Abstract
Several apparent advantages of vertical integration (i.e., capture of interior margins) disappear when carefully compared to the options available to non-integrated firms. The economic rationale for a firm to act as a substitute for the external capital market, in particular, has received serious criticism (Gertner et al. 1994; Billet and Mauer 2003), since independent firms can replicate the capabilities of mutual ownership using contracts except in very specific situations (e.g., asymmetric information, trade secrets, or an irregular funding pattern (Liebeskind 2000)). This paper argues that this view is too simplistic to capture the coordination benefits of vertical integration, particularly in selecting which customers to serve when both manufacturing and financing stages are required to consummate a sale. Previous analyses implicitly assume that coordination between the independent parties has already been accomplished (e.g., through negotiation or the price mechanism) and thus the integrated firm had no opportunity to demonstrate superior coordination. By beginning the analysis at the project-selection stage (i.e., which customers to serve, and under what terms), however, we show that superior coordination via vertical integration (suppressing the incentive for one independent unit to reject an opportunity that the other would accept) increases the range of projects that can be profitably undertaken, thereby increasing the quantity that can be provided without compromising the integrated firm’s value or cost positions (which determine its margins.) Such coordination ability therefore leads to higher returns on assets or equity vis-à-vis other competitors who are otherwise identical except for their organization.
The strategic management literature on vertical integration emphasizes that such integration has both benefits and costs. Several apparent advantages of vertical integration (i.e., the capture of interior margins from acquiring profitable downstream businesses) turn out to be illusory when carefully compared to the options available to non-integrated firms.

The economic rationale for a firm to deliberately act as a substitute for the external capital market, in particular, has received serious criticism (Walker 2008: 330; Gertner, Scharfstein and Stein 1994; Billet and Mauer 2003). Indeed, the ability of a firm to provide capital services to its component businesses is held to be necessary, but not sufficient, for these firms to compete effectively; particularly in financial markets, independent firms can replicate the capabilities of mutual ownership using contracts except in very specific situations such as asymmetric information about repayment prospects, trade secrets subject to inadvertent disclosure, or a sufficiently irregular funding pattern that precludes external capital investment (Liebeskind 2000; Walker, 2008: 310). Otherwise, independent producers that take advantage of competitive capital markets ought to perform no worse than those which utilize internal financing of their operations and/or their customers’ purchases.

This paper argues that this view is too simplistic to accurately capture the coordination benefits of vertical integration, particularly in the selection of which customers to serve and in the timing of capital transfers. Previous analyses of integrated vs. separate business units implicitly assume that coordination between the independent parties has already been accomplished (e.g., through negotiation or the price mechanism) and that no superior coordination can be demonstrated by the integrated firm. By beginning the analysis at the project selection stage (i.e., which customers to serve, and under what terms), however, we show that this coordination cannot be taken as given and that the vertically integrated firm can improve its financial performance vis-à-vis otherwise identically-configured but independent firms. We conclude by speculating that this superior coordination leads to a new form of competitive advantage (arising from the “governance” theory of profitability, as in Makadok (2011)) that, by increasing the range of projects that can be profitably undertaken and increases, effectively increases the quantity that can be provided (e.g., asset turnover) without compromising the firm’s value or cost positions (which determine its margins.) Such coordination ability therefore leads to long-term improvement in return on assets or equity vis-à-vis other competitors who are otherwise identical except for their organization (integrated vs. independent), which is the hallmark of competitive advantage achieved through organizational structure, even though it does not take the specific form of either increased value or reduced cost.

