Williamson’s Back Door: Transaction Costs and the Efficient Firm

Craig Freedman*

Abstract: Oliver Williamson has made an acknowledged contribution to the theory of the firm by incorporating Coase’s idea of transaction costs in his analysis. However, by identifying market efficiency with the minimisation of transaction costs, Williamson has linked rhetorically two concepts which are not by necessity connected.

[Morishima] attributes the continuous frustration which has beset the development of economic theory over the last thirty years or more to ‘failure of economic theorists to carry out sweeping systematic research into the actual mechanisms of the economy and economic organisation, despite being aware that their own models are inappropriate to an analysis of the actual economy.’ (Feiwel quoted in Williamson 1985, p. 386)

By drawing attention to transaction costs, Oliver Williamson (1985) deliberately took up the challenge first laid down by Coase (1937)\(^2\). According to this formulation, corporate configurations necessarily reflect decisions to retain certain operations in-house, while others remain the province of market exchange. Thus the competitive imperative of minimising the cost of transactions determines the limits of a firm. Though a useful categorisation, it still has the static feel of a catalogue featuring a typological division of existing organisational structures\(^3\). Fixing a butterfly on a mount may provide the chance for dissection but it need not convey an image of the insect in flight. We are left instead with a series of unrelated snapshots. Missing are the limitations imposed by an historical context and its concomitant prior choices. Because of this distinct approach, transaction costs under Williamson’s tutelage do not emerge as an alternative but rather as a complement to the standard neoclassical theory of the firm. What Williamson ultimately tries to demonstrate is that markets, when endowed with a richer economic environment of positive transaction costs, still yield efficient outcomes even if only in a relative rather than an absolute sense as Williamson puts it (1985, p. 22)\(^4\). Unfortunately, positing such an end result serves only to confuse Williamson’s analysis. The peculiar economic issue of optimal or even relative efficiency is not intrinsically related to the body of his analysis\(^5\). Nothing inherent in the notion of transaction costs leads inevitably to issues surrounding firm or even market efficiency. This will cause any serious student of Williamson’s work to suspect that its intrusion into what is essentially only a methodological approach has more the status of a conjuring trick than a necessary cornerstone supporting a theoretical framework.

My objective is to explain what Williamson did and to unearth a possible rationale which can explain why he would choose to adopt such an approach. By pointing this out, I am not denying the potential usefulness that the transaction cost
approach may yield. Instead, it serves as a clear example of how such analysis can be captured and enlisted for rather peripheral purposes. To use the language of von Clausewitz, some economists by capturing the high ground of economic debate can hope to determine how future campaigns will be conducted. In Williamson’s case, whatever interest the question of efficiency may have, it need not have come to be identified with or to dominate this quite separate issue of transaction costs. Moreover, by separating the two, we can make a definite break with the standard economic technique of postulating the *a priori* existence of economic efficiency⁶.

### 1 Williamson’s Crusade

... the transactions cost approach maintains that these institutions have the main purpose and effect of economizing on transaction costs (Williamson 1985, p. 1).

Oliver Williamson self-consciously sees himself as an institutionalist, an economist for whom the stylisation of the firm as a production function is an inadequate explanation of corporate activity⁷. Instead, Williamson starts by looking at the transaction as the basic building block of analysis. Economic activity depends on the successful completion of transactions. Therefore, the organisational shape firms develop must reflect to some degree the types of transactions required. Successful firms are those which are most efficient at conducting an array of necessary transactions. Williamson’s goal becomes the explanation of static organisational forms by exploring the possible transaction costs in each existing situation.

Williamson’s project starts with contracts and the nature of those contracts. These all-encompassing contractual arrangements are enveloped by a sea of limited information, or, as he prefers to call it, bounded rationality⁸. What a firm does depends just as much on what it does not know as what it does. Williamson’s worst-case view of human nature (opportunism), which motivates much of his work, is only significant in so far as it is not at all obvious ahead of time whether an opportunity will arise whereby devious human behaviour can flourish. Nor is it clear, given the informational asymmetry necessary for such opportunistic behaviour, when and where an economic actor will choose such a course. The need to explore alternative contractual forms, along with their concomitant transaction costs, becomes in part the necessary price paid to forestall the constraints imposed by limited information. The reason that opportunism poses a threat to a firm’s continued well-being lies in the nature of those contractual commitments, which leaves them resistant to rapid adaptation – what Williamson terms asset specificity⁹.

