Supplemental Readings in Strategy

Simon Rodan

Supplemental Readings in Strategy

Simon Rodan

Professor of Strategy and Organization Theory,
School of Organization and Management,
College of Business,
San Jose State University,
One Washington Square,
San Jose CA 95192-0070

About the author: Simon Rodan grew up in the UK, attending Imperial College of Science and Technology where he read physics. He worked for IBM for 8 years, 5 as a systems engineer and 3 in sales. He has an MBA and a PhD from INSEAD and after spending three years as a visiting scholar at Stanford, has been teaching at San Jose State University since 2002.

Introduction

This set of readings started life as a series of individual handouts to supplement the standard strategy text book. Some were topics that simply weren't covered, others seemed confusing and some called for a more slightly more detailed treatment.

The first three chapters provide some background to strategic thinking. Chapters 5 through 13 deal with different analytical frameworks (or theoretical models). These are a way of focusing attention on particular aspects of a firm's situation that research has suggested are important to a firm's financial performance. The last few are only tangentially related to strategy but reflect some of my research interests and influences[1] and may be helpful to managers in their careers.

It is customary to note that all errors and omissions are mine, and since I'm a terrible proof reader of my own writing, I am fairly sure you will find many that I have missed, for which I apologize in advance. If you find any, please let me know.

---

[1] In particular the work of James G. March.

Contents

Chapter 1 – What is strategy?

The Oxford English Dictionary defines strategy as "a plan for successful action based on the rationality and interdependence of the moves of the opposing participants"[2]. It's a simple and clean definition, highlighting three things; that the aim is to be successful, that there are other competing actors whose actions need to be taken into account and that all actors behave rationally (at least from their perspectives).

Henry Mintzberg (1987) suggests that strategies come in five flavors. First, a strategy might be seen as a plan of action. Second, it's a ploy to outmaneuver and out-wit competitors. Third, it might be the selection of a position, for example the niche or industry in which you have chosen to compete. Fourth, it might simple be a pattern of consistent and coherent actions that only becomes apparent in hindsight. And finally, strategy may be seen through different perspectives; alternative framings of the problem may reveal a strategy. There is no reason that a strategy can't be several or all of these; a perspective might show a new ploy or suggest the strength (or weakness) of a particular position. And a ploy might be planned in advance, or be seen as a pattern in retrospect.

Broadly, there are three pillars on which the development of good strategy depends. The first is strategy formulation. Good strategy is an act of creation, and imagination. This is one of the least well researched (and almost never taught areas of strategy). It is about thinking creatively and strategically. There has been some debate as to whether strategic thinking can be taught. Some maintain it can't and so don't try. Others think it can. I remain agnostic but there are tools and exercises that I believe can help in developing this capability.

Third is strategic analysis. This is the meat and potatoes of most strategy courses. Here you'll find the usual line up of analytical frameworks: Michael Porter's Five Forces model for industry level analysis, the Resource Based View for firm level analysis, Generic Strategies for a product-market perspective, Transaction Costs Economics for a transaction level perspective, and Game Theory with its focus on strategic interactions. All are useful in helping determine whether your strategy will work as intended, but they don't give you any easy solutions if the answer is 'no'. Indeed that's partly the point (and central to the RBV); if there were easy answers it would be hard for any one firm to develop a unique and defendable competitive advantage.

Finally there is the question of implementation, something that all too often has been ignored by strategy developers. But a strategy that can't be implemented will likely fail to deliver on its original promise. Lack of attention to implementation has led to the failure of many strategic change initiatives (and created a great deal of business for consulting firms).

---

[2] Kenneth Andrews (1980) developed a much longer definition: *"Corporate strategy is the pattern of decisions in a company that determines and reveals its objectives, purposes, or goals, produces the principal policies and plans for achieving those goals, and defines the range of business the company is to pursue, the kind of economic and human organization it is or intends to be, and the nature of the economic and non-economic contribution it intends to make to its shareholders, employees, customers, and communities".* While this definition is general and comprehensive, I find it too long to be memorable.

Chapter 2 – Who runs a public company?

You may have heard your manager say "this is not a democracy" when a decision they are making is questioned. But in an odd way, public companies are democracies; it's just that not everyone gets to vote.

Public companies are "owned" by its shareholders. In return for the money they invest, shareholders are generally guaranteed only one thing; a vote in the election of members of the company's board of directors. That vote matters because the directors appoint the Chief Executive Officer (CEO). So indirectly, shareholders, through their choice of directors, influence who is chosen to run the company and thus how the company is run. Why are voting rights granted to shareholders? The reason may in part be historical; when company owners began to appoint managers to run the company on their behalf they were clearly not about to relinquish all control over the company they owned. The owners, as the "residual claimants" were most at risk from the consequences of the decisions made by the manager they appointed. And as they sold an interest in the company to others, those people being similarly exposed to the risks of decisions made by the company's managers, also wanted a say in who was appointed to run the company in which they now had part ownership. So shareholders get as many votes as they have shares in the company, reflecting their degree of risk exposure. While some companies do sell stock without voting rights this is far from the norm for publicly traded companies. And non-voting shares trade at a discount to those with voting rights. One reason for this is that fund managers who collectively own the vast majority of shares in public companies, would be less willing to invest in companies that did not afford the option for them to oust a poorly performing CEO. Although "activist investors", those who take an active role in prodding a CEO to act in one way or another are in a minority, their existence illustrates that shareholders, at least those with a sizable stake in the company, do have more power than they would if their only recourse with a poorly performing manager in a company they owned was to sell the stock.

Public companies hold meetings annually to which shareholders are invited and at which they vote on resolutions; these will always include resolutions to appoint or reappoint members of the company's the board of directors, the company's auditor, and possibly resolutions regarding issuing or repurchasing the company's shares. But any shareholder[3] can propose a resolution for consideration at the annual shareholders' meeting[4], although these are non-binding.

So to recap, shareholders have a vote, one per share so the more shares you own the 'louder' your voice, and they vote each year to appoint or appoint the company's board of directors. The board they decides (also by a vote of the board members) whether to reappoint the company's top manager, its CEO. So in a sense, companies really are a democracy.

There is a branch of economics that deals with the issue of how shareholders' interests are protected called Principal Agent Theory. The Principal refers to the owner or shareholder, and the Agent the person appointed by the Principal to act on their behalf. The problem Principal Agent Theory looks at is

---

[3] Who has owned at least $2,000 of the company's stock for at least one year.

[4] Subject to the agreement of the chairman of the board to add the resolution to the meeting agenda

how to prevent self-interested agents acting in ways that benefit them rather than the Principal by whom they were appointed and for whom they are supposed to be working. Examples of the Principal-Agent problem abound; not only do public companies have this problem, so do credit rating agencies (e.g., S&P, Moodies, Fitch) who are paid by the companies whose bonds they are supposed to be rating, or investment banks advising their clients on an acquisition but who are paid based on the value of the deal (about which we will say more in the chapter on acquisitions).

Chapter 3 – Value Creation and Appropriation

Value, Price and Cost

Thinking about business in terms of just these three concepts, value, price, and cost, can be illuminating because it strips away the much of the complexity and pares a business back to the fundamentals. Cost is just what it says, the cost of making and delivering a product or service. Price is what the buyer pays for the product. Value is the worth the buyer places on the product thus setting the highest price the buyer will be prepared to pay. Value is important to the seller since it sets an upper limit on price, just as cost sets a lower limit on price, at least if they are to make a profit.

Value creation

When you buy a product, you generally exchange money for a product or service. In any such transaction or exchange, you are (or should be, assuming the decision was rational) happier after you have completed the transaction than you were before when you had the money. You may have been prepared to pay more than you did, meaning the value you ascribed to the product is higher than its price. The seller has received money (the price paid) which exceeds the cost of creating and selling the item. This representation of economic exchange – which by the way, ignores some serious problems including asymmetric and incomplete information, bounded rationality, and other cognitive limitations and or biases – is a fundamental idea in economics.

Figure 1
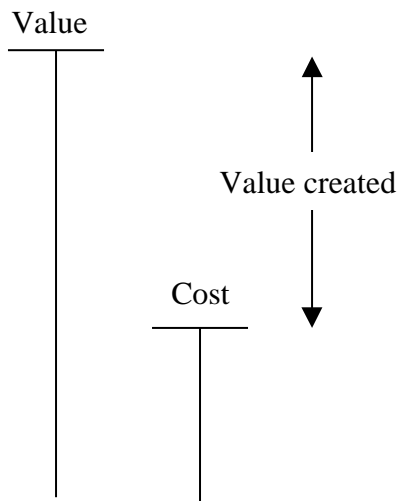
The *value created* in an exchange is the difference between the seller's cost and the buyer's valuation of the product (the *highest* price the buyer would have been prepared to pay for the product). Every economic transaction releases a potential store of value. In other words, exchange creates economic value. This is illustrated on the left (Figure 1). Value cannot be less than Cost; if it were the product would not be produced.

The value created is difference between the value buyers ascribe to the product and the cost of producing it.

Next we need to consider the price at which the transaction is made to see how much of the value created each of the parties to the exchange (buyer and seller) receive.

Value appropriation

The value captured or *appropriated* by the seller the difference between the price charged and the cost of production and delivery (Figure 2). This is essentially a firm's profit. If the price is lower than the value placed on the product by the buyer the seller only captures some of the value created. The

remainder, the difference between the buyer's valuation and the price, is the value appropriated by the buyer (economists refer to this as *consumer surplus*).

Value

Price

Cost

Value appropriated
by the buyer

Value appropriated
by the seller
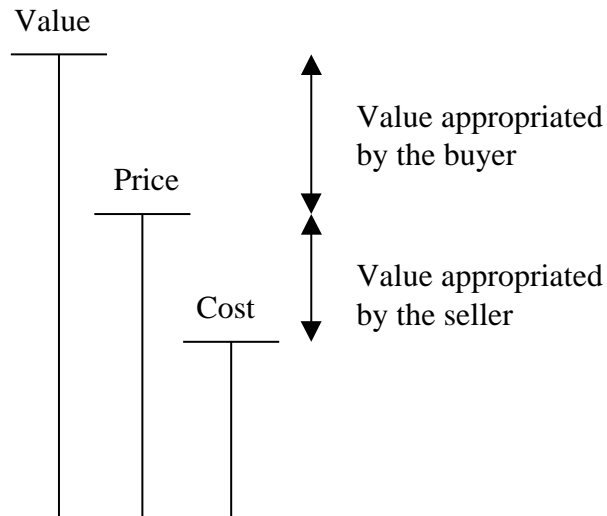
In a *perfectly competitive market*, the price paid by buyers will be exactly equal to the long run average cost the producers incur to make the product. That's the definition of a perfectly competitive market; the value appropriated by the firm is zero – their receipts are exactly equal to their long run average costs. Buyers, on the other hand, do well in perfectly competitive markets since they capture all the value created in the exchange.

Most industries, however, are not perfectly competitive; the majorly are concentrated with a relatively small number of large firms. In the extreme case, a monopoly with *perfect price discrimination*, the monopoly seller charges each buyer a price exactly equal to the highest value that buyer places on the good or service. In each transaction, the price charged is exactly equal to the highest price that buyer is willing to pay and buyers gain nothing while the firm appropriates all the value created. Normally, even in monopoly markets, price discrimination is not perfect so there is always some consumer surplus. Most product markets fall somewhere between the two extremes of perfect competition and perfect price discriminating monopolies. Most have several sellers and none are able to appropriate all the value from the products they sell.

The key take-aways from this model are that 1) value is created when a product's value is greater than the cost of its production 2) if no value is created ($V<C$) the product will (or should) not be produced and 3) price, which should like somewhere between cost at the low end and value at the other extreme, is what determines how much of that value is captured or appropriated by the each of the parties, the buyer and the seller.

[It is also worth noting that along the value chain, a firm's cost is its supplier's price and the value of its product to its buyer is exactly the price that buyer can get from its customers.]

Chapter 4 – Strategy: Analysis, Formulation, Implementation

Crafting strategy involves four different kinds of task. The starting point is the developing of a comprehensive understanding of the current situation, from the details to the big picture; this is "domain knowledge". Domain knowledge is not just about knowing 'what is', but also who it all work, how the pieces articulate. That matters, because second task, formulation or the creation of a strategy, requires knowing how any changes you might envision up will play out. Good strategy formulation is a creative task; if it were a simple, rote exercise competitors would be able to anticipate it and its usefulness would be very limited. That leads to strategic analysis, figuring out the consequences, good and bad of the strategy you are considering. Here there are a number of tools to help make that assessment simpler. The fewer unanticipated consequences there are, the more your successful you strategy is likely to be. One often overlooked aspect of strategy is designing (and carrying out) its implementation. In far too many cases, seemingly good strategies have foundered because insufficient thought (if any) had been given to what had to be done to make it work on the ground.

These four tasks are often presented as a sequence – comprehend, formulate, analyze, implement – but generally this is more of an iterative process. For example, any strategy that emerges from an exercise in formulation needs to be re-evaluated using the analysis tools used to diagnose and understand the situation or problem in the first place. And sometimes new facts come to light during implementation that change the analysis and require a change to the actions that are developed under 'formulation'.

No strategy lasts forever; the world changes and so strategies must adapt. So a strategic management process must continually cycle backwards and forwards through the three elements of implementation, analysis and formulation, often expanding domain knowledge in the process.

The organization of this text

Chapters 6 through 14 deals with frequently used tools and topics in strategic analysis. Chapter 6 deals with Michael Porter's Five Forces model, which deals with the industry in which firms operate and provides a way of looking and the firm's competitive environment. Chapter 7 deals with the Resource Based View, a way of thinking about the firm's internal resources and capabilities. And Chapter 8 discusses (the) two generic strategies, cost leadership and segmentation (sometimes confusingly called differentiation). Chapter 9 deals with vertical integration and introduces a branch of economics (Transaction Cost Economics) that can be helpful thinking about vertical integration. Chapter 10 discusses diversification and Chapter 11 addresses the way diversification is mostly commonly achieved, through acquisition. Chapter 12 looks principal agent theory and Chapter 13 deals with at network externalities (or network effects). Chapters 14 and 15 introduce topics that should be of interest to managers; the analysis of social networks, and artificial intelligence. To conclude, Chapter 15 sets out the mainstream economic paradigm on decision making in order to provide the context for Chapter 16 which takes more social/psychological perspective on the subject, one that has led to the growth of the relatively new field of behavioral economics[5].

---

[5] Much of this work was done by two academics whose work you may find intriguing and thought provoking; Herbert Simon and James G March.

Chapter 5 – Theoretical frameworks

In these notes, the terms "analytical" and "theoretical" are used interchangeably, as are "models" and "framework". Analytical frameworks are useful for several reasons. First they are based on extensive research that often spans multiple firms and industries. While our personal experience may derive from more than one firm, there are some advantages in relying on research based on large numbers of firms. One's own experience may be in rich and detailed, but the causal relationships we infer may rely on particular idiosyncrasies of those forms and therefore not apply more generally to other situations. Large sample research, while lacking some of the richness of personal experience, tends to highlight trends that are common across many firms and are therefore safer to generalize from.

Second, models provide some short-cuts to answers. Often, strategic choices can seem overwhelmingly complex and indeed there are often many moving parts; trying to think how they all fit together and what the consequences might be from particular courses of action can be difficult to say the least. Theoretical frameworks provide a way of prioritizing what matters, affording a simple high level indication of likely outcomes. Of course there will be details to be hammered out and problems to overcome but the general strategic direction suggested by the frameworks provides a road map that helps guide decisions and narrows the choice set to a more manageable size. And while they should not be seen as the "golden bullet", they do help in simplifying what would seem otherwise daunting problems.

Finally, models provide a coherent constraining structure, a foundation within which we can develop more elaborate mental models of causality. But by starting with these basic ideas, we can be relatively confident that the more elaborate mental models we build from experience don't violate important fundamental principles and are less likely to contain spurious relationships. For example, consider the illustration below (Figure 3). If at time T_0, a firm sets a price P_0, for its product and sells quantity Q_0. Then at some future point, T_1, it raises the price to P_1 and finds it has sold Q_1. Plotting these points (left hand figure) appears to show an upward sloping demand curve; in other words the higher the price the greater the quantity sold. But knowing that demand curves are generally downward-sloping should make one suspicious of such a conclusion, leading you to the more likely explanation, that overall demand had increased (right hand figure) shifting the demand curve at T_0 (blue line) to the right by T_1, (red line).