**Analogy with Multiple Marginalization**

An old (and mercifully brief) economist joke runs “What’s the only thing worse than a monopolist? TWO monopolists linked together in a chain.” This problem, which results from an input being priced at over its marginal cost and then “re-marked-up” by the eventually seller, is known as “double marginalization.” For a single firm that controls both the intermediate good and the final good (sold to the consumer), organizing both business units as profit centers falls unnecessarily into this double-marginalization trap. The problem with double marginalization inside a firm is twofold: (1) the final price to the consumer is not profit-maximizing, as the inclusion of the interior margin from the markup of the intermediate good adds an error to the
usual profit-maximization calculation faced by the downstream unit; the downstream business unit optimizes profit taking its costs as given, under the mistaken assumption that marginal cost to the firm is the same as the transfer price it receives from the intermediate-good producing business unit, whereas the true marginal cost to the firm is less (by the magnitude of the interior markup.) Optimizing based on incorrect cost data leads to incorrect prices to the end customer and (compared to the first-best prices) a subsequent reduction in overall firm profitability, (2) Some opportunities for the firm to make a profit (i.e., to sell additional quantity) are foreclosed to it because it is pricing too high; some consumers are “priced out” of the market who would otherwise consume the monopolist’s product. These customers have willingness-to-pay which is higher than the firm’s marginal costs of serving them; while some of these customers would be excluded by a single markup (leading to the classical deadweight loss from monopoly), still more of them are excluded by the additional markup caused by the interior marginalization (which effect is exacerbated by each additional layer of markup). Such missed opportunities to transact with customers are a second-order reduction in firm profit, as revenue from them would otherwise be captured by the firm at the lower (single-marginalized) price. In effect, the two independent firms are missing out on profitable projects to undertake together – not because of any lack of information, any capacity constraint, or inherent unprofitability of these excluded customers, but because their own pricing strategy prevents them from coming to an agreement with these potential buyers.

Instead of each business unit naively applying a markup over cost to its own outputs, the firm could increase its overall profits (and, coincidentally, social welfare) by suppressing this double marginalization internally -- in effect transferring the intermediate products at their marginal costs and applying a single optimal markup when the product finally leaves the firm and is sold to the consumer. Spengler (1950) showed (using a graphical technique) that the suppression of double marginalization also improved social welfare; since double marginalization restricts quantity even more than monopolistic (single) marginalization, it induces a higher level of deadweight loss. Croson (1996) demonstrated (using calculus) that this problem of multiple marginalization is exacerbated when more than two actors sequentially mark up their products -- leading to severe quantity restriction and accordingly higher deadweight loss from allocative inefficiency -- and that incentives for investing in cost reductions within the firm or value chain are severely attenuated by the double (or triple) marginalization activities on the revenue side.

Both Spengler (1950) and Croson (1996) noted that the advantages of coordination through vertical integration led to profitability improvements (in pricing and cost reductions, respectively) that were independent of the value-generating abilities of the underlying product or the specific customers to which it was sold. We will now argue that an analogous effect – of independent companies failing to consummate transactions with otherwise profitable potential buyers – can be overcome by vertical integration, even when prices are held fixed; when prices may vary, we argue that the effect is even stronger.

General Electric and GE Capital vs. Independent Electric and Independent Capital

General Electric (GE) is well-known for managing its disparate businesses using a strategic business-unit (SBU) system, in which each business unit operates as its own entity but coordinates with other SBUs on joint operations such as selling complex products to customers.
Walker (2008: 312) notes that GE, a much-studied firm [Aguilar and Hamermesh (1981); Bartlett and Wozny (2005)] has avoided the diversification discount common to other multibusiness firms (Berger and Ofek 1995; Villalonga 2004): “GE’s remarkable valuation as a corporation suggests strongly that it contributes significantly to its businesses...it also focuses intensely on the value it adds as a corporation to its current units.”