Commitments of this type reduce a firm’s ability to respond to unanticipated change, including opportunistic strategies. An attempt to minimise transactions costs is therefore not automatically consistent with maintaining an adequate level of flexibility¹⁰. Path dependency, an idea which Williamson (1993) accepts, precludes making such a simplistic identification. Being an efficient firm during period one may prevent that firm from achieving similar efficient results during period five or somewhere further down the line. Williamson relies on a natural selection-like process to finesse such inherent difficulties. Unfortunately, this particular biological analogy fails to see his argument through. The market environment in which firms operate is not exogenous, but partially a deliberate attempt to create a conducive sphere for business operations. An environment
which firms in part struggle to construct cannot at the same time be said to neutrally sort out the efficient from the also-rans.

2 Transaction Costs as a Rhetorical Wedge

The argument relies in a general, background way on the efficiency of competition to perform a sort between more and less efficient modes and to shift resources in favour of the former. This seems plausible, especially if the relevant outcomes are those which appear over intervals of five and ten years rather than in the very near term … One way of putting it is that I subscribe to weak-form rather than strong-form selection, the distinction being that in a relative sense, the fitter survive but there is no reason to suppose that they are fittest in any absolute sense (Williamson 1985, pp. 22-3).

Despite Williamson’s dissatisfaction with empty theorising, and his desire to have his analysis informed by actual case studies tested against alternative hypotheses, he still drifts toward the approach he most definitely has sought to avoid. Like those economists who seem content in dealing with ghosts, using stylised facts instead of plunging into the multicoloured stream of daily life, Williamson also slips into this nether world. It is not in starting with a static model that the trouble lies, but in failing to take sufficient steps beyond it when once the hard work of application begins: ‘The outcome of the method, at its best, is a body of logically consistent propositions concerning the normal relations of things – a system of economic taxonomy. At its worst, it is a body of maxims for the conduct of business and a polemical discussion of disputed points of policy’ (Veblen 1948, p. 227). Static interpretations display a regrettable tendency toward becoming bogged down in a frenzy of Aristotelian categorising. In such a framework, where all potential opportunities remain forever open to key economic players, classification of possible interactions generates a blizzard of definitions: ‘Political economy is said to have strangled itself with definitions’ (J.N. Keynes [1891] 1965, p. 153).

This problem exists to the extent that Williamson establishes, as an economic norm, an equilibrium structure whereby the market adapts form to substance. Williamson spends too much time being categorical, classifying rather than thinking out the dynamics of corporate change. He gives short shrift to explaining how structural arrangements evolve. Williamson does not explain why one particular structure has triumphed. His approach is rather an \( a \ posteriori \) justification for whatever arrangement survives. By this I mean that Williamson is basically content to fall back on the standard micro-analytics of economic advantage in a form that justifies rather than explains. What is missing is any evolutionary notion that current decisions constrain future actions – the idea of path dependency. By postulating his weak form of economic selection, he precludes any serious discussion of that issue. Once the end point is to some degree predestined, the path to it is only of secondary importance. There can be no surprises caused by any actual outcomes, since Williamson can never be wrong. By definition, those firms that survive after those crucial five or ten years of observation are inevitably those firms which minimise transaction costs.

The underlying assumption, indeed the very beginning point for Williamson, is that firms operate efficiently. Bounded rationality prevents any claims of absolute optimality, but still there is a good synchronisation between aims
and accomplishments. Given that peculiar, self-imposed point of departure, the task of the chronicler of such competitive markets can only be to lay bare the mechanism through which such agreement emerges. The transaction cost approach simply becomes an explanatory tool which accounts for any given existing economic organisation. Williamson’s departure is to claim that his particular methods address this question more succinctly and comprehensively than other existing alternatives. We are left to wonder whether at heart Williamson is only performing variations on an existence theorem theme.