Figure 3
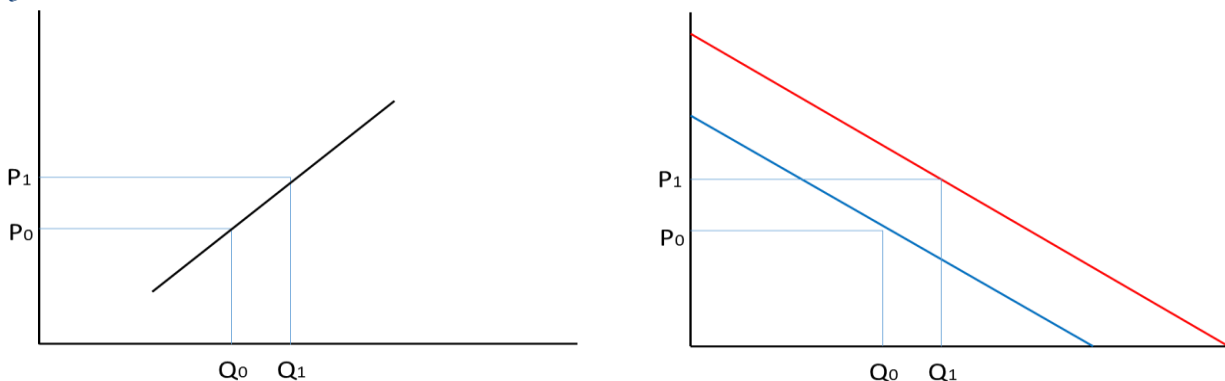
Chapter 6 – Porter's Five Forces

The central question in any strategic analysis is this: "is a firm making a profit/loss and, more importantly, why"? There are three places one can start looking; one can look outside the firm at its environment (competitors, markets, and its context more generally). Next there is an internal perspective looking at the firm's resources, structure and processes. And finally one can consider that nature of the exchanges between the firms and its buyers and supplier, the exchanges across the firm boundary. Michael Porter's 5-forces model (1980) deals with some (but not all) of the external factors. The Resource Based View addresses the internal analysis, and Transaction Costs Economics provides a way of looking at particular transaction to help decide what things the firm should do and what might be outsourced.

To begin a strategic analysis, we often start by looking at the industry and ask the question "does the *structure* of the industry explain the performance of the firms in it"? By structure we are talking about the size and concentration of the industry, the size and concentration of the buyers and suppliers' industries, and the possibility that buyers needs will be met by firms not (yet) in the focal industry. Michael Porter's model helps understanding the pressures that competitors, buyers, suppliers and potential new entrants (and substitutes) might exert on the firms in the industry we are analyzing. Porter's model has five "forces" all of which if strong, tend to depress industry profitability. They are: industry rivalry; bargaining power of buyers; bargaining power of suppliers; threat of new entry; and threat of substitution.

Industry rivalry

Rivalry captures the intensity or ruthlessness of competition. The more intense the rivalry, the lower the industry profitability is likely to be. There are several factors that affect rivalry: industry concentration, the nature of competition and the degree of market segmentation, the cost structure of firms and exit barriers.

Concentration

Concentrated industries are ones in which a few firms account for the majority of the total industry output. For example, in the tire industry four firms (Goodyear, Michelin, Bridgestone and Cooper) account for about 67% of the total worldwide industry output. Industries with many small firms are said to be fragmented. As the French economist, Cournot[6], showed in the 19th century, the more *fragmented* an industry is the less profitable the industry is likely to be. Conversely, the more concentrated the industry, the more likely it is that firm will be able to make a profit. Firms in concentrated industries are said to be more *interdependent* than firms in fragmented industries. While intuitively it might seem that the more interdependent firms are, the higher their rivalry will be and the less money they will make, this is a case where intuition isn't right. In concentrated industries, firms can more easily see who their less 'cooperative' rivals are, those who are willing to drop prices to steel market share and spoil things for everyone in the industry. The more firms there are, the more difficult it is to identify the spoiler; the transparency that comes with high concentration aids price signaling and tacit collusion, both of which help maintain "price discipline", where firms keep prices above their long run costs.

---

[6] See appendix for more on Cournot's model.

A widely-used measure of industry concentration is the *four-firm concentration ratio* (often referred to as CR4). It is the proportion of the industry's total output produced by the industry's four *largest* firms. The four firm concentration ratio of the tire industry is therefore 67%. The benefit of this measure is that it provides a more useful indicator of concentration than just counting the number of firms. For example an industry with a small number of very large firms and a large number of very small ones might appear fragmented if one were simply to count the number of firms. But the four firm concentration ratio would show that most of the industry's sales were actually in the hands of a small number of firms.

Basis of competition

The next question Porter's framework asks is: "do firms compete on the basis of price or on non-price factors?" If they compete on non-price factors, they will generally make more money than if they compete on price. Non-price based competition essentially means competing by product differentiation – there are two aspects to a successful differentiation strategy.

The two questions here are:

- Do the firms in the industry each sell to different market segments?
- Do they compete through innovation and or brand development, as opposed to price?

Market Segmentation

Segmentation means partitioning customers into groups with similar preferences and producing a product for each group that closely meets those preferences. Take the case of [Brithinee Electric](#) Ltd which rewinds industrial electric motors. There are two principal segments to the rewinding business; large motors (over 100 hp) and small motors. When Brithinee Electric gets an enquiry about rewinding a small motor (<100 HP), it refers it to its competitor, C&M. Conversely when C&M gets an enquiry about a large motor it refers the caller to Brithinee. C&M and Brithinee have specialized their operations and targeted their efforts towards different segments of the motor rewinding market; although there are two motor rewinding companies in the Greater Los Angeles area, they do not compete directly with one another. In effect, each has a quazi-monolopy, each being the only firm providing services to that market segment.

Innovation

When firms compete through innovation, they invest in research and development to bring new and improved products to market. Those new products, if well designed, offer buyers valuable attributes and functionality for which buyers are prepared to pay more. That increase in value allows firms to raise prices at least until a competitor brings out something equivalent. These periods of temporary price rises for new product innovations are referred to as *transient rents*. Sometimes, companies will take turns to be the innovation leader as was the case with Intel and AMD in microprocessors during the 2000s. Sometime one firm brings leading edged products to market consistently ahead of its rivals, as Intel managed to do the 2010s[7]. This is essentially differentiation.

---

[7] Although it seems as if AMD has regained its footing, and in 2019 had industry-leading products.

Brand

Developing brand loyalty also mitigates price competition. A strong brand creates a perception of greater value, and as value rises so can price. If brand loyalty increases the perceived differences between one brand and another firms with strong brands can raise prices without losing as many customers to competitors as it would without strong brand loyalty. This too is a means of creating differentiation. In general differentiation reduces the *inter-firm price elasticity*, or the degree to which customers are sensitive to prices differences between competing firms' product offerings.

Price competition

Next, consider whether competition is fought on the basis of price alone. If firms can avoid price based competition they should be able to keep prices above long run average costs. Even when differentiation is impossible and there is little or no scope to impose switching costs on customers, Cournot[8] suggests that by anticipating the actions of competitors and their anticipation of our anticipation of their actions, a tacit understanding can be arrived at in which output is restricted and oligopolistic quantities and prices ensue. However, the Cournot solution is not always stable[9]. Given sufficiently strong incentives to defect from this mutually advantageous cooperative outcome, individual firms are likely to produce more than the optimal quantity that would constitute the best outcome for the industry as a whole, in an attempt to secure a larger slice of the pie for themselves. This is in essence a prisoner's dilemma game in which the best solution would be corporate-cooperate but the outcome typically arrived at (the equilibrium solution) is defect-defect. What turns a potentially cozy situation of tacit collusion into a bloodbath of excess capacity and price war is an industry that has the characteristics of "winner takes all". Two examples of winner takes all situations might help illustrate this.

The first is an industry where minimum efficient scale is reached at higher quantities than the oligopolistic profit maximizing quantity. In this situation each firm can derive small but important cost advantages compared to its competitors by slightly increasing its output. The intuition here is that for a single firm the reduction in the price caused by a small increase in an output is more than compensated for by the reduction in costs associated with that increase in output that derive from the economies of scale in the production process.

A second example in which firms are subject to strong incentives to defect from cooperating to keep prices above long run average costs is where there are network externalities (See Chapter 14 for more on network externalities). In fact this works in a rather similar fashion to economies of scale except that it acts on value rather than on costs. Increases in market share have the effect of increasing the value to customers of the product or service you are offering. To the extent that your firm is the only one to benefit from the increase in the installed base of your product (in other words, the more of your product is sold in the market, the more valuable your product becomes for customers while at the same time this increase in your product's use has no impact on the usefulness of your competitor's product), your firm will try to put itself in the position where it has the largest installed base of any firm in the

---

[8] See the appendix for an explanation of Cournot's model

[9] Stability means tacit price collusion will persist even when there are slight disturbances that might knock firms off balance such as a minor down-turn in demand. When the Cournot-predicted outcome is not stable, small perturbations may lead to a cascade of price-cutting as one firm tries to grab market share from its competitors and red ink all over the profit and loss statement.

industry. The larger it's installed base, the greater the value of its products to customers, and the higher the price it will ultimately be able to charge. However, the first goal is to achieve this large installed base. One way to get there is to reduce prices initially in the order to capture the largest proportion of the market when a disproportionately large market share has been achieved then prices can be raised.

Exit barriers

The last factor that is likely to make competition particularly rivalrous is exit barriers. Exit barriers are any obstacles a firm faces when deciding to cease operations. When exit barriers are minimal, firms seeing prices drop below the long run average costs can close up shop and leave the industry; but when getting out of the industry is difficult, firms may continue to operate, even operating at a loss, in the hope that conditions will improve at some point. Barriers to exit can be financial, psychological or societal.

Financial exit barriers

Financial exit barriers can be additional costs the firm will incur should it decide to cease operations, or costs it currently has which cannot be avoided even when operations cease. Additional costs associated with exit include an obligation to pay severance, or legal requirements, for example in making a manufacturing site safe. Decommissioning a nuclear power plant is a particularly striking example. Decommissioning costs are high because the spent many of the reactor components will be radio-active and will require special handling, spent nuclear fuel must be dealt with and its radioactive half-life may be hundreds of years; all these factors make closing a nuclear power plant very expensive. Similarly, coal mines have to ensure that the mine is safe once it is closed down, and that the tailings from the mine do not pollute the ground water. Similar issues surrounding waste material exist when closing coal-fired power stations.

Ongoing costs that firms cannot avoid even when they stop producing might an obligation to fund pension plan liabilities. Those firms that offer their own pension plans to their employees generally set up separate entities to run the plan, that is to invest money provided by the firm and by employees so that pensions can be paid later when employees retire. Unfortunately, many firms have adequately funded pension plans so there is not enough capital to meet the plan's obligations. That means the firm will have to 'top up' the capital in its pension plan before exiting the industry. The cost of making the plan whole would be an exit barrier. Repaying bond holders is another. A firm's ability to repay its bond holders depends in large part on the how easily its fixed assets can be sold. If they can be sold close their book value if not their original cost a firm can leave its industry relatively easily. On the other hand, if its assets are specific to the industry, it is unlikely it will be able to sell them because the only firms who would have a use for them are its competitors. Indeed, if they will know that since firms trying to exit have no buyers outside the industry for their assets, they will offer cents on the dollar. In other words the likelihood of recouping any of the investment in plant and equipment a firms has made to participate in this industry with specialized assets it low.

If exiting the industry is not an option, the only decision firms are left with is what price to sell. Since their fixed costs are fixed, (lease payments, property taxes, interest to bond holders and so on) whether they produce a single widget or no widgets, all they need to do is sell what they make for more than the direct costs of production (cost or materials and labor), in other words a price that generates at least some *operating profit*. Every widget they sell for more than it costs to make is better than nothing, even

if the operating profit is smaller than the firm's fixed costs (SG&A, interest expense). If other firms in the industry face the same predicament, as they would be in an industry down-turn, likely outcome will be a price war in which firms price a little above or even equal to marginal cost but below long run average costs; and all firms in the industry make a loss.

This is a situation in which many airlines find themselves not infrequently. Commercial airplanes have few uses outside of passenger air transportation. These specialized assets cannot therefore be easily disposed of and constitute a significant exit barrier for an airline. Given the high fixed costs associated with this industry compared to the marginal costs associated with operating the fleet (the marginal cost of a single passenger seat on the plane is practically zero) firms have a strong incentive to defect from "price discipline",i.e. keeping prices high enough to cover long run average costs. Moreover, one airline seat is largely indistinguishable from one offered by a competitor so differentiation is difficult. Given high exit barriers, a commodity product and the high fixed costs and very low marginal costs, it is little wonder that making money in the airline industry is difficult when demand ceases to grow or declines.

So to review, if there are few players in the industry, industry rivalry will likely be low. Selling in non-overlapping market segments will reduce rivalry, as does competing on non-price factors. Non-price competition means one of two "sub-strategies". First, firms can exploit transient rents, for example by continually bringing innovative new products to market ahead of the competition until they catch up or over take. Second, you can try to prevent customers from switching to a competitor's product by designing your product in a way that it is differentiated from those of your competitor, is perceived by buyers as having greater value, or by creating switching costs.

The concentration of the industry, however, is not in and of itself sufficient to guarantee above normal returns[10]. When we noted Cournot's prediction relating industry concentration to profitability, we implicitly assumed that there was a demand curve representing a multitude of buyers with different degrees of willingness to pay. In industries with very many small buyers, this is a fair representation; Cournot's prediction depends on firms being able to pick to whom they sell. Although we have not explicitly said so to this point, buyers were assumed to be 'price-takers'; individual buyers on their own have no significant influence on the price they pay. The only thing they can do is to decide whether the prices charged by firms in the industry are sufficiently attractive for them to make the purchase or not. But when an industry is selling to a market where there are only a few buyers, they have to negotiate over price, which is where bargaining power comes in.

Bargaining Power

Who are the buyers and the suppliers? This is a topic that often causes confusion. First there is the question of who we are referring to when we talk about buyers and suppliers. Consider the value chain illustrated below (Figure 4). The blue arrows indicate the flow of goods and services.

---

[10] "Above normal returns" means returns in excess of the cost production, the firm's fixed costs and the cost of capital required in the production process – in other words the profit you earn on your investment in, for example, plant and equipment, exceeds the cost of the money you borrowed to buy that plant and equipment and its depreciation.

If we were analyzing the components industry (e.g., microchips) as the focus of our analysis (Figure 5), this industry's buyers would be the different product-producing industries that used microchips (and, for completeness, its suppliers would be the raw materials industries).

Figure 5
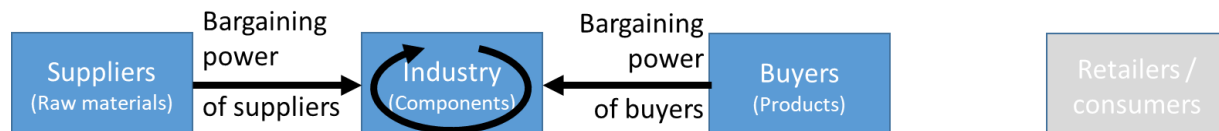
Now if we consider the next industry downstream in the value chain (the product manufacturers, Figure 6), the firms in the industry we were previously calling the "focal" industry become the suppliers, while what were the buyers become our new focal industry.

Figure 6
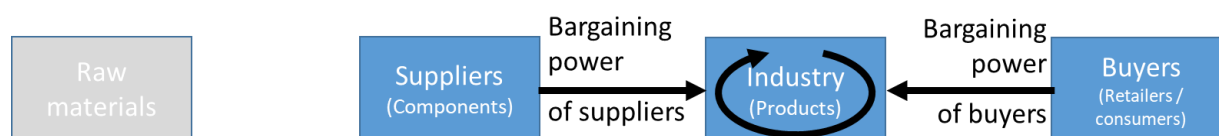
So the component industry is the supplier when our focus is the product-producing industry, and the product industry is the buyers when we are analyzing the component industry; it can all get quite confusing. So for clarity, by convention when we talk about buyers we will always be referring to the industry that buys from the stage in the value chain we are considering, and suppliers will always be the firms in the industry that sells to the industry we are analyzing.

Bargaining Power of Buyers

When the buyer market is also concentrated buyers are no longer price takers. Here, individual buyers can exert influence over firms in the industry. When a single buyer's decision to purchase or not has an appreciable effect on the focal firm, the focal firm can no longer adopt a take-it-or-leave-it approach towards its customers. In a competitive market, prices are set only with respect to the aggregate demand. However, with a concentrated buyer's market, it is no longer possible for the focal firm to treat customers in aggregate but must deal with individual buyers and come to a negotiated agreement about the price it charges for its products and services.

Does industry concentration change bargaining power?

The second area of common confusion is attributing the source of bargaining power. First, one has to acknowledge that bargaining is generally dyadic at the industry level, i.e. between a set of players in the industry and the set of players in the adjacent industry in the value chain. There is therefore a degree of symmetry in the relationship; what give ones side a strong position could equally be viewed as

making the other side weaker. So the attribution of sources of bargaining power is to some degree an arbitrary assignment of those causes to one or other side of the relationship.

The simple heuristic here is to exclude for the determination of buyers (or supplier) bargaining power any the factors that have already been accounted for when determining the degree of rivalry. Since we have already taken into account the concentration of firms in the industry itself when looking at rivalry, we could think of bargain power as being what's *not* accounted for by the structure of our industry, in other words the structure of the down-stream buyer industry.