Consider the following situation. Imagine that GE engages both in the manufacture of an industrial product with high gross margins and high barriers to entry (e.g., nuclear reactors) and in the financing of large-scale industrial projects (through its subsidiary GE Capital.) GE is not a monopolist in either area; however; a rival (“Independent Electric”) produces identical nuclear reactors and a financial firm (“Independent Capital”) offers financing of large-scale industrial projects, but neither outside firm offers both services under the same corporate aegis. The market price of a nuclear reactor is $1 billion and its variable cost is $500 million, leading to a 100% gross margin percentage (even after accounting for the price rivalry of having two firms in the market) as measured by the Lerner index. The finance market is much more competitive, however, to the point that a stand-alone finance firm earns only normal returns on capital investment according to the riskiness of its portfolio. Pricing of the financing is based both on time value of money and a risk premium that compensates the lender for the risk of default; while the exact magnitude of this premium (as a function of the amount at risk) is not particularly relevant, it has the property that if twice as much capital is put at risk, at least twice the risk premium (on an absolute dollar basis) must be paid to make the deal individually rational for the financier to participate. Further assume that the end customer puts down 10% of the purchase price of a nuclear reactor and wishes to finance the rest ($900 million).

Now consider a buyer who wishes to purchase and finance a nuclear reactor who has a willingness to pay in excess of $1 billion but who, compared to other potential buyers, has a relatively high risk of default – e.g., 25% or 50%. Are the two methods of industrial organization (vertical integration vs. separate) equivalent for the profitable servicing of such a customer and, if not, which structure does this risky investment favor?

GE, as an integrated company, faces a different gamble than does the uneasy, arm’s-length alliance of Independent Electric and Independent Capital. From Independent Capital’s perspective, there is either $900 million at risk from a customer’s failure to pay (the original $1 billion purchase price, less the $100 million downpayment) or the prospect of being required to “claw back” hundreds of millions of dollars from its client, Independent Electric, in the event of default – all for zero expected profit! From GE’s perspective, only $400 million is at risk from any given inframarginal customer; the $500 million of production cost is offset by the $100 million downpayment, leaving $400 million at risk (and $500 million of profit to be gained from a successful repayment by the customer.)

In the presence of any transactions costs of clawback whatsoever and any difference in the cost of capital of the two independent firms, vertical integration is clearly dominant. GE faces a gamble which is first-order stochastically dominant (indeed, statewise dominant, as in Hadar and Russell (1969)) over that faced by the IE/IC alliance; with probability \( p \), GE gains $500 million and with probability \( 1-p \) loses $400 million, whereas IE/IC with probability \( p \) gains $500 million (minus any costs of misallocated capital, as described below) and with probability \( (1-p) \) loses...
$400 million between the two of them, which is distributed between IE and IC according to the terms of their insurance/clawback agreement with some nonnegative amount of transactions cost incurred along the way. Each percent of default risk lowers General Electric’s expected profit by $4 million (1% of the unpaid production cost, after accounting for the downpayment from the buyer.) Assuming the entire transfer from Independent Electric to Independent Capital has been collected as a fixed insurance premium (X), each percentage of default risk lowers Independent Capital’s expected profit by $9 million (1% of the $900 million loan.) Our intuition tells us accordingly that there may be high default-risk customers that must be served by the integrated firm.

In the absence of any clawback provision, it is clear that Independent Capital must collect not only the pure time value of the $900 million to be lent but also a risk premium proportionate to the entire loan principal amount. If a certain portion of the buyer’s payment to Independent Electric (funded by the Independent Capital loan, but paid to Independent Electric as revenue) can be “clawed back” in the case of buyer default, it is possible to share the risk of nonpayment between Independent Electric and Independent Capital in the ratio of 5:4; even if Independent Capital retains the entire $100 million downpayment and can claw back any profits paid to Independent Electric, it must still bear the risk of $400 million of its capital being lost. Instead of retaining the amount subject to clawback in a segregated account (which is inefficient, since presumably Independent Capital’s cost of capital is lower than that of the manufacturing firm, given their relative specializations), Independent Electric may wish to employ part of all of this capital in the operations of its business (treating it as equity and not an asset offset by a contingent liability) at which point Independent Electric must pay Independent Capital an insurance premium in lieu of a pledge to repay this capital in the event of a buyer default -- effectively, the actuarially fair insurance premium representing protection against loss of somewhere between $400 million and $900 million.