But there are no compelling reasons why applying transaction cost analysis should lead a researcher irresistibly to make such assumptions about firms and markets, save for a long tradition in economic theory. Though in general, at any given time, a subset of firms will be adapted to the current economic regime, individually those firms will not necessarily be able to adjust to an unforeseen change in the economic climate. Williamson does seem to be at least partially aware of this, but its importance tends to get lost under the weight of his economic taxonomy. His use of asset specificity implies a trade-off between short-term efficiency and flexibility without explicating or developing it. Williamson subsumes this ability to adapt – what I would term flexibility – under the umbrella of efficiency. This disguises the inherent conflicts lodging within this assumption of efficient outcomes. In a drive to become a least-cost producer, to maintain or enlarge market share under increased competition, cost efficiency, including economising on transaction costs, may be at odds with flexibility. Firms can over-extend themselves, not out of stupidity, but in the desire to achieve present needs measured against uncertain future adjustments. The path that institutional development takes under such constraints need not be optimal, even given a loose definition of that term.

By assuming that firms are efficient, Williamson largely ignores the possible contradiction between the joint goals of efficiency and flexibility. However, his emphatic insistence on asset specificity, taken on its own, is sufficient to indicate that adaptations cannot by definition be frictionless. In Williamson’s own terms, minimising transaction costs includes making the sort of asset-specific investment needed for efficient production. But such commitments can constrain a firm’s ability to adjust to future unanticipated changes. Minimising transaction costs now may interfere with a firm’s ability to do so at a later stage. Accepting the legitimacy of this objective as a general goal will not necessarily yield efficient firms, even under Williamson’s weak formulation. That firms try to reduce costs is a rather unexceptional claim. That such attempts, given bounded rationality, create an inevitable movement toward efficiency is a much more ambiguous and ambitious hypothesis to defend. Williamson asserts rather than argues this link. However, should this assertion prove to be indefensible, our loss would not be particularly large. We would find ourselves instead returning to an examination of rather important relations between corporate structure and transaction costs. This approach should continue to provide increased insights. Whatever value that inheres to such methodology remains largely intact, no matter what efficiency pose is adopted. Why Williamson then insists on making such an efficiency assertion must, at least at first glance, appear baffling to his more discerning readers.

Yet, in Williamson’s work, there is a decided sense in which firms adjust to a pre-existing environment without in turn altering that environment in any unpredictable way. Thus over time, firms, or at least representative firms, approach a level of what he terms ‘comparative efficiency’. Disruptions are restricted to the
familiar exogenous type. But for such shocks, markets would run relatively smoothly. Maladjustment must originate in corporate error rather than from the logic of the market itself.

Given this somewhat gratuitous espousal of market efficiency, Williamson, despite his explicit embrace of bounded rationality, must also fail to take issue with the traditional role that competition has held in economic theory, since competition ultimately drives market efficiency. Retrospective accounts of economic change necessarily tend to be clear, coherent and inevitable, a case of running the film backwards. Based on the proper transaction-cost calculus, firms develop first one organisational structure and then the next. The storms and confusion that swirl around this process disappear. Such deviations become white noise to be effectively filtered by competition over time. Instead, the market proceeds in a stately fashion, moving from strength to strength. Unfortunately, this should be a matter of fact and analysis rather than an article of faith.

Williamson willingly dons the self-imposed straitjacket of predetermined market efficiency favoured by standard forms of economic analysis. By ascribing results solely to controllable individual actions, by portioning out blame and praise, economic analysis of this sort surrounds itself with a certain naïveté. This is perhaps a residue from the past, a direct derivative of the steadfast Victorian belief in growth and progress that still colours much of economic analysis, even that of self-proclaimed institutional and managerial economists. Far from being a dismal science, the true struggle has usually been to squelch the somewhat Pollyanna traditions of the profession.

This firmly-held belief in progress helps to explain the persistent lack of attention paid to a possible trade-off between flexibility and efficiency and thus the uni-dimensional character granted to competition. Williamson implicitly assumes that pursuing one of these goals automatically accomplishes the other. Even if only in a relative sense, the economics trade has dismissed these potential costs of competition for years, either by trivialising them as the mere higgling and haggling of the market place or – what is not inconsistent with that view – equating the consequent lack of resiliency with the competitive price paid for economic growth in a somewhat Social Darwinist context.