Figure 7
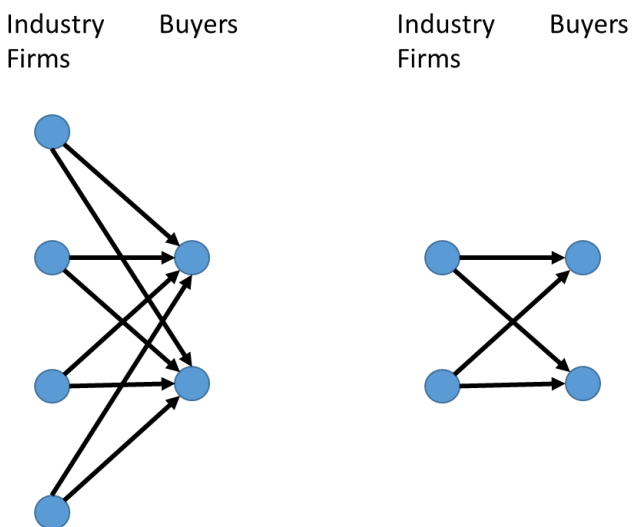
To illustrate this consider these two stylized examples (Figure 7). On the left, four firms sell to two buyers; on the right two firms sell to two buyers. The industry on the left is less concentrated than the one on the right.

However, consider what proportion of the firms production goes to each buyer, an indication of the bargaining power of the buyers. In the left hand example, each buyer takes 50% of the output from each firm. But that's the same in the industry on the right; in other words regardless of industry concentration, the proportion of each firms output is only a function of the number of buyers.

One way to think about the bargaining power of buyers in general is to imagine you are a sales rep for a firm in the industry. A concrete example might be commercial aircraft engines; since airlines choose the engines when they are ordering the body of the aircraft (Boeing and Airbus offer their planes with whichever engines their customers choose), if you are negotiating the sale of engines you have only a few buyers, each contract with one airline will be a large proportion of your annual sales. And if you are selling military engines your choice of buyers is smaller still. It also makes little difference to you if you have one competitors, three competitors or more; if your buyer choses to go with a different firm you lose the same amount of business, regardless of the number of firms in the industry.

Imagining you are in a meeting with the buyer's purchasing manager also helps in thinking the threat of backward vertical integration. If your buyer tells you they are considering making the product you are selling for themselves you are under pressure to do something, such as reducing the price, or providing additional services to keep their business; all of which reduces your firm's profit. Similarly, if you know that your product is slightly different from that of your competitors, and that your buyers would need to make changes to their product design were they to switch from your product to a competitor's, that switching cost takes some of the pressure off. IN other words, switching costs reduce buyer bargaining power. And finally, if your product costs only a few dollars and your buyer's product uses thousands of dollars in materials, the impact of what may be a substantial price change for your firm will be only very minor from the buyers' perspective. Buyers may therefore not feel it worthwhile

spending hours negotiating a small change in price and that reduces the downward pressure on price for firms in the industry. In summary, as buyers become more concentrated, their bargaining power increases and so they are able to negotiate lower prices.

Other things than structure influence bargaining power. For example, if the product we are selling represents only a small portion of our customers' cost structures, then they may be less worried about spending time and effort to negotiate a better deal, so we can raise prices more easily. Conversely, when our product is a large proportion of our customers costs, their profitability may well improve dramatically if they can pressure us to lower our prices. Similarly, if buyers' margins are thin, they will negotiate more aggressively to push down the price firms in the industry charge.

Switching costs reduce buyer power. If it costs money to re-tool to use another supplier's product, we may be able to raise prices. If customers incur costs when they switch from one supplier to another, a firm can raise prices to the point where the difference between their price and their cheaper competitor's is just lower than the customer's cost of moving to that competitor. Cable companies often offer discounts to attract new customers but don't offer them to existing ones. They know that for many people the trouble involved in switching to another cable company (if there is one) is high enough that the saving from a similar offer from a competitor will not outweigh the costs of switching. Credit card companies do this by awarding redeemable points, airlines do it with air miles, and stores of all sorts with their store cards. Any device that increases the cost of changing to another firm allows an industry to maintain price discipline.

Another factor that adds to buyer bargaining power is the possibility of customers backward integrating into our industry. If a buyer can credibly threaten to begin making the product it currently buys from the focal industry, firms in the industry will be under great pressure to accommodate buyers demands for better service or lower prices; both will depress industry profitability.

The Bargaining Power of Suppliers

Clearly, a symmetrical argument can be made on the input side. Previously we considered the effect of increasing concentration on the output (buyer's) side; the more concentrated and thus the more powerful buyers become the more they push down the prices firms in the industry can charge. On the input side, the argument runs in an exactly analogous manner. Here the price a firm in the focal industry would like to pay for our input is the competitive market price, where price is equal to our suppliers' long run average costs. But as our options get fewer and as our suppliers coordinate their prices more effectively, the price we have to pay for our inputs rises. Just as with buyer bargaining power, the threat of forward integration increases supplier power. So do switching costs; if firms in the industry can't easily switch suppliers, supplier power increases. If suppliers have other industries to which they can sell, so firms in the focal industry are not very large customers, they will be less willing to negotiate better prices.

Threat of new entry

If new firms can easily enter the industry, and they do, the industry concentration changes, effectively altering industry rivalry. But even when they do not actually enter, the threat of entry effects prices and industry profits. If an industry is relatively concentrated it will appear attractive to potential new entrants. In order to deter them from entering the industry, incumbents may reduce prices to a level

that just makes it unattractive for new firms to enter. This is often termed *limit pricing*. Another way in which firms may try to make it harder for new entry is to spend on advertising to create brand loyalty, on R&D to create new products or offer customers additional services. All these actions create barriers to entry but come at a cost and reduce industry profits.

There may exists external barriers to entry. Sometimes industries are protected from new entry by law. Inter-city bus routes were regulated until 1973 such that only one company could operate busses between any two cities, effectively a government guaranteed monopoly. The same legal protection existed in trans-Atlantic air travel until 1978. There are other less direct barriers to entry; pharmaceutical companies must complete extensive clinical trials of new drugs before the FDA will allow them to be marketed. That takes time and money, and expertise, making it hard for new companies to enter.

Threat of substitution

The threat of substitution has a somewhat similar effect on our ability to maintain prices in excess of our long run average costs. You could think of substitutes as increasing the number of competitors in the industry (even though they're not in the industry) because they give buyers more options from which to choose. If there are five firms in our industry and 25 firms in an industry that produces products that are in effect perfect substitutes for ours, the outcome is likely to be little different in terms of the prices we can charge that if our industry have had 30 firms rather than only five. Examples of substitutes would be plastic taking the place of glass in bottling, electronic sensors replacing film in photography, high-fructose corn syrup replacing sugar as a sweetener in the soft drinks industry, and ride-sharing apps as a substitute for taxis. What makes identifying substitutes so difficult is that because they come from outside the focal industry, it's hard to know where to look; and because they generally depend on quite different production technologies, their decision making criteria are likely to be different and thus their behavior hard to predict.

Competing against new foreign competitors is not substitution. If you make cars and your new competitors also make care they are in your industry; cars, even if they are made outside the US are not substitutes for cars; they are still cars and the firms that make them are still in the same industry. A new transit system (the Altamont Commuter Express for example), which provides commuters with an alternative to driving, is a substitute.

Chapter 7 - The Resource Based View of the Firm

Making Above Normal Returns without Barriers to Entry

In the previous chapter, we looked at the effect of industry concentration on a firm's potential profitability. The Porter Five Forces framework neatly captures both the focal industry's concentration (rivalry) and the effect of concentration both upstream (the bargaining power of suppliers), and downstream (the bargaining power of buyers).

An analysis of the industry does not tell us everything about *relative* firm performance within an industry. For example, even within a concentrated industry some firms may consistently earn higher returns than others. This is not something that Cournot's model addresses since it assumes that all firms are identical and therefore equally capable. While this is a useful simplification in that it generates valuable insights about particular industries, common sense tells us that this is seldom the case. General Motors and Ford may both be large multinational automotive manufacturers but each has its own unique heritage, culture, values, management practices, resources and bodies of accumulated knowledge. While they may look alike and behave alike in many respects, there will nevertheless be ways in which their actions will differ significantly and this is likely to lead to differences in performance between the two firms.

Richard Rumelt of UCLA presented compelling evidence (1991) showing that the industry in which a firm operated accounted for only about 8% of the variance in profitability among business while stable business unit effects accounted for 46%. In other words the effect of business unit differences matters almost six times more than the industry.

Rumelt's original insight (1984) was that even in industries with no barriers to entry, firms do make returns above the industry's long run average costs, an observation that ran counter to the predictions of Porter's Five Forces model. He explained this by making two key assumptions; one was that firms may have to live with their initial endowments because imitating better performing competitors is often hard if not impossible. This is a central idea in the Resource Based View (RBV). The second was before firms enter the industry they face uncertainty as to which technology to choose and how to organize their operations. These, combined with a longitudinal perspective led Rumelt to some interesting conclusions.

Imagine an industry at its inception: that industry comes into existence at the moment the first firm begins producing a product no one has seen before, one that defines this new industry. At this point, the firm is a monopoly. It has made some choices regarding its production technology which Rumelt characterized as a random draw from an 'urn of technological choices'. Some technologies are better than others, though firms have no way of knowing before they start producing how good the technology they have chosen to develop will be. Entering the industry costs money, needed to build and buy plant and equipment. Only when the firm begins operating does it see how good (or bad) was its technology choice. Before entering the industry, firms consider the range of technologies with which they could enter, and since they do not know until after they begin producing with that technology what the cost or production will be they assign probabilities and a cost to each and calculate an expected cost.

Suppose the firm has a choice of two production technologies, one high cost, C_{hi}, the other low cost, C_{lo}. (Figure 8). Before putting the technology into use it does know whether the one it has chosen is C_{hi} or C_{lo}. Only once it has entered the industry does it discover which one it has.

Figure 8

Enter — EV[C] — p_{hi} → C_{hi} ; p_{lo} → C_{lo}

Do not Enter
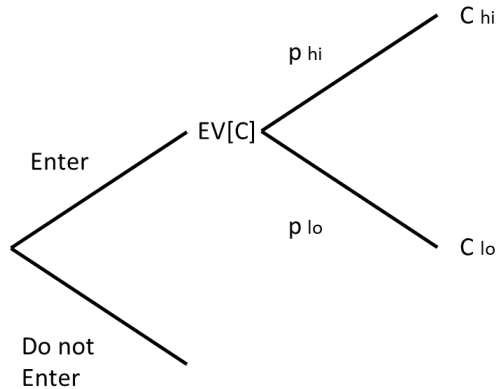
The probability of the chosen technology being high cost is p_{hi}, and the probability of it being low cost, p_{lo}.

The expected value for the firms costs before the choice is made is $EV[C] = C_{hi} * p_{hi} + C_{lo} * p_{lo}$. For simplicity, assume that $p_{hi} = p_{lo} = 0.5$ and $C_{hi} > C_{lo}$.

EV[C] is therefore $(C_{hi} + C_{lo}) / 2$, $C_{hi} > EV[C]$ and $C_{lo} < EV[C]$.

Firms will decide to enter if the price, P is greater than the expected cost EV[C].

As each successive firm enters, the price all firms can charge falls (that's Cournot's result again). Eventually a point will be reached at which the price no longer exceeds the expected cost and at that point rational firms will decide against entering; new entry stops. Any firm already in the industry that has a technology for which unit costs are greater than price, i.e. if the firm drew C_{hi}, it will exit the industry. That will leave only firms that drew C_{lo}.

Prior to the first firm deciding not to enter, P > EV[C]. After it enters there are two possibilities:

$EV[C] > P > C_{lo}$ or $EV[C] > C_{lo} > P$

In the first case, all firms that have entered remain in the industry, $P > C_{lo}$, the industry has a limited number of firms which all make above normal returns. In the second, P falls below when the last firm enters, and it (or another firm) will exit. Firms considering entering will see that a firm that drew $P > C_{lo}$, was unable to remain in the industry as the price P after its entry and therefore conclude that either draw (P_{hi} or P_{lo}) would be unprofitable on entry, and thus be deterred from making the attempt. The result would again be $P > C_{lo}$ (although in this case C is also greater than EV[C]), with no further entry and firms remaining making above normal returns.

Recall that we assumed at the outset that this industry has no barriers to entry to keep newcomers out, so this result is surprising. First, despite the absence of barriers to entry, there is not the infinite number of firms in the industry that a simple neo-classical model would predict. Second, among those firms that remain in the industry all make positive returns, again something unexpected for a market that neo-classical economics would suggest should be perfectly competitive.

Accounting for variance in profitability

A key factor that makes one firm different from another is its management, and the systems and processes its management has put in place. If management matters, the firm that owns a particular line of business operation should affect its profitability.

Figure 9
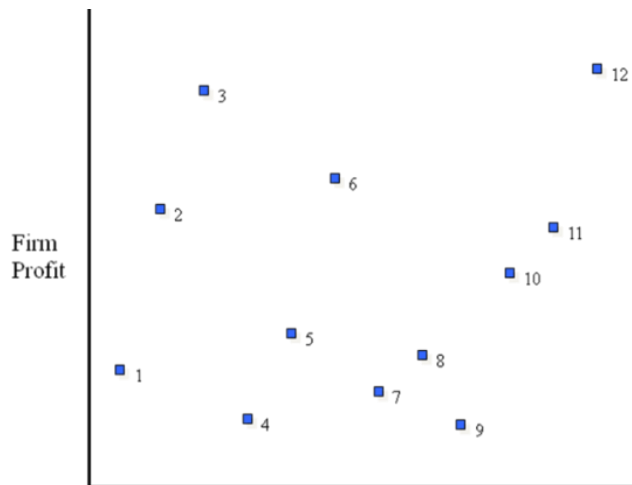
In 1991, Rumelt tested this prediction by looking at firms with different corporate parents (and therefore different management systems) but in the same industry, and he did this for a number of different industries (1991).

Suppose we look at a number of firms whose profitability is quite different (Figure 9).

If now we rearrange them, grouping them into industries, Cournot (and Porter) would predict that those firms in concentrated industries with weak buyers and suppliers and no threat of entry or substitution would make high profits. Firms in less concentrated industries and with some bargaining power from buyers would do less well and firms in fragmented industries would perform worst. Arranging them graphically into these three categories, we might see something like Figure 10.

Industry A is the concentrated industry with low bargaining power and no threat of entry or substitution. Industry B is the one with some bargaining power and Industry C is the fragmented industry. Firms are clustered around the mean from the industry meaning that which industry one is in is a good predictor of profitability.

Figure 10
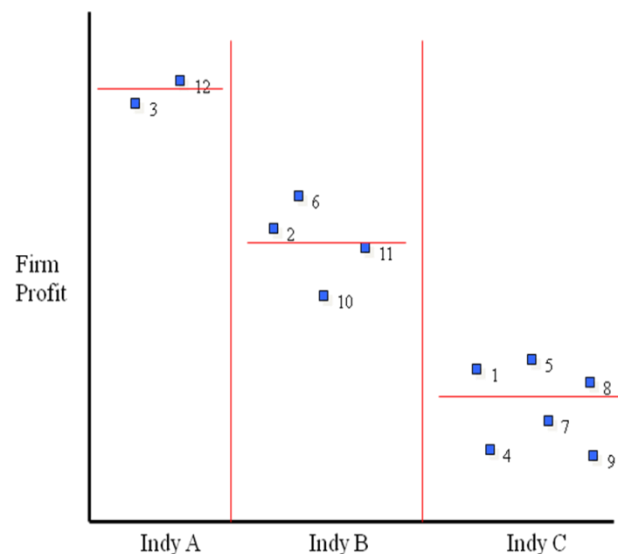
However, in his 1992 paper, Rumelt showed that in fact which corporate parent a firm belonged to mattered twice as much as the industry in which it was situated. Graphically, when firms are grouped by industry the results would look much more like Figure 11.

While the averages are still consistent with Porter's predictions about industry, there is considerable variation in profitability within each industry, and corporate parents accounted for twice the variance that industry accounted for. In other words, management matters more than industry.
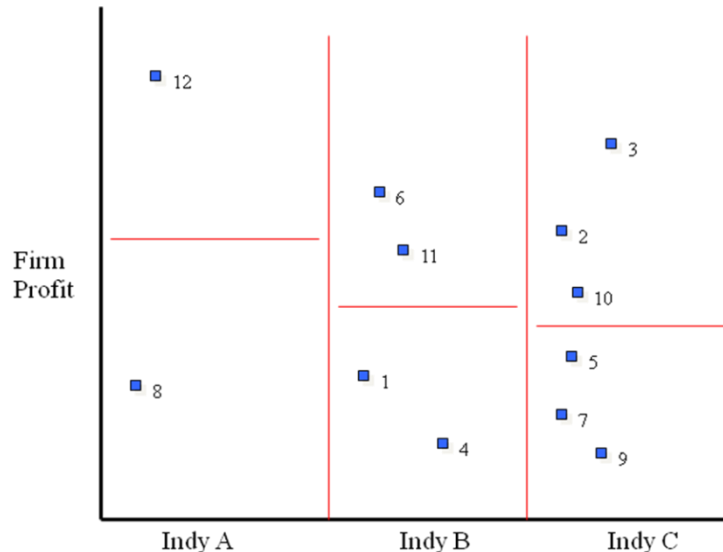
Figure 11

The Resource-Based View of the Firm

In the Resource-Based view the focus is not on products, but on the resources used to create them. A firm can be thought of as a bundle of resources (Figure 12); some are tangible (physical) assets, some are intangible (generally, knowledge based) resources (which the RBV calls *capabilities*).