Sketch of Formal Model

Consider a product with marginal cost $c$ to produce which sells for $R$ with a downpayment of $d$ delivered at the beginning of period 1. The probability that the client will default is all-or-nothing; either $R-d$ or 0 is repaid at the end of period 1, with probability $p$ and $1-p$ respectively. The risk-free rate of interest is assumed to be 0 for convenience and all parties are risk-neutral; to completely hedge themselves against risk, they need merely receive the expected value of any such loss. Contracts between an independent seller and an independent financial-services firm are of the following form: a premium of $X$ is paid at the time of sale (i.e., at the beginning of period 1) and/or a certain amount of capital $B$ is set aside as being vulnerable to clawback in case of buyer’s default; without loss of generality this can include the downpayment $d$, such that $B \geq d \geq 0$. The capital market is assumed to make normal returns (as in perfect competition) such that the expected profit of the capital provider is normalized to zero after accounting for its cost of capital and the actuarial expectation of the risk of loss. Thus, the payoff to the capital provider is $X + (p)[d + (R-d)] + (1-p)[d + (B-d)] - R$.

The values of $X$ and $B$ may vary so that the zero-profit condition is maintained. Implicit differentiation of this payoff function yields $dX/dB = -(1-p) = (p-1)<0$ which formalizes the
intuition that a risk with probability \((1-p)\) and a consequence of -$1 can be exchanged for a \((1-p)\) up-front payment.

The payoff to the independent producer is \((R-c) - X - (1-p)[B-d] - Bk\); note that the product market is imperfectly competitive and thus this expected profit need not be driven to zero, reflecting rents from market position, VRIO resources, etc. (see Makadok 2011 for a survey of alternative mechanisms). In particular, \(R\) is the outcome of the competitive dynamics between rival firms rather than a mechanical accrual of costs and risk premia in the (assumed perfectly competitive) financial market. The manufacturing firm earns a positive economic return on capital of \(k\), over and above the risk-free rate, for every dollar invested in its superior competitive position (at least for small, local increases or decreases in capital base that do not involve substantial changes in scale or scope.)

When choosing whether to pay an insurance premium \(X\) to the financier or whether to self-insure (by segregating capital into a liquid escrow facility, which can be clawed back without significant transactions cost), the producer seeks to maximize its expected profit given the probability \((1-p)\) of buyer default.

Consider the combination \((X^*, B^*)\) which maximizes the producer’s profit subject to satisfying the financier’s requirements for security against default. Differentiating its profit function with respect to \(B\) yields the first-order condition for profit maximization for the producer: \(\frac{dP}{dB} = -\frac{dX}{dB} - (1-p) - k\). Since the financier’s requirements (analyzed above) show that \(\frac{dX}{dB} = -(1-p)\), the producer can thus maximize its profits at any combination of \(X\) and \(B\) that satisfies the financier if \(k=0\); otherwise it strictly prefers to pay up front and chooses \((X^*, 0)\) to avoid having \(B\) “tied up.” If \(B>0\), some capital will be necessarily misallocated in the producer’s operations: monies that could have generated value commensurate with the firm’s overall return on equity would instead be constrained to reside in a liquid vehicle subject to clawback – in effect, in escrow pending the realization of the buyer’s payment or default.

While a more detailed model that examines the nonlinear relationship of opportunity cost (i.e., the exact schedule of marginal opportunity costs of misallocated capital, as a function of capital already in place) could yield a specific interior solution, for this paper’s purposes it suffices that \(B\) and \(X\) are either interchangeable according to the \((p-1)\) ratio with respect to both their effects on the producer’s profit and on the financier’s willingness to provide financing or the producer strictly prefers to pay \(X\) and to set \(B=0\).

This substitutability of the implicit cost from escrowing for the explicit cost of paying the insurance premium makes clear that the producer does not save money by earmarking \(B\) in lieu of paying \(X\): indeed, \(X\) ought to be counted as a cost to the independent producer, whether it is incurred directly (in the form of a payment to the financier) or indirectly (in the form of being required to segregate \(B\) from its equity capital in order to be conspicuously available for clawback in case of buyer default.) \(X\) is thus just as much a cost as any capital, labor, energy, or materials input required to produce the underlying industrial product; without the producer incurring \(X\) (or its equivalent \(B\)), the sale cannot be consummated. This already suggests that the necessity of incurring \(X\), which itself is an up-front payment and thus not available for
investment at rate \( k \), can conceivably be an implicit cost advantage for a manufacturer who does not need to pay it.