If, though, decisions taken today inform a firm’s choices tomorrow, then the ideas behind the transaction-cost approach must be refined. Given the fact that any decision must have an attached opportunity cost, this alternative cannot be simply rejected out of hand. As it stands now, if we accept Williamson’s approach, any successful organisational structure is by definition transaction-cost efficient. Those costs, however, cannot be fully known at the time in which they are incurred. Bounded rationality ensures that, by giving weight to Williamson’s asset specificity problem. The value of flexibility, existing in terms of an uncertain future, which Williamson includes under his idea of efficiency, can only be guessed at, and under competitive pressure may be slighted for more immediate gains. To keep the idea of economising on transaction costs from acquiring a tautological edge, whereby whatever happens is simply identified with this one overall functional goal, the problem of adaptation and flexibility needs to be brought to the fore. To do so requires a more sceptical, though by no means dismissive, approach to the workings of a competitive market.

What I have indicated is that a desire to economise on transaction costs need not be synonymous with Williamson’s weak form of natural selection. No mechanism exists in his work to link these two terms.
market efficiency is not inherent in the concept or use of transaction costs as an analytical tool. The link he makes between the two is rhetorical in nature, imbuing a simple idea with an unnecessary dimension. It is in fact a rhetorical device which deliberately fails to distinguish between two quite separable ideas. The weak or strong assumption of efficiency could easily be disassociated from Williamson’s work. It is not intrinsic to his analysis. Most of his work, his explanation of organisational structure, would be little changed if shorn of such a belief. In which case, this issue is reduced to self-interested banner waving for a preferred idea. Williamson is trying to use a back door to slip in as inevitable a familiar pet assumption.

At first glance Williamson does appear to be critical of standard neoclassical analysis. What he in fact fashions is a different, and what he hopes is a more appealing, rhetorical twist to support the key conclusions of standard economic theory. He is in fact a true believer, rescuing price theory from the clutches of its false prophets. What Williamson has devised is a more sophisticated and appealing way to demonstrate market efficiency.

Notes

1 I would like to thank Kathy Donneson for demonstrating an uncommon enthusiasm for the ideas expressed in this work. Unfortunately, enthusiasm alone is no guarantee of the contents of any paper. Also I would be remiss in not acknowledging the assistance rendered by two idiosyncratic referees who must remain anonymous at least to me if not to themselves.

2 Throughout, I use Williamson’s most representative work, The Economic Institutions of Capitalism (1985). At the time, this work best summed up his extensive writing on transaction costs. Like most economists, Williamson has often surrendered to an ecological urge of recycling old material, making the book a near-compilation of his greatest hits. Though still active today, Williamson has not significantly deviated from the views espoused in that volume.

3 Williamson himself quotes Stanley Fischer’s withering comment on the tautological nature of transaction cost analysis. ‘Transaction costs have a well-deserved bad name as a theoretical device … [partly] because there is a suspicion that almost anything can be rationalized by invoking suitably specified transaction costs.’ (Fischer quoted by Williamson 1979, p. 233)

4 Like Williamson, I will limit my focus to the realm inhabited by firms. However, similar transaction cost analysis is applicable to the consumption side of the equation. If Williamson is willing to assert that firms operate in a relatively efficient way, the same judgement could and should be extended to the whole market if Williamson is to be consistent. Williamson uses ‘relative’ to indicate achievable rather than some sort of nirvana level of efficiency.

5 Williamson may prefer to hedge by eschewing optimally efficient outcomes for the cosier realm of relativity. This, though, resolves nothing. The question then becomes, ‘relative to what?’ The only logical answer is relative to a Pareto-optimal system where transaction costs are zero. In which case, one wonders if Williamson has not substituted one benchmark for another. Given positive transaction costs, market competition yields results that are as efficient as possible. This harks back to Armen
Alchian’s assertion that ‘What is, is efficient’, first demonstrated in his 1950 *Journal of Political Economy* paper.

For instance, I will not attempt to conclusively prove or disprove that a firm’s attempt to minimise transaction costs leads to efficient firm or market outcomes. I will raise an objection but do not intend to present a comprehensive summary of the issue. Those interested in such question might start with the text by Milgrom and Roberts (1992, pp. 33-9). My objective is at least to indicate that Williamson simply makes this assertion of relative efficiency and that it is not intrinsically related to transaction cost analysis.

