynamic effects of Type I errors.

The costs of type II errors are primarily caused by allowing welfare-reducing mergers. Allocative inefficiencies will be reflected in higher prices and lower quantities. But there is also a dynamic aspect to type II errors: if companies expect a liberal decision-practice in merger control, this will affect the number and quality of the mergers notified. It is possible that mergers will be attempted for other reasons than for improvements in efficiency such as market power. Here, the competition authorities do not send signals that would reduce an adverse selection in mergers (Basenko/Spulber 1993, 11).

For identifying an optimum, the costs of both error types need to be compared. It is in the evaluation of the costs expected with committing the two error types that the approaches presented in this chapter differ: Representatives of the Chicago approach would rather commit an error of type II than type I because they believe that errors of type I are – at least on average – more costly. Type II errors can be corrected ex post but there is no clearly identifiable ex post correction mechanism with type I errors: mergers that are not efficiency-enhancing but that are passed nevertheless are still subject to the market test: if other producers are more productive or meet consumer preferences better than the merged company the new company will lose market shares – and profits. If it is too large, capital markets are expected to correct for this (Manne 1965). In many jurisdictions, competition authorities can let mergers pass but can check the behaviour of firms that are supposed to dispose of a market dominant position. This is thus an additional channel to keep the costs of type II errors low. But if efficiency-enhancing mergers are wrongly prohibited there is no ex post market test. Efficiencies can simply not be realised.

Representatives of the Chicago approach thus argue that costs of type I errors regularly outweigh costs of type II errors. Judge Easterbrook, e.g., argues (1984, 15): „... the economic system corrects monopoly more readily than it corrects judicial errors ... in many cases the costs of monopoly wrongly permitted are small, while the costs of competition wrongly condemned are large.“

Representatives of the Harvard approach seem to be more likely to argue in favour of committing type I rather than type II errors. Traditionally, representatives of the Harvard approach have been much more critical with regard to the market than have representatives of Chicago. This is obviously reflected in their evaluation of the costs due to type I errors in comparison to type II errors.

Representatives of Transaction Cost Economics have also explicitly dealt with the issue of wrong decisions in merger policy. In a recent paper, Joskow (2002, 6) writes: „The test of a good legal rule is not primarily whether it leads to the correct decision in a particular case, but rather whether it does a good job deterring anti-competitive behaviour throughout the economy given all of the relevant costs, benefits, and uncertainties associated with diagnosis and remedies.“ The dynamic effects of errors are clearly recognised here. Moreover, Joskow clearly recognises that our knowledge concerning cause-effect-relationships is very limited and that enforcement agencies have only very limited knowledge at their disposal. Rather than taking a clear stance on what type of error rather to commit, the policy advice seems to point to broad and general rules. This can, however, not be easily reconciled with the specific type of efficiency defence that TCE stands for, namely efficiencies based on asset specificity, uncertainty, and frequency.

In concluding, it can be said that with regard to competition *theory*, a lot has been learned over the last couple of decades. In theory, a competition *policy* based on sound economic reasoning should thus be possible. The problem to be solved is to devise rules that allow taking the intricacies of a specific case explicitly into account but is yet general and robust enough to allow for a high degree of predictability. Succumbing to economic trends is not a good advice here as they have often turned out to be short-lived fads. Before we develop some proposals how this could possibly be achieved in chapter IV, we turn to the description of some business trends that an up-to-date merger policy should probably take explicitly into account.