**Graphical Analogy of Formal Model**

Consider a distribution of customers characterized by differences in \( p \) as well as in their value (in the sense of willingness to pay) for the product. Competition for their custom will result in these customers paying an amount which is decreasing in both \( p \) and \( V \). Certain combinations of \( V \) and \( p \) will generate acceptable projects to Independent Electric, Independent Capital, both, or neither. The intersection of the two acceptability zones – i.e., those deals which are acceptable to both Electric and Capital – represents deals which can be consummated. Some high-risk (low \( p \)) but high-value (high \( R \)) customers will fall outside the financeable zone, however – analogous to consumers who need a car to commute to a high-paying job but who have no credit.

As long as \( V \) exceeds \( c \) and \( p \) exceeds a particular threshold as a function of \( R \), all deals are, in fact, profitable in an expected-value sense and therefore should be done. The difficulties in making them occur arise because of the risk allotment between the two companies and the challenges of ex post recovery, not the underlying merit of making the sale. (It is, of course, precisely because vertical integration economizes on the costs of risk allotment between the two business units and the challenges of ex post recovery that it has an organizational advantage in such a situation.)

Now consider what changes in the vertical-integration situation. By unifying the decision, must satisfy only one simple condition instead of two complex ones: now \( p \) must be at least \( (R-d)/(R-c) \). All deals in which \( V \) exceeds \( c \) and \( p \) exceeds this threshold will now be acceptable to the integrated firm, ensuring that they can serve any customer which can be profitably served at all. In particular, customers with high \( R \) and high \( p \) can now be served, whereas previously they failed the financeability test. The advantage of vertical integration as a mechanism to price-discriminate was also previously treated in Perry (1980, 1988) with respect to Alcoa, who (to avoid resale issues) elected to integrate into the service of customers who would otherwise not pay higher prices but to use market relations to serve predictably low-value customers.

Adding this wedge of incremental customers to the previous market achieved by the two-constraint system means that a strictly larger set of potential customers that can be served by the vertically-integrated team than the separately-organized units. Such a proportional increase in quantity (turnover, in the DuPont system framework treated in Kline and Hessler (1951)) has the same effect on profitability as a proportional increase in margins. We note that existing “V-C” theories of competitive advantage (Peteraf and Barney 2003; Postrel 2006; Walker 2008) do not recognize this increase in turnover as a legitimate criterion for concluding that the manufacturing arm of the integrated firm has a competitive advantage over the independent manufacturing company, whereas an increase in margins would immediately be accepted as such – an inconsistency similar to that noted by Amit and Livnat (1990) in their analysis of firm- vs. industry-level profitability.

Vertical integration can thus improve profitability by increasing the quantity of customers who can be profitably served. Once default risk reaches a critical level (which occurs in the example
at p=4/9, or approximately 56% risk of default), however, the project cannot be financed under either form of organization as the expected revenues from the client do not cover the construction cost of the product. Given that the financial component does not itself earn extraordinary returns (whether independent or integrated), no cross-subsidy is possible.

**Empirical Consequences of Capital Misallocation in Independent Firms**

If this cost were an explicit component of the producer’s income statement – analogous to a franchise fee or a license to a critical patent required to operate -- we would have no difficulty in noting that this cost advantage would constitute a competitive advantage for the manufacturer portion of a vertically-integrated firm. [Of course, it does not make sense to speak of a competitive advantage for the financier, who can earn only normal profits.] Any misallocation of capital through a choice of $B>0$ generates a cost (through the deadweight loss from allocative inefficiency) that must be borne by the independent manufacturer but not by the vertically-integrated producer.