Figure 12
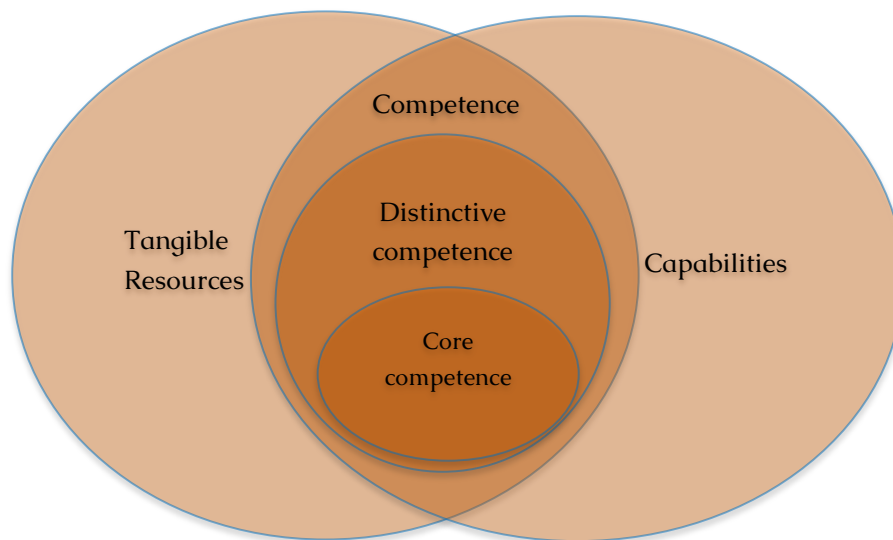
Think of a capability as the ability to execute some or all of the transformation process by which inputs that the firm purchases are transformed into a higher value outputs that are sold for more than the cost of inputs from which they were created. Together, a bundle of physical resources used knowledgeably (in other words, a combination of physical resources and capabilities) constitute a *competence*. A competence is the ability to provide a good or service that is valued by customers or in the VPC framework, the ability to create value.

However, just because a competence enables a firm to create value, there is no guarantee that any of that value will be *appropriated* by the firm. If a large number of firms have the same ability (in other words the same competence) they may compete until the price is reduced to the level of their long run average cost. At best, they will make the level of return predicted by Cournot for an oligopolistic commodity, but the more firms have this competence, the lower their profits. For simplicity we will consider here the case of a competitive industry with no entry barriers which in the neoclassical model would have a very large number of firms. In this case, the buyers of the product or service appropriate all the value created. The seller makes only his long run average costs or normal economic returns.

For competence to confer above normal returns, a firm must both create more value than its competitors and be able to appropriate that value by being the only firm in the industry able to provide that product or service; in other words it must also be distinctive. A *distinctive competence* could be the ability to make design products which have particular aesthetic and functional appeal like the iPhone, or to design and build advanced microprocessors, as in the case of Intel and AMD. However, value creation and rarity on their own isn't sufficient; a distinctive competence must also be hard to copy, or else it won't remain rare for very long. Thus a *distinctive competence*, a competence that confers competitive advantage and allows the firm to generate sustained above normal returns, must be not only create value and be rare; it must also be *inimitable*.

So the test to apply to any resource bundle (tangible assets and capabilities) is first, does it create **value**, second, is it **rare**, and third is it **inimitable**? A resource bundle that creates value is a competence, but a competence that is valuable, rare and inimitable is a distinctive competence. Valuable and rare resources are protected from imitation by other firms by *barriers to imitation.* While Rumelt's 1884 model assumes inimitability, in practices some resources are more imitable than others. Without barriers to imitation, valuable resources will not be rare for long.

Core competences

A distinctive competence is a bundle of capabilities that allow the firm to do something better than its competitors, that is rare, and that cannot be easily imitated. Firms with a distinctive competence will earn higher returns than their competitors in the same industry and do so in the medium to long term. Any distinctive competence that can be applied in a *number of different product markets* is termed a *core competence*. A core competence allows a firm to exploit its distinctive competence in a wide variety of domains. Typically a core competence is not the ability to manufacture a particular product well but something with wide applicability. The ability to assemble a burger inside a minute might be McDonald's distinctive competence – but a more general one might be the ability design a food assembly process for any kind of fast food for example tacos and burritos.

Can distinctive competencies be bought?

In 1986, Jay Barney made another interesting observation. If one could buy such valuable rare and inimitable resources in a market (like the stock market except this one is a market for "strategic factors" i.e., resources that are valuable, rare and inimitable) then the market price for such a strategic factor would be the present value of the expected profits it would yield. If so the net present value of buying such a factor, if it is correctly priced, will be zero. If the NPV of a strategic factor is zero, firms buying them will not make above normal returns as a result of the acquisition. But firms do make above

normal returns, which means they cannot have bought them in this hypothetical market for strategic factors. Implication: if you want to make a profit, you need to develop a distinctive competence in-house; buying one will do you no good. This is something worth thinking about when considering mergers and acquisitions, and in part, answers the question as to why so few acquisitions make money for the acquirer.

There is however an assumption in this argument that a strategic factor will generate the same returns no matter who owns it. That may not be the case; a factory and its employees may be unprofitable when operated by one company, but when acquired by that same resource bundle may be profitable. NUMMI might be an example. The Freemont factory (now operated by Tesla) was once a loss-making GM plant. But when the New United Motor Manufacturers Inc. was created, a joint venture between GM and Toyota to operate the Freemont facility, the same plant with the same employees was profitable. There is another possible problem with Barney's argument, and that is that the market for strategic factors is inefficient, in other words it doesn't always price the factors correctly. Here luck plays a role. If you are lucky enough to buy such a factor for less than the present value of the cash flows it will generate, you can make above normal returns. That is just what happened to Alcoa which bought mineral rights to a deposit of bauxite (the mineral from which aluminum is extracted) that is 30% richer than any other deposit found before or after. At the time no one knew just how rich the deposit was so Alcoa bought it for much less than it was really worth. Alcoa now has a cost advantage over its competitors, none of whom have access to deposits that are anything like as rich.

The Knowledge-based View of the firm

Intangible resources (knowledge or capabilities) are harder to copy than physical ones. For example a competence based on a new machine acquired from a third party is almost inevitably going to be imitated, since any of the firms competitors can acquire the same machine. New product designs can be reverse engineered, and unless elements of the design are protected by patents, the product can be copied. Knowledge that cannot be codified down is particularly difficult to imitate. Some things are hard to explain in words, like riding a bicycle: people who can are usually unable to explain to those who can't what is involved in a way that enables the non-rider to master that art. Such knowledge is termed *tacit knowledge*. Tacit knowledge refers to things we know but that can't be easily explained. It is usually gained though first-hand experience rather than from books, manuals or explanations from others. A capability that is based on tacit knowledge will be particularly hard to imitate.

A capability that relies on the knowledge in many different peoples' heads is also a good barrier to imitation. Such distributed knowledge requires 'picking the brains' of not one person but many. Capabilities that are built on such distributed knowledge systems in which a single element in and of itself will not do anything particularly useful without all other parts of the system being present are very good sources of competitive advantage. Toyota's lean production system is an example. Each part seems simple enough, yet the whole is greater than the sum of the parts. Implement 90% of the things Toyota does and you have nothing – it works only when all elements are present.

In some ways the most tacit knowledge system imaginable is a company's culture. Culture is by definition institutionalized, a taken for granted set of shared understandings, that often support idiosyncratic (to the firm) communication and information interpretation. Because such shared understandings are almost never explicit or documented and are seldom perceived as essential

elements of the system by those who work for the firm, copying a company culture is almost impossible. If culture is the underpinning of competitive advantage, it generally meets the criteria of a distinctive competence, being valuable, rare and inimitable.

It is also worth noting that core competencies can become *core rigidities*. Firms that have relied on a distinctive competence as the foundation of their competitive advantage are often slow to change when the environment changes. And firms whose distinctiveness derives from its culture are particularly prone to this problem, in part because cultures, like routines, often encode the lessons of history without there remaining an understanding of the reasons that the practices they enshrine were adopted. They are repeated "because that the way we do thing round here". Adherence to practices whose rational has been lost make change difficult.

Chapter 8 – Generic Strategies

Most markets can be grouped into clusters of buyers with similar tastes, needs, and requirements; the market can be *segmented*. However this doesn't mean that the most profitable approach is to build a product for each market segment. Generic strategies deals with the question of how to approach and segment product markets.

The central idea is that firms can meet customer needs in two ways; either by offering them almost exactly the product they are looking for, which is generally costly since economies of scale decline with a proliferation of product variants; or by offering them something not ideal but making up for that by offering it at a low price, made possible because of economies of scale. The first is often (confusingly) termed *differentiation*[11] and the second *cost leadership*.

Differentiation

Although producing different products for each market segment is often expensive, as long as customers are willing to pay a premium for a product that almost exactly meets their requirements, prices can often be raised in excess of these higher costs. If in addition to creating a product for each segment, competing firms are careful to target different segments, buyers in each segment face a quazi-monopoly supplier which allows firm to further increase prices.

Cost leadership

If buyers willingness to isn't much affected by producing a product that exactly meets their needs (relatively to one that mostly does) differentiation may be unprofitable. Instead, costs can be driven down by producing just one kind of product – think of Ford's Model T: "you can have any color you like as long as it's black". Reducing product variety leads to greater economies of scale and lower costs mean that profits are still possible when prices are lowered; as they have to be to attract customers who may not think that a one-size-fits-all product is ideal but who appreciate that they can get it for a relatively low price.

Stuck in the middle

Michael Porter cautioned against trying to cut costs and meet each market segment's needs simultaneously. This would result, he suggested, in firms being 'stuck in the middle'.

Some of starting assumptions underpinning the theory have changed. For example, the assumption that production cost rises steeply with increasing product variety has been called into question with two advancements; the 'platform' strategy and computer aided design and manufacturing.

Platform Strategy

In the 1990s Ford realized that it could create a variety of products without making them all completely unique. If different models shared a common set of parts, a common 'platform', economies of scale could be derived from those shared parts and variety achieved in the way they were packaged and

---

[11] Differentiation here refers to the selling of products that are different from others in the firm's own catalog, but are not necessarily differentiated from those of a competitor.

accessorized. The Ford Explorer shared the chassis, engine, gearbox and drive train with the Ford F150. The models differed only in their interiors and body designs. Ford achieved economies of scale with these common components, but sold the final product to two different market segments. Product variety was gained without the large increase in costs otherwise comes with increasing variety.

CAD/CAM

Computer Aided Design / Computer Aided Manufacturing (CAD/CAM) has also enabled firms to increase variety without increasing unit costs. Linking numerically controlled (NC) machine tools and robotic assembly lines to computer aided design software enables production lines to be switched easily from one model to another and for variations in models to be created and manufactured without major disruption, re-tooling or cost.

Lean manufacturing

Lean manufacturing was an approach to pioneered by Toyota that dramatically changed the way manufacturing facilities were operated. Until lean manufacturing, it had been considered imperative that the production line never stop – to do so would lead to under-utilization of what was usually the company's most expensive asset. In order to ensure that the line worked 24x7, parts inventory was made or purchased in large batches and buffer stocks created between production stages to prevent a problem in an upstream stage causing stoppages downstream. While the approach did maximize the utilization of the production line it often led to large inventory holdings and tied up working capital. It also meant that before model changes could be implemented inventory holdings had to be built into now obsolete finished goods or scrapped. It also meant that any defects in an upstream batch might go un-detected until a large number of parts with that defect had already been made. Given the cost of scrapping the parts or making do, many firms choose to use the sub-standard parts, which ultimately led to quality and reliability issue in the finished product and a loss of reputation and brand loyalty.

Toyota reasoned that the costs of holding inventory and of preserving hard a won reputation for quality and reliability were high enough to justify a different approach; the "Kanban" system, where parts were only built when a downstream stage in the assembly process needed them, meant that intermediate inventory holding was effectively reduced to zero, almost no inventory had to be scrapped as defects were identified immediately and no products needed to be shipped with sub-standard components.

Quality vs cost

It is often assumed that quality products must cost more to make. However, Toyota's Kanban system allowed it to maintain and improve quality while simultaneously reducing costs from defective parts and waste. Counter intuitively, a work process that raised quality also reduced costs. Another example is Lexus (a Toyota brand) which maintained that it's straight as an arrow exhaust pipe improved performance and fuel consumption; it was also *much* easier to manufacture than a pipe with lots of bends; higher quality and cost reduction going hand in hand.

Chapter 9 – Economies of Scale, Experience and Learning

Economies of scale experience and learning seems a fairly simple idea; the cost of making something gets less the more you produce. But it's worth unpacking this because although scale and experience both lead to lower unit costs, they are not the same and that has implications for strategy.

Chapter 10 – Vertical Integration, Transaction Cost Economics

When a firm operates in several stages in the value chain it is said to be vertically integrated. A firm that produces a product and operates the retail outlets though which it is sold is vertically integrated. So is a firm that makes the components if needs for a products it produces, even if those components could be bought from a supplier.

Transaction Cost Economics provides a useful way to think about vertical integration. While not the only consideration, it covers many situations which were previously considered in a somewhat ad-hoc manner and provide a systematic way of approaching "make or buy" decisions.

Transaction Cost Economics (TCE)

Background

In the 1930's Ronald Coase, an economist, left the LSE where he was an undergraduate, to investigate the structure of different American industries. "What came out of my enquiries was not a complete theory answering the questions with which I started but the introduction of a new concept into economic analysis, transaction costs, and an explanation of why there are firms."

The question of why firms exist seems a simple question, one that almost anyone could answer, yet it turns out that economists didn't actually have a good answer. The problem centers on the firm as an institutional arrangement for organizing work. Coase was suggesting that one could accomplish the same things by outsourcing almost all the activities involved in creating a product and paying those independent subcontractors for their work instead of setting up a firm and hiring those same people as employees. Such a subcontracting-based arrangement was theorized to be less expensive than employing the same people (for example you don't need a large HR department, monitoring costs are eliminated, no need to pay employees when orders decline), in which case the existence of large firms employing lots of people was a puzzle. In a free market the less expensive mode of production, subcontracting, should beat out the more expensive (i.e., large firms). The fact that large firms exist therefore suggests something else was going on, something that made subcontracting more expensive than firms.

Coase' work lay relatively unnoticed for many years till another economist, Oliver Williamson, 'discovered' Coase' paper, "The Nature of the Firm" (1937), and began to develop it in a wide array of applications of the theory (Williamson 1981; Williamson 1996; Williamson 2006). This new strand of economic theory became known as Transaction Cost Economics (TCE). In 1991, Ronald Coase received the Nobel Prize in economics for his 1937 paper. Williamson was awarded the price in 2009 for his contribution to the development TCE.

TCE

In TCE, the focus of attention is not the industry as in Porter's model, or the firm as in the RBV, but the *transaction*. An exchange is an agreement struck between two parties to trade goods or services. Exchange can take place in two different institutional contexts: the market or the firm. An institutional context is the set of rules, regulations, and norms that govern behavior, in this case who exchange is

regulated. For example, an exchange carried out in the market allows for the possibility of one party to the exchange suing the other in court if they believe the terms of the deal have not been kept; but a manager in a firm can't sue an employee for not getting their work done satisfactorily, nor can an employee sue the manager for requiring work that wasn't originally specified out when a project was begun. That's because employment contracts are quite general; as an employee you agree to accept a salary in return for allowing your manager (i.e., the firm) tell you what to do and how to do it. Conversely as an independent contractor doing the same task, you reserve the right to get the job done in whatever way you want as long as it meets the terms of the contract you agreed with the other party to the exchange. Firms and markets therefore differ in ways that alter the cost of monitoring and enforcing the terms of a transactio and it is those costs that are the *transaction costs*.

Coase's starting point is to assume that all economic transactions can be most efficiently carried out in a market. 'Efficient' means that the cost of the transaction is as low as it can be. Holding the cost of making and delivering the good or service to one side, Coase considered the difference in total cost of the exchange depending on whether it was conducted inside a firm or in the market. The total cost comprises, in addition to the cost of making the product, the cost of writing and enforcing the contract and any associated information gathering or search costs; these are transaction costs. These are not always negligible, and vary depending on the context (firm or market) in which the exchange takes place and two other characteristics of the transaction.

If market are so efficient, Coase asked, why are they firms at all? His answer, and Williamson's, is that markets are not always the most efficient way of handling economic exchange. Under certain conditions, economic exchange between actors (firms or people) is less efficient when undertaken in the market than when some other institutional arrangement is used to support the exchange.

The institutional context in which exchange takes place
Any economic exchange takes place in an institutional framework, a system laws, enforcement mechanisms and procedures designed to protect the parties to the exchange from opportunistic behavior by the other. The institutional framework that supports transactions conducted in the market is different from that for the exact same exchange undertaken within the legal entity we call the firm. For example, someone working as a contractor for a firm would be protected by contact law, while the same person working as an employee would be subject to employment law. While the work done might be the same, the means (and the associated costs) by which the terms of the exchange are governed and enforced are different.