‘The primitive state of our knowledge is at least equally explained by a reluctance to admit the details of organizational matter. The widespread conception of the modern corporation as a “black box” is the epitome of the noninstitutional (or pre-analytic) research tradition’ (Williamson 1985, p. 15).

Bounded rationality is not even trivially interesting in the absence of information problems. As Williamson (1979, p. 234, n. 5) himself claims, ‘But for the limited ability of human agents to receive, store, retrieve, and process data, interesting economic problems vanish’.

For Williamson, asset specificity revolves around the sunk cost component of any investment: ‘Four types of asset specificity are usefully distinguished: site specificity – e.g. successive stations that are located in a cheek-by-jowl relation to each other so as to economize on inventory and transportation expenses; physical asset specificity – e.g. specialized dies that are required to produce a component; human asset specificity that arises in a learning by doing fashion; and dedicated assets, which represent a discrete investment in generalized (as contrasted with special purpose) production capacity that would not be made but for the prospect of selling a significant amount of product to a specific customer’ (Williamson 1985, p. 95).

Under given circumstances, asset specificity can involve a hold-up problem, giving scope for profitable opportunistic behaviour. See Klein, Crawford and Alchian (1978) for an early exposition of this issue. The hold-up problem is only one issue arising from asset specificity. In general what may seem the most efficient contractual arrangement in a given economic environment may prove detrimental should that environment change in unanticipated ways.

Flexibility measures a firm’s ability to adjust to an unforeseen future, in effect a measurement of a firm’s margin for error. Lack of flexibility can potentially draw on a firm’s resources in terms of direct financial burden and/or adjustment time. Saying that in the long run successful firms must be efficient, which implicitly also means flexible, skirts the issue. One is simply painting bull’s eyes around pre-existing arrows. Flexibility is needed so that firms can reduce the cost of adjusting to unanticipated changes in their economic environment. In effect, resources need to be devoted to achieving flexibility as well as to achieving productive efficiency. But the existence of uncertainty implies that any optimal allocation is impossible. If it were, there would be no need for the sort of flexibility defined here. Uncertainty then calls forth provisions for flexibility while denying the possibility of knowing beforehand what a sufficient level of flexibility defined here might be. The problem of deciding on an appropriate allocation level of resources does not vanish by hypothesising a dynamic concept of efficiency. Instead it muddles the difference between the two resource uses. This glosses over the potential problem.

This lack of testable hypotheses drew measured scorn from the doyenne of Chicago economics, George Stigler:

He (Stigler) wasn’t hostile to that sort of thing (transaction cost analysis) *per se*. It had a very narrow pathway. It’s still true today. And that is a sore point. Williamson himself thinks it has a lot of applications, that would be
his view. But he’s looking at the world in a particular way. If you look at the larger ‘what do we know as a result of this approach’, of course these approaches have a certain amount of application, there are a few things we’ve learned about the firm, but I’m talking relative to the time spent. I mean there are serious folks now, his students, who understand that there is a problem and are trying to fix it. But you know what George would have said. ‘Twenty years! … Or whatever it is … thirty years! hey’ve been working on this stuff. That’s getting to the end of the short run’. Quote, unquote, he would say that (Conversation between Craig Freedman and Sam Peltzman, October 1997).

Kay (1992), in a critical analysis of Williamson’s work, notes a disparity between Williamson’s claims and what he actually delivers in his analysis of corporations. Kay dissects Williamson’s economic rationalisations of the conglomerate and multi-divisional form to illustrate the limitations of a comparative static approach to such matters:

The problems arise when this is offered as a rationale for the evolution of the conglomerate. In this context, Williamson is comparing the wrong things. Instead of analysing why the conglomerate may be superior to the external capital market, transaction cost economics has to explain why it may be superior to the specialized firm. By Williamson’s own chosen criterion of efficiency, there must be some advantage that conglomerateness has over specialization that encourages this strategy to evolve at the expense of the specialized firm (Kay 1992, pp. 324-5).