Given that the difference in performance caused by inefficient capital allocation is hidden as an opportunity cost with an unobservable counterfactual (higher performance by Independent Electric in its absence), empirical verification is problematic. This proposition would gain support if independent producers displayed persistently inefficient capital allocations, for example, or if vertical integration occurred more often in settings with higher risk of buyer default. As the probability of buyer default increases, the more incentive there is to pay an explicit risk premium ($X$) rather than misallocate large amounts of capital ($B$). Similarly, as the length of time between sale and the resolution of the buyer’s uncertain payment increases, the more attractive it will be to pay rather than escrow, since the payment is made only once and is based on the probability of default (indeed, the payment may be less than its face value, since it can be invested by its recipient pending the outcome of the uncertainty) whereas the opportunity costs of escrowing capital occur yearly.

Certainly US automotive manufacturers have displayed a historical tendency to co-organize the manufacturing and financing of consumer auto purchases (particularly in the high-risk “buy here, pay here” segment), whereas homebuilders have generally used an arm’s-length contracting process for financing the sales of their products. At the other extreme, third-party financing at highly competitive rates is available for large-scale purchases of Treasury bills (often into the tens of billions of dollars).

**Conclusions, Extensions, and Opportunities for Future Research**

As noted above, a constant opportunity cost of capital to the manufacturing firm is a simplification that assumes that any given sale is small compared to its capital base. Perhaps a more detailed model that examined the opportunity cost of removing equity capital (which would, unsurprisingly, equal the return on this equity) would lead to an interior solution that represents the cost-minimizing combination of a partial payment of $X$ and a partial segregation of $B$. For the purposes of this paper, however, it suffices to show that $X$ is a required payment to the independent financier which is not required to the manufacturing arm of the integrated firm. It is not $X$ per se, but the method of paying $X$, which is the cost advantage; $X$ is transferred from
producer to financier, but the same amount is lost to Nature through the probabilistic default, and thus is correctly classified as a cost to the producer-financier system whether integrated or not. Similarly, the inefficient allocation of any capital segregated into B which cannot be deployed by the producer as equity capital is a deadweight loss rather than a transfer.

Under the assumption that the financial component does not itself earn extraordinary returns (whether independent or integrated), no cross-subsidy is possible. Note that if, instead of the actuarially-fair rate of (1-p) per $1, the insurance premium is priced at a higher rate (i.e., such that the financier and producer split, in some mutually agreed-upon proportion, the benefits from avoiding the inefficient escrowing of capital) then vertical integration becomes even more efficient because of suppressing multiple marginalization, as in the original Spengler (1950) analysis. In addition, the excess insurance premium would represent an additional cost to the independent producer, reducing its competitiveness in the product market -- a competitive disadvantage even if no cross-subsidy (beyond actuarially-fair pricing) is implemented in the integrated firm.

We believe that framing this use of the complementary service of providing capital to finance purchase of the primary industrial outputs of a firm can justify vertical integration even in the absence of information asymmetries, explicit transactions costs, or other classic motivations for vertical integration. Characterizing the internalization of financing as a remedy for a dispute over the distribution of economic gain (in terms of selection of which projects to pursue rather than the usual conflict over price determination, as in double marginalization) highlights the control problem for independently-organized and separately owned members of the supply chain and allows the analysis of an implicit cost-based competitive advantage, even though the variable costs of production are identical across the two organizational forms. Whether a deterministic set of outcomes could always be efficiently accomplished through separate ownership (or whether risk is either necessary or sufficient to motivate integration) is a topic for future research.

References


Acknowledgements: We thank Gordon Walker and Steve Postrel for illuminating (and amiably contentious) discussions on this matter. MBA Students in multiple sections of the SMU Cox School of Business STRA 6202 (Advanced Strategic Management) class suggested key refinements to the example presented herein during discussion of two General Electric cases [Aguilar and Hamermesh (1981) and Bartlett and Wozny (2005).] All mistakes, errors and omissions are our own. The first author owns shares in General Electric.