Briefly stated, TCE suggests that in settings where there is *uncertainty* about the future, market contracting is necessarily *incomplete*, that is to say not every eventuality can be accounted for in the contract and those unaccounted for outcomes have significant consequences for one or other party to the exchange. When there is *incomplete market contracting*, and where substantial investments must be made in *transaction specific assets*, the possibility for *ex-post opportunistic behavior* will increase the risk associated with the transaction and that raises transaction costs. O*pportunistic behavior* or opportunism is when one or other party to the exchange finds a loophole, something that wasn't anticipated, that they can exploit to their advantage and to the detriment of the other party. Williamson (1977) states that while not all parties will behave opportunistically all of the time, some will do so some of the time, and thus the risk of opportunism must be considered when exchange is entered into. Given

the risk of opportunistic behavior, the cost of marketing contracting under uncertainty with transaction specific assets may rise to the point where it is more expensive compared to the same exchange carried out in another context, namely the firm. In some cases the cost of market contracting becomes so high that it prevents the exchange for taking place in the market at all.

The way to deal with high transaction costs that arise with asset specificity and uncertainty, Coase and Williamson suggest, is for the transaction to take place within a different institutional context – namely within a single firm where the incentives for opportunism disappear. The solution to the hold-up problem is for the mine to 'forward integrate' into transportation. Vertical integration, the topic we are considering this week, is thus seen as a solution suggested by TCE to the problem of incomplete market contracting under uncertainty where one or both parties need to make investments in transaction specific assets.

How does the firm reduce opportunism and thus transaction costs? Williamson suggests a number of ways in which exchange based on within-firm exchange help. First employees do not have strong incentives to appropriate (since they are usually on some kind of salary or wage, which reduces the degree to which there is conflict of interests in the incentives of the transacting parties. Second, when disputes do arise they are settled by managerial 'fiat' or the imposition of hierarchical power. Not having to involve lawyers, either ex-ante to draft complex contracts or ex-post not having to pay large legal fees to settle disputes reduces the costs of contract enforcement. Because of the way employment law is written, courts intervene far less in transactional disputes within the firm than between two parties who contract in the market; that makes a wide range of transactions possible within the firm that would not be economically rational to undertake in the market.

More generally, even when absolute market failure is not predicted, the lower cost of contracting for a variety of transactions means that though they might be undertaken in the market, the firm is a cheaper context in which to carry them out. Hence we have an answer to the question why do firms exist – they exist precisely because they are an institutional arrangement that creates lower transaction costs for these types of exchange.

More transactions are executed when firms exist to facilitate them than would be the case were firms not there for enable them; in this way the existence of firms gives rise to economic activity that would not otherwise take place. In addition, transactions that would be more expensive absent firms are now completed at lower cost. Finally, firms create economic value both through increased economic exchange and exchange completed at lower cost and in a market.

Vertical Integration

TCE thus answers the "make or buy" decision; should firms buy its inputs or backward integrate and make them itself. When an input requires the firm (or its supplier) to make investments that are transaction specific, and if there is significant uncertainty about possible eventualities, the combination of uncertainty and asset specificity indicates "make" (vertical integration) over "buy".

While TCE provides one indicator of when vertical integration is appropriate, there are other considerations. Cutting off supply or raising the price of inputs competitors is another. By backward integrating, firms can reduce the availability of inputs to its competitors. Similarly, by forward

integrating into distribution firms can make it difficult for competitors to reach their customers. This has in the past been viewed as anti-competitive practice. In the UK, brewers once owned a significant proportion of the public houses through which beer was sold, and in doing so stifled competition from independent breweries; the rise of microbreweries was impeded by the established firms being forward integrated into retail distribution. To promote competition the UK Government passed legislation forcing the brewers to allow other firms to sell their beer in "tied" or brewer-owned pubs.

A third reason to vertically integrate draws on the RBV. If a firm believes is that it has or could develop a distinctive capability either upstream or down-stream then it can differentiate its product by vertically integrating. Sony provides an example. Akio Morita, Sony's founder and CEO was not happy with the quality of the cathode ray tubes (CRT) (used in color TV sets before the advent of flat-screen TVs) available from third party manufacturers and believed that Sony could create a better tube; Sony's R&D department invented the "Trinitron" CRT which had better picture quality than the older 'shadow mask' tubes. The Trinitron design, which was 25% brighter than the 'shadow mask' tubes ("valuable"), came to market in the late 1960s, was available only from Sony ("rare") and was protected by a patent ("inimitable"). Sony's backward integration into CRT manufacture provided it with proprietary technology that afforded it a competitive advantage.

Vertical integration does not come without risks however. For example it reduces flexibility. If a firm is vertically integrated, switching to different inputs that might be required when a new product is developed is more difficult than when those inputs are sourced from third parties. Moreover, the costs of those inputs, when produced internally, can be hard to control. Managers in one part of the company can make a convincing case that the costs involved in making that part are reasonable, and managers in the division that is "buying" them, because they are at an informational disadvantage, may not be able to get lower transfer prices and inefficiencies in the upstream production side may develop. It is often suggested that what is missing from efficient transfer pricing is market discipline, that is the ability to points to a market price as a benchmark against which transfer prices can be assessed, but where there is market failure, a lack of market discipline is inevitable.

Chapter 11 – Diversification and Conglomerates

Diversification occurs when a company extends its activities into new product markets. Diversification can include vertical integration, when firms take on the activities of some of the companies from whom they used to buy or to whom they used to sell. Diversification is generally divided broadly into two categories; related and unrelated.

Related and unrelated diversification

As the names suggest, related diversification is where a firm exploits assets, either physical or knowledge-based, it already has in the product market it is entering. Using an unused fertilizer production facility to make explosives would be an example of using existing physical assets in a new domain. Cannon used its knowledge of optics from its camera business and electronics from its calculator division to begin building photocopiers. Honda used its knowledge of small efficient combustion engines to diversify from motorcycles, where it began, into cars, marine outboard motors, portable electrical generators, ATVs and lawnmowers. Related diversification may prove challenging when two lines of business with differing demands are dependent on a single centralized or shared asset. Nevertheless, research suggests that firms that diversify into related businesses are more successful than ones that undertake unrelated diversification.

GE is an example of a highly diversified corporation with activates spanning consumer durables (white goods and brown goods), light bulbs, civil and military aircraft engines, locomotives, turbines, electrical generating and transmission equipment, medical devices and finance. It is also an example of an M-form (or multi-divisional) corporation, where business units with somewhat similar characteristics are grouped together in a division, for example consumer durables and light bulbs in the Home and Business Solutions division, and turbines, electrical generating and transmission equipment in Power and Water, Transportation for locomotives and rolling stock and aircraft engines in the Aviation division.

When a firm begins operating in a product market with which it has no prior experience and there are no linkages between any of its current activities and those into which it is diversifying, diversification is said to be unrelated. Examples of unrelated diversified firms include ITT, Beatrice Foods, Greyhound, United Technologies, The Swire Group, The Tata Group, Kawasaki, Mitsubishi, Hitachi, Sumitomo, Mitsui and Hanson Trust.

Rationale for diversification

In the 1950s and 60s, many firms diversified to maintain revenue growth as their core markets matured. However, coordination and agency costs rose as firms became more diversified, often resulting in declining profitability. Diversification was sometimes undertaken to achieve lower earnings volatility; the lower the variability in earnings, the better was ratio of risk to reward. But this provided only a 'one size fits all' solution for investors who might want a portfolio more specifically tailored to their specific needs. With the growth of the mutual fund industry and discount stock brokers, individual investors could build their own customized portfolio without the coordination and agency costs associated with a single diversified firm.

A final reason companies diversify is that their size and financial stability provides access to cheaper capital than the individual businesses would be able to gain on their own. There are two problems with this argument; first, high potential businesses will get access to capital with little difficulty so the only capital requests that this opens funding for are those that quite possibly shouldn't be funded anyway. Second, and relatedly, research shows that R&D spending by 758 of the largest US corporations yielded negative returns in more than half of those firms. In other words, the study concludes that managers in generally can't be trusted to make large investment decisions without a more direct form of market discipline (Michael Jensen, the study's author, advocates high leverage (debt), since high leverage reduces managerial discretion[12]).

"Sticking to the knitting"

Diversification fell out of favor in the 1990s and many large firms changed strategies to "stick to the knitting", meaning focusing on the things they were good at (in other words thing in which they had a distinctive competence) and divesting the rest. This change was in part caused by the use of the leveraged buyout in the 1980s, illustrated in the Hansen Trust case. A leveraged buyout is when the financing to acquire a company's shares is secured using the target company's assets and cash flows as collateral; those arranging the buyout generally bring only a small proportion of the financing to the deal themselves (hence the term "leveraged"). During the 1980s corporate raiders were able to raise money in the capital markets to buy controlling stakes in many large conglomerates and break them up into smaller individual businesses. What made these kinds of deals attractive was that fact that conglomerates generally traded at a discount relative to the combined value of their individual business units. In other words the whole was worth less than the sum of the parts. This so called "conglomerate discount" arises from three issues: agency and governance; coordination costs and market legitimacy.

Agency and governance

When business units are part of a conglomerate they are not directly accountable to financial markets for their performance. When their products are sold internally to other business units, for example when the conglomerate is vertically integrated, neither are they exposed to product market forces. Incentives are often insufficiently aligned with the superordinate goals of the corporate parent. This lack of accountability and transparency can lead to decisions being made for reasons other than purely financial or strategic; politics and personalities cloud judgment, and performance suffers as a result.

Coordination costs and flexibility

As firms get larger and more complex, the corporate home office has less and less time to devote to oversight of each individual business unit and the home office's "value-add" in terms of expertise and guidance declines. When the businesses are linked, for example along the value chain or through sharing of common resources such as logistics and distribution systems, divesting an underperforming business is difficult. Contrast this to an individual investor or fund manager who can sell an underperforming stock at will. It is this lack of flexibility that finance theory suggests accounts in part for the under-performance of conglomerate relative the sum of the parts, and the discount at which they therefore trade in the stock market.

---

[12] Jensen, M. C. (1993). "The Modern Industrial Revolution: Exit and the Failure of Internal Control Systems." Journal of Finance 158(3): 831-880

Categorical imperative and market legitimacy

The final reason for the conglomerate discount is the lack of transparency to investors and the consequent difficult in valuing the conglomerate as a whole. Instead of benchmarking performance against a single industry, conglomerates have to be benchmarked against a raft of industries and even then the comparisons aren't perfect since there often isn't enough information about each individual line of business to do a thorough enough job. Consequently there are fewer market analysts who provide research on conglomerates which in turn depressed market demand for conglomerate stocks. Hence they tend to trade at a lower PE than sum of the sum of industry average PEs for each of the individual businesses.

Chapter 12 – Acquisitions

An acquisition is when one firm (the acquirer) buys another (the target). This involves buying all the shares in the target company, thereby acquiring the rights to control the company and its assets. Acquisitions can be hostile or friendly. They are said to be friendly when the board of directors of the target company recommend the take-over to its shareholders. To make this recommendation they should be convinced that the acquisition on the proposed terms is the best available for the shareholders whose interest they were elected to represent. A takeover is said to be hostile when the directors of the target oppose to the acquisition. Opposition is not unexpected given that the target's board is answerable to its shareholders and is therefore charged with getting the highest price for the shares as it can; and the initial offer from the acquirer will likely not be the best offer available.

Asymmetric information

Generally, for a variety of reasons, the acquirer doesn't know all that there is to know – or indeed all they arguably ought to know – about the target company. Despite the requirement for due diligence, not only may it simply not be possible to know everything, some of the information will be coming from the management of the target company which has an interest in presenting the most optimistic picture of the firm. Of course, the acquiring company (or companies) know this, but there is nevertheless uncertainty about the true state of affairs, and that uncertainty allows decision making bias to distort the picture.

Secrecy

Part of the reason many merger negotiations are based on insufficient information is the need to keep them as secret as possible. Secrecy is important for two reasons. First, it may help avoid a bidding war which will push up the price of target's shares and reduce the financial rewards to the acquirer. Second, if the employees of the target company (and to some extent the acquiring company) know that they are to be merged, many will often rightly assume there is a risk they will lose their jobs.

Bidding wars

One reason for secrecy is to avoid bidding wars. When one company tries to acquire another, whether the takeover is friendly or hostile, the target company's management will try to ensure it get as high a price for the company as they can. Indeed they are bound by a fiduciary duty to ensure that their actions are in the best financial interest of their company's shareholders. One way to ensure this is to solicit bids from other potential suitors. Even when they are not actively sought, the acquirer's competitors may be alarmed by the prospect of the acquirer gaining an advantage in terms of economies of scale or scope and may enter the fray to try and buy the target themselves. Now you have what amounts to an auction; and auctions are designed to ensure the highest possible amount is paid for the things being sold. First one company will make a bid, another will make a counter offer; this may repeated several times until one of the bidders drops out because the price has risen to a point at which it no longer makes financial sense to buy the target company. What determines the final price of the sales, then, is the reservation price of the acquiring companies. We'll come back to that in a bit.

Uncertainty and 'asset' flight

Another reason not to go public with the acquisition plans is that uncertainty may trigger departures; and they may not be random. Those who are risk-seeing and successful may be more likely to leave while those who are risk averse or less successful may be more likely to stay. In the period of change that follows an acquisition, the risk-takers and highly accomplished may be just the people the firm needs most to make the post-merger integration successful.

Negotiation dynamics of a bidding war

One of the problems with budding wars is the combining of the dynamics of an auction with attribution error and uncertainty. When multiple bidders are bidding for the same item (the target company in this case), the winner will pay just over the second highest bidder's reservation price. Their reservation prices will be the value the bidders place on the target. In other words the acquisition price depends on the two highest bidders' valuations. The true value of the target can never be determined absolutely; it depends on a host of unknowns, but some estimates are possible with certain assumptions. For example an assumption of the ability to lower the target's cost structure will affect the valuation. The greater the value of savings assumed, the higher the valuation.

Given some uncertainty as to the prospects and therefore the actual worth of the target, managers in the acquirer may be overly optimistic in their predictions of future cash flows. They may be aware of risks that would lead to lower expected values but may underplay them and therefore overvalue the company. This is particularly problematic in a bidding situation where one acquirer's hubris may spur its competitor to up their bid too. A competitor's high valuation may suggest to the other bidder that the target's prospects are actually rosier than they had estimated, so leading them to up their bid. One over-confident estimate of future cash flows trumping another. Two other emotions may also play a role: ego (pride) and fear. When acquired had locked horns, not wanting to be seen to loose may drive bidding higher than might be seen as objectively reasonable. And fear of letting ones competitors take the prize ("FOMO") may be another factor.

Advisor incentives

Acquirers are often advised in the deal making process by banks. Banks get paid a percentage of the deal price, so their incentive is to get as high a price as possible and to get any deal rather than no deal at all. This potentially skews the interpretation of incomplete information towards the optimistic and a higher valuation. It's an example of a principal-agent problem.

Post-merger integration

Research suggests that most acquisition fail to meet their goals, goals on which the target's value was predicated, because of problems in the post-acquisition phase. Good people often leave. Those that are left often become embroiled in political battles to save their departments or their jobs in the rationalization that almost inevitably occurs (and on which the bid is often predicated). And since the pre-acquisition planning is often done at a very aggregate level, important details can easily be missed that later emerge as major obstacles in the integration phase.

In summary, it is not unusual for acquisitions to be lucrative for the shareholders of the target company but not so for those holding share in the acquiring company.

Valuation

An important aspect of acquisitions is valuing the target appropriately. We have looked at reasons why managers often over-value the companies they are trying to acquire, but we haven't yet said anything about how one might value a company.

One way to see how much a company might be worth is to look at the stock price. Multiplying the price by the number of share outstanding (which you can find and the company's SEC filings) will give you the value that investors collectively ascribe to the company. Using the stock market as a proxy for value assumes that all investors are completely rational and that there is no other information available that the information available to those investors.

Another way is to start from first principles and use information on the company's finances to arrive at a valuation. "Appendix C – Valuation" provides an approach for very simple situations (constant cash flows). For or complicated cases, McKinsey, a consulting firm, suggested a set of methods in a book titled "Valuation: Measuring and Managing the Value of Companies" originally published in 1990 and now in its 7th edition that you may refer to should the need arise.

Chapter 13 – Principal-Agent Theory

If you have to make a decision how to invest your resources (e.g., time or money), you are likely to give the decision careful thought and make a decision that increases your 'utility', to use the economists term. In many cases increasing your utility means making your better off. But what if it was someone else's money, and you were being employed to make them better off? Would you act purely in their best interests? That is the question at the heart of Principal-Agent Theory. It's a problem that arises fairly frequently. For example, is the life insurance policy sales person selling you the policy that is best for you or the policy that makes them the most commission? It's also an issue with the way companies are run. When operational and strategic decisions regarding a company's operation were made by its owner, it is likely that the decisions will benefit the owner. But when owners hire a manager to make decisions on their behalf, the manager may be making decision that benefit them and not the owner. With few exceptions, that's the situation facing public companies.

Public companies are owned by their shareholders (pension and mutual funds and private investors), but run by managers appointed by the directors who the owners elect. And there is an information asymmetry here; managers know more than the board about the details of the company's finances and the state of its operations, and the can paint an overly optimistic picture when they account for their decisions to the board. The board, while charged with doing due diligence on behalf of the shareholders has less information to draw on and less time to process it. So what guarantees do the owners have that managers are working in the owners' their best interests? Principal-agent theory suggests that aligning manager's incentives with that of the owners is the answer. Since owners are generally interested in returns to their investment in shares of the company, either though stock price appreciation or dividends, giving mangers a stake in the company helps address the problem.