… Williamson’s theory rests on comparative statics and cannot be used to describe dynamic evolutionary phenomena, despite his claims to the contrary (Kay 1992, p. 330).

While Williamson (1992) concedes the need to make his theory more evolutionary, readers are left to wonder what his vague use of this term is intended to mean. He makes light of whether or not natural selection is used in the standard biological sense, or if what he describes covers a more consciously intentional process where economic agents deliberately try to alter the economic environment to suit their own pre-selected ends: ‘Such an argument (Kay 1992) seems to me to be needlessly deferential to biological reasoning. If economic actors behave intentionally – which, within limits, they do – then why eschew intentionality in the study of organizational innovation?’ (Williamson 1992, p. 348).

The problem however is not one of eschewing intentionality. Such intentionality is a problem if Williamson is deferring to some type of natural selection process to ensure his favoured weak form of market efficiency. The biological approach posits some impersonal source of arbitration that selects winners from losers. However, it is wrong even in this case to conclude that any absolute sense of efficiency or progress obtains. The surviving organisms or structures are simply those best adapted to the prevailing environment in a comparative sense. Williamson’s argument for weak form market efficiency would seem unsustainable once intentionality prevails to any significant extent.

‘The fundamental difference in the case of animal or conscious life is that it can react to a situation before that situation materializes; it can see things coming’ (Knight 1971, p. 200).

‘First, not all future contingencies for which adaptations are required can be anticipated at the outset. Second, the appropriate adaptations will not be evident for many contingencies until the circumstances materialize’ (Williamson 1985, p. 70).

Firm efficiency in the short run is undermined by Williamson’s (1964) own early work on discretionary managerial decision-making. Add this to the idiosyncratic nature
of those decisions involving asset-specific investments and it is difficult to see that a competitive evolutionary process will necessarily yield even relatively efficient firms or market outcomes five or ten years down the track. This may in fact be the result, but an emphatic assertion that this is the case does not make it so.

15 See Freedman (1991) for a discussion of why this issue (the trade-off between efficiency and flexibility) has been relatively ignored in the past.

16 ‘The survival of an organisation depends upon the maintenance of an equilibrium of complex character in a continuously fluctuating environment of physical, biological and social materials, elements, and forces which calls for readjustment of processes internal to the organisation’ (Barnard quoted in Williamson 1985, pp. 5-6).

17 Williamson acknowledges the need to delineate the mechanism which creates this desired connection (1985, p. 23). But any such attempt would undermine the very thing he wishes to demonstrate. The economics profession still awaits this missing demonstration.

18 There is a subsequent pointed exchange (Journal of Economic Perspectives, Winter 1993) between Williamson, Stiglitz, and Bowles and Gintis concerning the status of Williamson’s concept of ‘comparative efficiency’:

Since Williamson recognizes that a competitive equilibrium does not satisfy the Pareto criterion when enforcement is endogenous, he is obliged to rely on a more Darwinian sense of efficiency: institutions are deemed efficient if they survive in competition and if superior alternatives cannot be found. The inference that survival entails efficiency is unwarranted, for it ignores the path dependent nature of evolution and the possibility of multiple equilibria. In any model with multiple stable equilibria, biological or economic, where you end up depends on where you’ve been, and whatever optimality properties may be claimed for the equilibria are at most local rather than global (Bowles and Gintis 1993, p. 97).

He [Williamson] holds that institutions emerging from the competitive process will be ‘comparatively’ efficient. But compared to what? Surely a voluntarily signed contract improves the lot of both parties, but that does not demonstrate that the general equilibrium of the economy is Pareto efficient. The theorems referred to earlier, establishing that markets are not constrained Pareto efficient, show that there is no general basis for Williamson’s assertion (Stiglitz 1993, p. 97 n. 4).

In his defence, Williamson looks to waffle to see him through: ‘I also agree with the proposition that evolutionary processes are path dependent. That does not preclude comparative efficiency assessments of alternative feasible forms, but some forms may be privileged by their evolutionary origins’ (Williamson 1993, p. 104).

The point I am making is that, whether the claim is groundless or not, it is a statement of faith that adds nothing to Williamson’s analysis. Remove it and nothing is lost. Or rather, nothing is lost except a half-hearted rhetorical cheer for market efficiency.

References