One solution to the principal-agent problem is to grant managers stock or stock options in the company. By granting managers shares in the company as part of their compensation, their decisions should be more closely aligned with the interests of external shareholders. Stock options are another popular ways of aligning manger's interests with those of shareholders. A stock option gives its owner the right to buy shares in the company at a set price, called the "strike price". If the price of the shares rises above the strike price, the option is said to be "in the money" and its owner can exercise the option and buy the shares or they can sell the option to someone else. The advantage of stock options is that they afford a stronger incentive for managers that shares alone. A manager may be happy with a share price that hardly rises at all if they are simply given the stock as compensation. But when they are given an option that is worth nothing until a share price target is met, managers may be more motivated.

However, research suggests that options have several drawbacks. First the strike price can be negotiated by a manager such that the option comes into the money with relatively little effort. Second, even when the strike price is fairly set, and leads, as intended, to more risk seeking decision making, those decisions aren't always good for shareholders. And some of those 'not-so-good decisions' even

involved misrepresenting the company's financial performance to artificially inflate the stock price and so trigger the stock option[13].

---

[13] "Why Stock Options Are a Bad Option" Freek Vermeulen, Harvard Business Revirew, April 21, 2009
https://hbr.org/2009/04/why-stock-options-are-a-bad-op.html

Chapter 14 – Network Effects / Externalities[14]

Sometimes products are worth more than "just what's inside the box". The very first fax machine sold would have been no more than an expensive paper weight since there would be no one else to send or receive faxes (it would probably have made sense to sell them as a pair). If only you and one other person has a fax that could be useful but your fax machine's utility increases with every new machine sold to someone who, at some point, you might want to send a fax to or receive from. The value of your fax machine is therefore not just in the features the manufacturer has put into the machine itself, but increases as more people own fax machines. That is a network effect.

Network effects have become increasingly important over the last two decade, since more and more products and services rely on them. Products that rely on network effects include MS Office, Apple's products (Mac / iOS). Services such as Facebook, Uber and Lyft, dating apps or on-line markets like eBay also rely on network effects. Indeed it is this that in many instances drives companies to bring users to their platforms while losing oodles of money in the hope that once they are established, prices can rise, costs will decline and profits can be made. While Facebook is profitable, Uber and Lytf are still (in 2019) losing money.

There are two kinds of network effect; direct and indirect. Direct network effects refer to cases like the telephone or the fax where the value of the product increases with the adoption of the product itself. Another example is Microsoft Office. While for many years there were several incompatible (or only somewhat compatible) alternatives in Star Office and Open Office, the ease of file exchange within the Microsoft Office network of users essentially walled off the Star Office and Open Office users from Microsoft Office users and vice versa. So the value of the Microsoft Office, with more users than Star Office and Open Office was greater, not because the product was batter but because more people used it. And because it was seen as more useful / valuable, installations grew faster (even though Star Office and Open Office were free!), making the difference in network-derived value greater still[15].

The other kind of network effect is "indirect"; here the value of the product depends not on the installed base of the product itself but the availability and value of complementary products. An example from the 1980s is IBM's OS/2.

OS/2 was a mainframe-strength operating system written for the PC. It was designed to replace DOS which was architecturally limited. It was written by a team of software developers who had years of experience designing operating system for the largest and most reliable systems IBM sold (the MVS operating system). When OS/2 was launched, Microsoft had been building a competing product (originally based on DOS) called Windows. OS/2 was (in my view) without doubt a better built, more robust product. But there were lots more applications that would run under Windows than under OS/2. A big driver of Windows sales was Microsoft Office, its "killer app" which almost everyone who had a

---

[14] The terms "network externality" and "network effect" are often used interchangeably but economist have argued that externality should only be applied when the benefit of the effect is not captured by the product's (or service's) producer (provider).

[15] Open Office and another variant, "Libre Office" are still available today; they have many of the same features as Microsoft Office, and are still free. And with Microsoft's adoption of the OpenXML international standard for office files, documents can be easily exchanged between them.

desktop computer would install and turned the PC from a specialized tool running niche applications such as accounting or inventory management into a general purpose tool useful for practically everyone. Microsoft also set up exclusive deals with PC clone manufacturers to have Windows pre-installed, while IBM provided OS/2 at significant cost with its own PCs. Microsoft's smarter marketing led to Windows becoming the most widely installed desktop operating system. And as Microsoft Window outstripped OS/2 in installations, application developers saw little point in developing (or maintaining) application software for platforms other than Windows and eventually abandoned OS/2. That only made OS/2 less attractive to buyer of desktops and helped sales of Windows.

Interestingly, Apple was in a similar position in the 1990s. It's Mac operating system was not being used as much as Windows and there were fewer and fewer apps available for the Mac relative to Windows. Apple's response was to license the Mac OS to clone manufacturers. While the decision made sense in terms of trying to get the Mac OS into the hands of more users, that decision was reversed when Steve Jobs came back to Apple in 1997. Jobs was concerned that Apple's image would be diluted if it continued to provide its operating system on machines that didn't have Apple's attention to design and attention to detail. While that may have seemed an odd decision, ultimately Job's later innovations proved that it was probably the right one. With the launch of the iPod and Apple music, Jobs was creating a network around Apple products that went far beyond users of computers. When he took the iPod Touch and launched a mobile phone, the iPhone, based on the same UI design, he further embedded iOS (the operating system Apple's mobile devices use) into the consumer market. When people talk about the Apple ecosystem, they are really referring to the multiple interlocking networks that create value for Apple users. Ultimately network effects depend on market share.

In the smart phone market, the battle for market share is now between Android (built by Google/Alphabet) and Apple's iOS. Microsoft missed the boat and it's attempt to port Windows onto a phone and unify the user experience (beginning with Windows 7) was unsuccessful.

Uber (and Lyft) provide another interesting example of an indirect network effect. Uber's value depends on the availability of rides, which in turn depends on the number (and availability) of drivers. If drivers don't get enough work they may stop driving for either company, and if they do rides are harder to get and fewer riders will use the app; a vicious circle. Conversely as more user use the app, and drivers can make more money, more drivers will come to the platform, make rides more readily available and wait times shorter, increasing the value to riders and attracting more riders to the platform. In a sense there are two interlocking networks here; the value to riders increase with the number of drivers and the value to drivers is a function of the number of riders.

Chapter 15 – The Analysis of Social Networks

Social networks are more than Facebook, WhatsApp, Instagram, or Twitter. They are examples of platforms that are intended to facilitate the development and maintenance of social network ties. Social networks means simply a set of social relations and interactions. And they have been studied by academics long before Mark Zuckerberg was even dreamt of.

The study of social network combined the mathematics of graph theory with sociology (and generally resides within a sub-field of sociology know as economic sociology). Networks comprise all the ties between network 'nodes' (i.e. people), who are often treated as profit maximizes / cost minimizers in their network "exchanges". Properties such as network density, reachability (how many degrees of separation are they on average between people) diameter (what is the longest path between two nodes) have been the subject of considerable research. Reachability (and density to which it is often related) are important for understanding the spread of information as well as disease. Structural holes, gaps in an individual's network, have been lined to advancement in organizations[16].

Another important idea is centrality. There are several definitions of network centrality a simple one is how many ties (friends) a person has; the more friends, the more central and by implication the more influential that person is (that is termed 'degree centrality'). Another is "betweeness centrality" which is a measure of how many direct paths through the network a node lies on. Yet another is information centrality which looks at all the possible paths information might take though a network and rather like betweeness centrality, measures the information that might flow though a particular node, but instead of considering only direct paths it looks at all possible paths.

My preferred centrality measure is eigenvector centrality. Eigenvector depends as much on who you know as how many people you know. The idea is that it may be better to have a few influential (i.e., central) friends than lots of friends who are isolates or at the periphery of the network. As an aside Page and Brin, Google's founders used eigenvector centrality for their original page rank algorithm.

Understanding networks matters in many ways, not just for predicting the spread of information (or disease). Anyone who wants to make change happen needs to understand the structure of the network. Where are the clusters of people who are resistant to change? Are they connected to influential (central) people? Who needs to be persuaded so that others will follow? Where are the cliques who will vote as a block? Those who are successful change agents generally have a good feel for social networks even if they don't apply the analytical tools academic researchers might use.

---

[16] Burt, Ronald S., "Structural Holes: The Social Structure of Competition" (1992) Harvard University Press.

Chapter 16 – Artificial Intelligence

Artificial intelligence (AI) might be thought of as the ability of a machine to make inferences and choices by learning from large quantities of data. It can predict what movies we are likely to want to see on Netflix, what products we will want to buy on Amazon, and what advertisements might appeal on Google. AI recognizes objects and people, steers self-driving cars, detects uncharacteristic spending patterns that might signal fraudulent activity on our credit cards, and chooses what items to show us in our social media feeds. It might even be persuading us to make donations to charity.

It has been over 20 years since Deep Blue, an IBM chess-playing computer, beat chess grandmaster Gary Kasparov[17] . The victory of 'AlphaGo', an AI program developed by Google, over the world's top Go player[18] provides more recent evidence that AI can solve highly complex strategic problems of a kind that had been considered analytically intractable and therefore amenable only to human intuition. Activities that require computation are now routinely done with computers; many of us, when asked to multiply two numbers immediately reach for our phones (a pocket sized computer with more computational power than the large mainframes used to calculate orbital trajectories required to put a two men on the moon in 1969).

Complex knowledge-related tasks will almost certainly be supplanted by AI. And it is no longer just routine repetitive information processing tasks that AI will replace. White-collar jobs requiring years of training could, in the not-too-distant future, be performed by AI, with greater accuracy and consistency than by humans. By way of example, medical image diagnostics, once considered a job that required years of experience and human judgment, is becoming much more automated[19] [20] [21] [22]. Recognizing malignant blood cells or early stage tumors will soon be done by AI. More importantly perhaps, in addition to mimicking human visual recognition and interpretation, AI may soon prove to be capable of a variety of creative tasks. As BusinessWeek reported, AI is beginning to generate creative artifacts that could easily be mistaken for the work of little known Impressionists[23]. An AI generated song has been widely popular on streaming platforms[24]. Dall-e, has made AI image generation available to everyone.

ChatGPT3.5 caused a sensation when it was made available to the public in December 2022. ChatGPT was had the fastest adoption of any application, reaching over 100m users within just two months of its public launch[25]; almost overnight AI was in the news almost every day. That will only increase the rate

[17]https://www.scientificamerican.com/article/20-years-after-deep-blue-how-ai-has-advanced-since-conquering-chess/

[18] https://www.nytimes.com/2017/05/23/business/google-deepmind-alphago-go-champion-defeat.html

[19]https://www.technologyreview.com/the-download/610853/fda-approves-first-ai-powered-diagnostic-that-doesnt-need-a-doctors-help/

[20] https://www.technologyreview.com/s/610397/ai-could-alleviate-chinas-doctor-shortage/

[21] https://arxiv.org/pdf/1703.02442.pdf

[22] https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/45732.pdf

[23]https://www.bloomberg.com/news/articles/2018-05-17/ai-made-incredible-paintings-in-about-two-weeks

[24] https://www.rollingstone.com/music/music-features/ai-generated-drake-the-weeknd-hip-hop-fandom-1234720440/

[25] https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

of adoption of AI; firms whose work can be enhanced using AI but who are slow to adopt AI to increase their efficiency and effectiveness will fail and the fear of missing out will drive adoption.

In a 2015 study by McKinsey[26] designed to better understand AI's impact in the workplace, the authors looked not at jobs but at the activities within them. In summarizing their findings, the authors noted that: "about 60 percent of all occupations have at least 30 percent of activities that are technically automatable, based on [then] currently demonstrated technologies". The impediments to a significant degree of automation in white collar jobs are institutional rather than technological.

The extent of the disruption AI will bring about is perhaps unprecedented not so much because of its scale, but because of the speed with which change may occur. Before the industrial revolution, the vast majority of the labor force, around 80-90%, worked in agriculture; by the time the industrial revolution was over, the majority work in manufacturing and by the end of the 20ᵗʰ century, it was under 5%.. What allowed the transition to take place without creating massive social dislocation and unrest was that it took place over six to eight generations. By contrast, AI may transform the workplace in under a generation or a quarter of a person's normal working life.

It is certain that AI will bring enormous benefits. But while optimists suggest that AI will be used by humans to enhance their capabilities, it should be fairly clear that when one person, assisted by AI, can do the work of 10 people not using AI, two things are likely. First is that nine people are no longer needed to accomplish the same task as before. And unless demand for whatever that task is increases, many people will lose their jobs. Second, AI may well depress wages and salaries. That may lead to an increase in demand (if demand is elastic), but it will also mean that lots of well-paid jobs may command far lower wages and salaries. Both the lowering of salaries and the rise in unemployment are likely to happen quickly leaving many people struggling financially.

These are just the simple economic impacts of AI. A wide variety of other societally problematic effects should be anticipated, from the generation of fake news and deep-fake videos, to increasingly sophisticated cyber-attacks. And as we become increasingly dependent on AI, its lack of transparency—it's effetely a black box—will pose serious issues. As with any new technology. there are both opportunities and risks; but what makes AI different from previous technological discontinuities is not its impact but the speed at which it is developing. The rise of the internet has shown how difficult it is for societies and governments to keep up with and regulate new technology; the coming decade looks likely to present us with what may be one of our most daunting challenges yet.

---

[26]https://www.mckinsey.com/featured-insights/employment-and-growth/technology-jobs-and-the-future-of-work

Rational Choice

Economists (except Behavioral Economists) have a theory of choice that is more prescriptive than descriptive, but which whey argue yields the most economically efficient outcomes and should therefore be the way in which decision are made. They mean this in two senses; first, that were decisions to be made any other way, because that method is less efficient, it would be selected out (like unfit species in an the process of evolution) leaving only the most efficient. Thus only the most efficient method of making decision survives and thus this must be what is in use. The second sense is that economically rational individuals would never choose anything but the most economically efficient means of choosing the best course of action. Thus to the extent that people behave rationally, this is the way they should try to use when making choices. A large body of work, much owing its genesis the genius of Herbert Simon and James G March, points to considerable evidence that this is a relatively poor description of the way people actually make choices. Nevertheless, without some understanding of rational choice models, March and Simon's critiques will mean very little, so we begin with this as a foundation even if it is a straw man in a house of cards.

Expected value

Simple decision theory requires that you are able to calculate the expected worth or value of a course of action. This is essential if you are to make choices—which will be what you are expected to do as managers. One of the difficulties facing decision makers is commensurability. For example, how do you compare 10 years of life with the cost of a heart transplant operation? The answer is to turn everything into a dollar equivalent. For some problem this turns out to be quite simple. Is it worth recalling all the Ford Explorers fitted with Bridgestone/Firestone tires? The first step is to estimate the cost of the recall, based on the cost recalling a single vehicle and multiplying by the number of vehicles the company believes are on the road fitter with these tires. The alternative to the recall is to do nothing. However, this is not a costless option – there are estimable costs associated with this course of (in)action. Suppose one in a thousand Explorers is involved in an accident, and in half those cases the tires are the most likely cause, and in 1 in five of these, death or serious injury is sustained as a result. Based on the typical size of court settlement, one can estimate the likely cost of not issuing a recall. One could also add many other things to the mix such as the cost in future sales by appearing not to act in the customers' best interests. If the cost of the recall exceed the expected cost of not doing anything, rationally, Ford should simply deny that there is a problem, and if it comes to a class action law suit, reach a settlement with the victims and their families. Whether you think this is the right way to make this kind of decision is not the question here – it is simply an example of the rational decision theory approach. Next we'll look at a simpler (and less value laden) numerical example.

Suppose you have a lottery ticket in which there is a single prize of $100 and 100 tickets have been sold. You have a 1 in 100 chance of winning. In other words the probability of your ticket being the winning one is 1 / 100 or 0.01. If you were to play the lottery 100 times you would expect to win once. If you played 1,000 times you would expect to win 10 times. What then would be your average winnings per play? It would be the total prize money you could expect to win divided by the number of times you played the lottery.

If you played 100 times and won once, your expected winnings would be $100 divide by the 100 times you played or $1. If you played 1,000 times and therefore expected to win 10 times your winnings would be 10 x $100 or $1,000. Your average winnings are therefore $1,000 divided by the 1,000 times you played or $1. Notice that your average expected winnings remain the same irrespective of the number of times you play the lottery.

The expected value for this lottery, the average expected winnings for a single play is $1, and is simply the prize ($100) multiplied by the average probability of winning (0.01).

Suppose that each ticket costs $1. The total amount of money taken in by the person running the lottery will be $100. Imagine (as is usually the case) that the amount paid out in prizes is less than the total receipts from the sale of tickets – the person running the lottery takes $10 for their trouble. The prize is now only $90. The probability of winning is still 0.01 as 100 tickets are sold. So now your expected value is only 90 cents. Since this is less than the price of a ticket, you should probably not play the lottery. There is a possibility, albeit rather remote, that you will buy the winning ticket first time around and walk away with an $89 profit. But more likely you will win nothing and incur a $1 loss. You could improve your odds by buying 2 tickets rather than one. Your probability of winning is now 0.02 but the cost has risen to $2. The expected value of your two tickets is $1.80, while your outlay is $2. Suppose you bought all the tickets – does this help? You will increase your chances of winning to 1, (i.e., certainty) but you will win $90 for a cost of $100. So things don't get any better the more tickets you buy. Much the same argument applies if you play the lottery many times. If you play once, the expected value (your expected winnings) is $0.9 and the cost is a dollar net you are likely to be out of pocket by 10 cents. Overall then, this has a net expected value (the EV of the lottery less the price of the ticket) of -$0.1.

Rational Choice Decision Trees
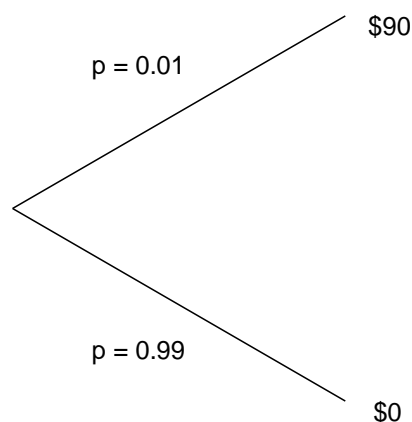
Now you can cast this as a simple decision tree. In the tree, each branch is a possible outcome. There are two here, win or lose. Each branch has two things associated with it – the probability of that outcome occurring and the value of the outcome. First we calculate the expected value of each branch. The value of the branch is the product of the value of the outcome and its likelihood. For the top branch this is $90 x 0.01 = 0.9 and for the lower branch it is 0.99 x 0 = 0. Now we add them together and this is the expected value of the lottery.

$90

p = 0.01

p = 0.99

$0

We write this: $EV = \sum_{i=1}^{n} p_i x_i = p_1 * x_1 + p_2 + x_2 = 0.01 * 90 + 0.99 * 0 = 0.9$

Here n, the number of branches, is 2, i is a particular outcome, p_i is the probability if that outcome and x_i the value associated with it.

Now we can work back to a choice or decision as to whether to play the lottery or not by adding another branch before the lottery.
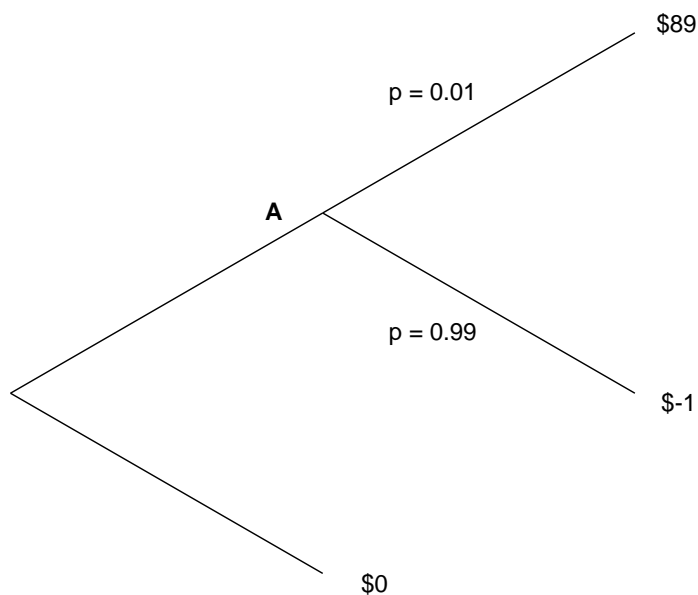
The first 'fork' represents buying a ticket (up) and not buying a ticket (down). Once you have bought a ticket there are two possible outcomes, win or lose as above. Here the outcome numbers reflect the choice of buying the ticket ($1). The net winnings in the top branch is only $89 and in the lower branch shows winning nothing less the $1 for the ticket.

The EV of the lottery should be the same as before.

$89

$p = 0.01$

A

$p = 0.99$

$-1

$0

$$EV = \sum_{i=1}^{n} p_i x_i = p_1 * x_1 + p_2 + x_2 = 0.01*89 + 0.99*-1 = 0.89 - 0.99 = -0.1$$

A   EV = $-0.1
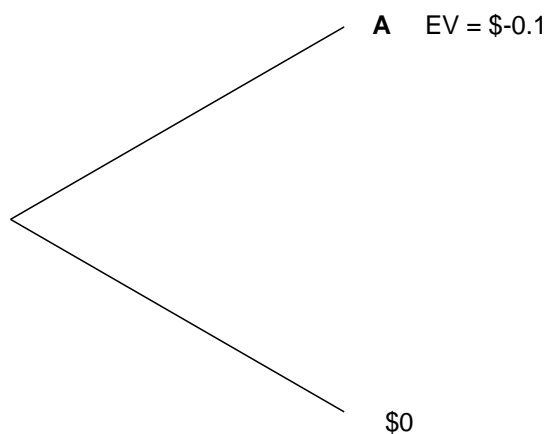
Working back from the outcomes on the right we can 'condense' the lottery part of the three to a single number (its expected value) at point A which is $-0.1. Now we can discard the lottery branches because they are represented by this single expected outcome.

The lottery is now represented in by its expected value and the decision facing us is to buy a ticket and expect to lose 10 c or no buy a ticket and lose nothing.

$0

Suppose you have two lotteries, one with a prize of $1,000,000 dollars in which a million tickets have been sold each for $1 and a smaller lottery in which the prize is $10,010, and only 1000 tickets are sold each for $10. The first lottery has a probability of winning of one in a million. Multiply this by the prize and you have an EV for the top branch of $1 (the price of the ticket). Ignore the lower branch because you are using zero as the outcome value. The second lottery has an EV of $10.01. From each you can subtract the price of the ticket to get a net EV for the lottery of 0 in lottery 1 and 1 cent in lottery 2. Now there are three possibilities – buy a ticket in lottery 1, buy a ticket in lottery 2, or do nothing.

Doing nothing and lottery 1 are equivalent. Each has an EV of 0. However lottery 2 is slightly more attractive, having a net EV of a cent, so rational choice decision theory suggests you should play lottery 2.

Chapter 18 – Decision Making and Rationality

Economics, or at least neoclassical economics, makes some simplifying assumptions, often to help make the models tractable. Rational choice theory, in its purest form, assumes that decision makers are able to process all available information. They are assumed to have all the information they need to make a decision. They are assumed to be rational and not swayed by emotion, and not to be subject to cognitive bias. That's a pretty bold set of assumptions.

A more nuanced version of rational choice theory adds risk or uncertainty; a decision no longer has a single outcome, but several possibilities are envisioned. The decision make does know the probabilities of each however. This was the variant of rational choice theory discussed in the last chapter.

Decision making

There are however some problems with the model, namely that there is now 50 years of research showing that this isn't generally how decision are made, even though we may like to think it is how decision should be made. For example, not all alternatives are exhaustively enumerated; neither are all the consequences of each alternative. Not all alternatives and consequences are considered simultaneously, as some are only considered when others have been ruled out. Decision makers seldom optimize (that is chose the very best alternative but 'satisfice', that is choose the first acceptable solution they come to in a sequential search.

There is also the problem of cognitive bias. A common one is the inability to correctly estimate likelihoods of outcomes. For example, the probability of a highly salient event may be greatly over-estimated, particularly for rare events. For example, are you more likely to be killed in a plane crash than stuck by lightening[27]? Because very rare events are never experienced by most of us, there is a tendency to ignore them completely (essentially assuming their probability is 0). But if there is a very large number of highly improbable outcomes, although each individually is unlikely, one of them is probably going to happen; imagine you have a thousand outcomes each with a one in a million chance of happening; there is a one in a thousand chance, not by any means negligible of one of them happening.

Another cognitive limitation is attribution bias. We generally attribute successful outcomes to our own competence and negative outcomes to forces beyond our control. After a string of successes we are generally likely to over-estimate the likelihood of a successful outcome, since we consider the situation as more controllable and manageable than it really is. This bias often comes into play in decision to acquire other companies where for example, the odds of achieving the anticipated savings, are generally over-estimated particularly by companies who have made successful acquisitions in the past.

Rule-following

Many choices, however, aren't made this way at all. Rather, we follow rules about what ought to be done in a given situation. Rules are useful because they short-circuit the need to acquire knowledge

---

[27] Odds of being stuck by lightening in a year are about 1 in a million; odds of being killed in a plane crash about 1 in 4.7 million,

and experience. Other people mishaps, when enshrined in company policy for example, guide action without our having either to search extensively for an optimal solution (so economizing on effort and time) or go through the same experience ourselves to figure out what will happen.

For example, IBM, had tools to help ensure that large systems were correctly configured, that bundled the hardware and software designers' knowledge into a tool that require only that the field engineer answered some fairly simple questions about a customer's requirements. And before each installation, field engineers went over a detailed pre-installation systems assurance checklist with a customer before the equipment was delivered: Was the false floor in place? Was the air conditioning and cooling sufficient? Was air the filtration adequate? And these are the obvious ones I can remember. But the actual checklists were several pages of questions I wouldn't have even thought to ask.

"The experiential lessons of history are captured by routines [including checklists, rules, policies and standard operating procedures] in a way that makes the lessons, but not the history, available to organizational members who have not themselves experienced the history". (Levitt, 1988)

Decision making though the application of rules follows a different logic. If rational choice is seen as a logic of *consequences*, rule following is a logic of *appropriateness:* does the situation meet a set of criteria to evoke a certain rule. Once evoked, the choice is pre-determined. A great deal of organizational activity looks rather like this. Of course, these are two extremes, and many decisions may be part rule following to prescribe a set of allowed alternatives with some boundedly rational choice within that set. ('Thinking outside the box' is a way of describing looking for alternatives not generally evoked in a given situation).

It's also worth remembering that rules aren't always the outcome of experience and learning. Sometimes they are simply adopted for reasons of legitimacy; an accounting firm has to follow accounting rules that may not make sense but that are needed if it is to retain its standing as an accounting firm. Sometimes they are copied from other organizations; benchmarking is an example. And sometimes they are the outcome of bargaining that may have more to do with distribution than optimization. Because much of the decision making in organizations is the outcome of rule following, it would be foolish to assume that firms can necessarily be treated as rational actors.

Emotion

Emotion is also important. While Mr. Spock may epitomize the ideal rational decision maker, in reality, many seemingly simple decisions, like buying the optimal kind of breakfast cereal, simple are intractable when approached though logic alone. Studies of people who have though brain injury lost the capacity to feel emotion are often incapable of making even the most trivial decisions, like which color pencil to use. Emotion plays a crucial role in facilitating choice when the task required to solve a problem logically exceeds our cognitive capacity.

Intuition falls somewhere between a purely emotional choice and a purely logical decision. We often have a hunch based on experience that our subconscious tells us is the way to go even though we may not be able to articulate precisely why. And since many managerial decisions have to be made without complete information, intuition is an important part of the manager's tool kit. And its accuracy grows with experience.

Summary

We tend to think of decision makers as heroically gathering all the available information, mulling over all the alternatives and their consequences, and picking the very best course of action. But in reality that not how all or even most decisions are made. Even when a conscious attempt is made to follow this model there are limitations of cognitive capacity and available information, not to mention other cognitive biases, that lead to a departure from this ideal. And many decision aren't this way at all; some are made though heuristics, for example, dealing with options sequentially until a satisfactory one is found; some are made by following rules, and some are made through gut feel and intuition.

Appendix A – Financial Statements

Profit and Loss (P&L) or Income Statement

A Typical Income Statement
Sales (or revenue)
COGS
Gross Profit
SG&A
Depreciation
EBIT
Interest Expense
EBT
Tax
Net income

The P&L is an accounting statement of flows or changes between two points in time general a quarter or a year. It is the complement to the company's balance sheet which shows a snapshot of value at a point in time. The changes between one snapshot (balance sheet) and the next generate the record of changes, the 'flows' reported on the P&L.

The top line (see inset) is always the money booked as flowing into the company (though it may not actually arrive – tracking what actually comes in is the cash flow statement). All the other lines are (generally) outward flows; labor and materials costs, utilities, rent, salaries and commission, R&D, and advertising. There are often totals at different points as you work down the statement, shown in italics in the illustration that provide different insights into the company. These are typically gross profit, operating profit or earnings before interest and taxes (EBIT), earnings before taxes (EBT) and net income. For example, a company may have a healthy operating profit but could still be unprofitable if its interest payments are too high. Or, it may have income from investments in other companies and report a profit even though its operations are not profitable, as was in essence, the case for Yahoo.

All P&Ls look something like this; while seeming rather sparse and uninformative, this standard allows you to make simple comparisons across companies. The structure also allows you to separate the operational costs from the financing costs and, broadly the direct costs (that vary directly with the level of activity) from fixed costs that don't, For example Cost of Goods Sold (COGS) would be expected to doubles when the number of units sold doubled, while sales, General and Administrative Expenses (SGA) will be the same whether you make lots or very few of whatever you are making. That has implications for strategy. A firm with mostly fixed costs (such as a company that writes software) may be expected to respond differently to downturns in the market than a firm which has mostly direct costs. The former needs to keep volumes high to keep the assets (which are in effect a fixed cost) adequately utilized. To do this the firm may cut prices to (or sometimes below) long run average costs. Conversely, firms with mostly direct costs need not worry as much about changes in demand and may be less prone to price cutting in a downturn. A firm whose "Times Interest Earned", the ratio of EBIT over interest expense is close to one is a firm that can barely cover its interest payments (not good). "Times Interest Earned" less than 1 probably means the firms is going broke.

Cash Flow

Like the P&L, the reflects firms cash flow statement the firm's activity over a given period, but unlike the P&L it omits things that do not actually involve money being spent even if it is "recognized" in the P&L. For example, the P&L will generally include depreciation, but depreciation does not involve money being spent; rather it reflects the change in the value of an asset the company owns reported on the balance sheet and is intended to reflect the declining value of those assets over time.

Cash flow statements are generally divided into three parts; cash flows from operations, cash flows from investing activities and cash flows from financing activities. Because the last two are not operational cash flow the items there are not recorded in the profit and loss statement which reflects the profit (or loss) associated with the operation of the business. Operational cash flows will include money actually received from customers. Sales on credit are booked as sales in the P&L, and will be added to the accounts receivable in the balance sheet but are not cash flows. Similarly, money received from sales on credit in an earlier accounting period will appear as a positive cash flows but will not appear in the P&L (that money has already be recognized in the prior period); and it will show up as a reduction in accounts receivable. Similarly, purchases from suppliers paid on credit will not show up in the cash flow, but will add to accounts payable in the balance sheet.

To calculate the cash flow from a profit and loss statement (the "indirect method"), you must remove any expenses shown in the P&L that are not cash flows. The three most obvious items are depreciation, previsions for future liabilities (for example expected settlement of legal actions) and any tax loss carry forward. All of these reduce taxable income but do not involve any cash outflow.   So, to "reconstruct" the cash flow you need to add back any expenses from the income statement that do not involve money (cash) being spent.

Balance sheet

While the P&L reflects the firm's activity over a given period, the balance sheet is a snapshot of what the company owns and owes at a point in time. The sum of the things the company owns less the debts its owe to others is the amount that belongs to shareholders, the company's equity. Assets are generally ordered by how easily they might be liquidated. Cash which is already "liquid" comes first, then things that are easily sold like marketable securities (which is different from the an ownership of a subsidiary since the latter might be much harder to dispose of than a few shares in an exchange traded money market fund). Accounts receivable is an assets (it's money owed to the firm) that should relatively easily accessible but does depend on whether the customer can actually pay up. Generally towards the bottom are things like plant and equipment and fixed assets like buildings which are perhaps the hardest to dispose of.

On the liabilities side you generally have two categories; current liabilities which are debts that fall due within a year, and non-current liabilities which is any money owed that does not fall due within a year. Under current liabilities you will typically see accounts payable, that is money owed to suppliers, and short term debt (money owed to banks). Long term debt, a non-current liability, generally refers to corporate bonds the company has issued.

Equity, the difference between what an company owns (its assets) and what it owes (its liabilities) often is divided into paid-in capital and retained earnings. Paid in capital is the money the company has taken in from sales of its shares, while retainer earnings, as the name implies, is the sum of contributions from its business operations after all expenses, taxes and dividends have been paid.

Connections

When accountants talk about financial statements "tying together" they are looking for the connection between the three sets of statements to make sure that they are consistent. For example every year the company makes a profit, what is left after paying shareholders a dividend (if the company issues one) gets added to retained earnings; so by comparing two successive balance sheets, the difference in retained earnings from one period to the next should equal the net income, less dividends paid, reported on the P&L for that period.

Similarly, the net cash flow in the cash flow statement should equal the difference in cash shown on the balance sheet from the end of the prior period to the cash reported on the current periods' balance sheet. The cash flow statement typically shows the beginning cash balance (i.e. the prior period's ending cash balance) and the current period's ending cash balance with the net cash flow being the difference between them.

Another connection between financial statements is depreciation. Suppose your company buys a machine tool for $100k. The tool is expected to last 5 years before needing to be replaced. Using the straight line method you would reduce the value of the asset on the company's balance sheet by $20k every year so that at the end of 5 years it is, at least in accounting terms, not worth anything (even if it is still working). Every year for the 5 years you depreciate the assets, you also enter the $20k as a depreciation expense on the P&L. Depreciation in the P&L is not a real expense, no money changes hands; instead it reflects the change in the value of the asset reported on the balance sheet.

Appendix B – Financial Metrics

Return on Sales (or profitability) is the ratio of income or earnings to sales. Another term for sales is revenues. Income can be before interest and taxes. If we are talking about earnings before interest and taxes, I will refer to this as 'EBIT'. ROS can also be calculated using earnings; after deducting taxes and interest expense. If we are talking about earnings after interest and tax, I will use the term 'net income' but sometimes refer to this simply as earnings or income. When you present an ROS number you should say which you have used.

$$ROS = \frac{EBIT}{revenue} \ \text{or} \ ROS = \frac{net \ income}{revenue}$$

Having a low return on sales isn't necessarily a bad thing. Imagine you are Wall Mart and you earn only 1% on sales. This looks bad at first glance. But suppose you keep each item for a month only (i.e., your inventory turnover is 12) then you are probably doing better in terms of ROI than a firm with 10% return on sales but which keeps items in stock for an average of a year. ROS is easy to calculate but useful only to benchmark the company against other firms in its industry and against its own past performance.

Return on Equity is the ratio of income to shareholders equity (as presented in the balance sheet).

$$ROE = \frac{net \ income}{shareholders \ equity}$$

This is perhaps a more useful measure since it doesn't depend on the rate of inventory turnover as ROS does. Because a highly leveraged company (one with a large proportion of its liabilities in debt rather than equity) will have higher earnings before tax the more debt is has, it's probably better to use net income (i.e. after the interest and tax have been deducted) in this ratio.

Image a firm with no debt earns $10 million before interest and taxes and has $100 million in equity. Calculated using EBIT, its ROE will be 10%. Now imagine that its managers go to the bank and borrow another $100 million to buy a competitor of about the same size. Its EBIT will double to $20 million. If we used EBIT here the ROE would rise to 20%, but we would be ignoring the interest payments it has to make on the money it has borrowed.

Using net income (i.e., after interest expense and taxes) we get quite a different picture. If the tax rate is 30%, then ROE after interest and taxes prior to the loan would be 7%. If the interest rate on the loan is 10%, then interest expense is $10 million and net income after taking out the loan and buying the competitor is still $10 million. So the ROE after interest and taxes would remain the same at 7%. However, if the money was borrowed at a lower rate, say 5%, then net income would be $15 million and ROE would be 10.5%. This measure is particularly interesting to shareholders because it is an indication of what they can (or should) expect in terms of dividends and/or the residual value of the firm.

Return on Assets is the ratio of income to the total assets of the company (as presented in the balance sheet).

$$ROA = \frac{EBIT}{total\ assets}\ or\ ROA = \frac{net\ income}{total\ assets}$$

Return on assets is not sensitive to the company's level of debt because it includes both debt and equity in the denominator. It is also not sensitive to inventory turnover like ROS. As far as I'm concerned you can calculate this using either EBIT or net earnings. It gives a measure of how efficiently the company is using its capital asset base. If you are a retailer, think of this as how much money you make per square foot of floor space as well as how much you make on the value of your inventory.

If a company takes assets of its balance sheet, for example by selling its production facility to another company and then contracting with that company for the use of what was once its own equipment, its ROA might well increase. This assumes that in selling its assets, the reduction in debt service payments and the cost of operating this facility were more than it will now be paying the company to whom it has sold the facility and from whom it is now contracting for its use.

Share price

In a project presentation a few semesters ago, someone asserted that company X was better than company Y because X's share price was higher than Y's. Imagine that company X's share are trading at $120 a share and Y's at $70. Now, X announces a 2 for 1 stock split. It shares should now trade at around $60. Does this mean that X as a company is worth only half what it was before the split? No. Does it mean that X is doing only half as well after the split as before? No again. This simple example shows that share price alone offers no real indication of firm performance.

A change in share price is another matter. If shares increase in value this suggests that investors value the company more highly now than they did. That's a good sign. But does it tell us anything about the company? Perhaps, but more data are needed. Suppose the company's share price has risen 20% in the last year but all other firms in its industry have risen by 30%. That tells you that the company may be doing better but it is not doing as well as its competitors. And one only has too look at the huge rises in share prices of many of the firms in the Valley during the dot-com book to see why share price rise alone is not enough to tell us if a company is likely to be successful.

One indicator I particularly like is the price-earning (PE) ratio. The PE ratio links the company's present performance to the market's expectations about the future. PR ratio is the ratio of the firm's share price to its earnings per share. The earnings per share is a measure of current performance while the share price is an indicator of market expectations about future earnings, in a crude sense a measure of the net present value (see next section) of the company's discounted future earnings.

$$PE = \frac{price\ per\ share}{earnings\ per\ share} = \frac{price\ per\ share}{net\ earnings \Big/ number\ of\ shares\ outstanding}$$

Appendix C – Valuation

Valuing a company is not an easy task. McKinsey, a strategy consulting firm, wrote an entire book on the subject in 1990[28]. A lower bound might be the value of the company's assets less the amount of debt is owes (in other words the value of its equity). However that is a backward-looking value in that the equity reflects decisions that have been already been made regarding for example, payments to shareholders as dividends, or investments in advertising and brand development. At its simplest, McKinsey suggests using the future value of the company cash flows to determine what it is worth. The approach works for companies that are either not growing, so the cash flows are constant or for companies that are growing relatively slowly, specifically at a lower rate than the discount rate to be used in this simple approach. It is usual for purposes of valuation to apply the prevailing tax rate to the cash flow to see what the income would be without the non-cash deductions and use this figure to calculate a company's value. With the cash flow net of tax, and assuming that cash flows are constant over time, one then needs only to use the handy formula below for calculating the present value of a constant in-perpetuity cash flow to arrive at a valuation.

$$PV = \sum_{t=1}^{\infty} \frac{CF_t}{(1+r)^t} = \frac{CF}{r}$$

The present value is the sum of all future cash flows, *CF*$_t$, discounted by 1+r for each year in the future the cash flow arrives. For cash that arrives 3 years out (t=3) the present value of that particular cash flow is CF$_3$ / (1+r)^3. A cash flow that is anticipated to arrive in 10 years would be worth CF$_{10}$ / (1+r)^{10}. Adding them all up out to ∞ (i.e., in perpetuity) gives you the value of the expected cash flows and hence an estimate of the value of the company. This is greatly simplified by assuming the company will not grow, and hence the cash flows are constant, hence the present value can be simplified to

    PV = Cash flow / discount rate

In this course, that's the formula I will expect you to use.

---

[28] Valuation: Measuring and Managing the Value of Companies, Copeland, Koller and Murrin, McKinsey and Company, 1990, John Wiley & Sons.

Appendix D – Bertrand and Cournot: Two Models of Oligopoly Pricing

Background

Bertrand and Cournot were both economists; and, coincidentally, both French. They held different views on the way prices should vary with *industry concentration*. Bertrand's view was monopoly or nothing; his logic was intuitive and compelling. Cournot's prediction was harder to grasp but is more appropriate (and gives better results) in situations in which operations are preceded by significant up-front investments.

In brief

Bertrand predicted that profits would be earned (i.e., equilibrium prices above long run average costs) only by monopolists. Once a competitor entered the market, prices would inevitably be competed down to long run average costs and neither firm would turn a profit.

Cournot predicted that more firms would lower profits, but not as dramatically as Bertrand predicted. Cournot's model suggests that profits decline as the number of firms in an industry rises.

Bertrand's model of oligopolistic competition

Bertrand suggested that while one should expect a market supplied by a single firm would command a high monopoly price and generate monopoly profits, when two or more firms were competing prices would, in equilibrium, fall to the firms' long run average costs.

Suppose we have two firms. Each does the calculation for profit maximization to determine the optimal quantity this industry should be putting onto the market (that is optimal from the point of view of the firms in the industry, though clearly not from the buyer's point of view). Assuming they have good information they should both arrive at the same answer, $q*$. Since there are two of them, their best joint outcome is to divide the quantity in two, and hence divide the profits equally between them.

However, Bertrand reasoned that they now have a problem. The first potential buyer appears and both firms would like to have him as a customer. First one firm lowers its price to try to make the sale; then the other firm responds in kind and before long, both firms have reduced their price to their long run average costs. The same logic holds for three firms and above. Thus Bertrand suggested that the relationship between the prices we see and the number of firms in the market would look like this.
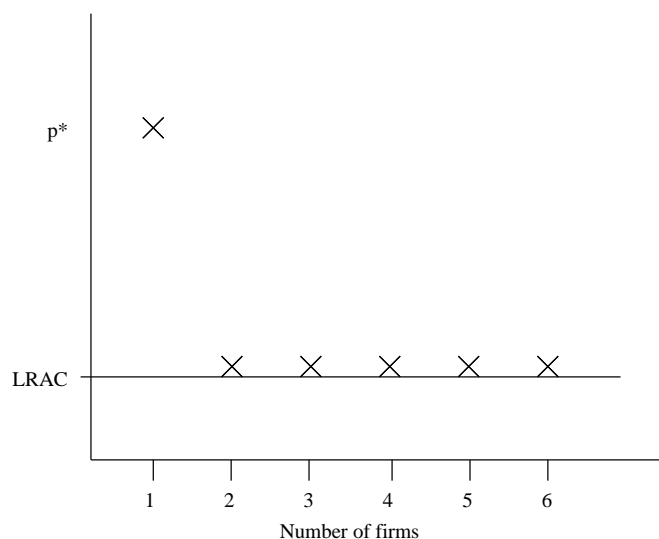
In other words, when there is only a single firm, it will set a monopoly price, make a large sum of money and thus reduce consumer welfare. In all other cases, firms make 'normal returns', that is, they just cover their long run average costs.

Cournot's model of oligopolistic competition

Cournot's model of competitive behavior in some ways anticipates a much more recent development in economics; non-cooperative game theory. John Nash (you may well have seen the film A Beautiful Mind) won the Nobel Prize in economics in 1994 for his work in game theory.

In order to see how Cournot's model works we need to begin with the simplest possible case of a single firm trying to find the price that maximizes its profits. The take-away here is that there is a sweet spot (neither to high, nor too low) at which the firms' profits are highest.

The simple case is of a profit maximizing firm facing a downward sloping demand curve. A monopoly means there is only one supplier of a good or service. Buyers are price takers meaning that no one customer buys enough to exert any bargaining power over the supplier. There are many examples of monopolies: AT&T used to have a monopoly in long distance telecommunications, IBM was investigated in the 1970s for monopolizing the System/360 architecture market place and more recently Microsoft for its possible abuse of a monopoly position in desktop operating systems. And consumers are generally price-takers, at least when buying things like groceries, or anything from a store where one customer's specific purchase has a very small, almost insignificant, impact on the firm's profits (although consumers as a whole do matter).

Since price and quantity are related (any given quantity produced will have a unique price at which no one that wants to buy at that price can't get the product and the firm's inventory is completely used), we want to find the quantity it should produce and the price is should set to make the highest profit.

If you are not particularly interested in the mathematics, please skip to the conclusions.

Monopoly profit maximization

When the market is supplied by only one firm, the firm calculates the monopoly quantity and market clearing price (or monopoly price and market clearing quantity).The starting point is the demand curve:

$$p = 1 - q \tag{1}$$

Price, p, is equal to 1 – quantity, q. This is a simple downward sloping linear demand curve. Real demand curves are not likely to be linear not have a slope of 1, but this simplification makes it easier to see what is going on.

Profit is calculated by subtracting costs from revenues. Another simplification is to assume that all costs are variable (no fixed costs). Again the intuition is unchanged by this simplification.

Profit, π, is given by $\pi = p*q - c*q = q*(p-c)$ (2)

Substituting for p from equation 1, we have an equation in c, which is fixed, and q.

$$\pi = q*(p-c) = q*(1-q-c) = q - q^2 - qc \qquad (3)$$

This can be seen in graphical form in the following figure:
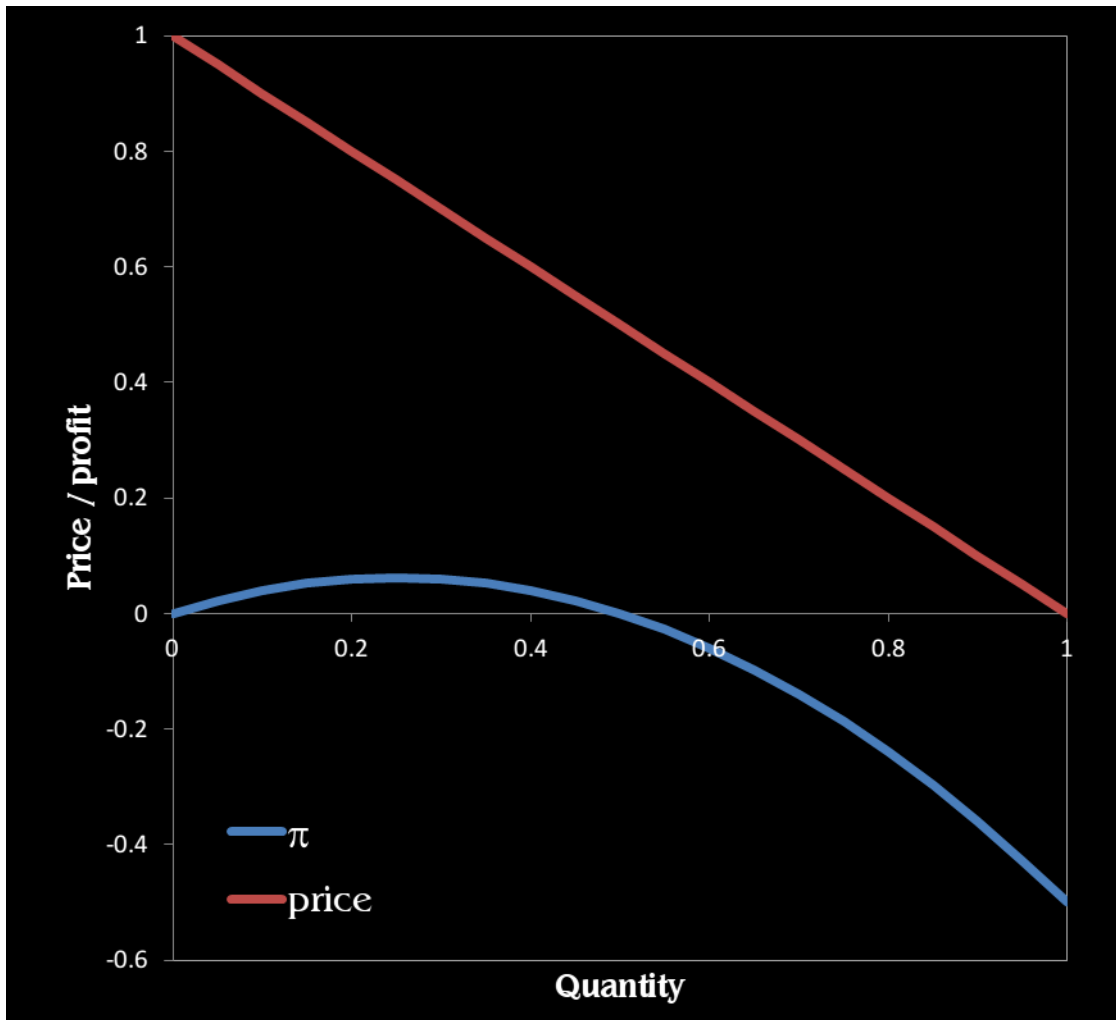
Profit, from equation 2 is the blue line; the demand curve, equation 1, is shown in red.

To find the quantity at which the most profit is made, we differentiate equation 3 with respect to quantity.

$$\frac{d\pi}{dq} = 1 - 2q - c \qquad (4)$$

This first order differential with respect to quantity is the gradient of the profit / quantity curve.

To find the maximum profit, we need to find the point at which the gradient is zero, so we set equation 4 to zero.[29]

$$1 - 2q - c = 0$$

$$q = \frac{1-c}{2}$$

If we call this optimal quantity $q*$ the corresponding monopoly price, $p*$ is $1 - q*$. These are the profit-maximizing quantity and price.

In the graphical example, costs c were set at 0.5, so q = 0.25. Note also that the profit line crosses the axis at q = 0.5, which is also p = 0.5. This you would expect; with p = 0.5, and c = 0.5, profit would be zero.

This is when one firm commands the entire market. With two firms one might expect they would each make half as much money. However, the answer is not quite that simple.

Duopoly profit maximization

In the basic Cournot model, two firms set the quantities of product they will manufacture and then adjust prices to maximize their own profits once their production quantities are set. This requires each firm to anticipate the quantity the other firm will produce in order to estimate the price that will ultimately prevail and then chose its own production volume accordingly. The notion of making a choice based on the anticipated choice of an opponent (who is herself making a symmetric calculation about your likely choice) is central to game theory.

Given two firms, *i*, and *j*, and a demand function *p=1-Q* where *Q* is the total quantity of the good supplied and *Q* is the sum of the quantities supplied by each firm, $Q = q_i + q_j$

First we calculate firm *i's* profit, π_i

$$\pi_i = pq_i - c_i q_i \tag{1}$$

Again, p is the price (note that $p = p_i = p_j$ - since this is a commodity product, were one firm to charge more than the other it would sell nothing).

Substituting the demand function, $p = 1 - Q = 1 - q_i - q_j$

$$\pi_i = (1 - q_i - q_j)*q_i - cq_i = q_i - q_i^2 - q_i q_j - c_i q_i \tag{2}$$

Differentiating with respect to q_i, and setting the first order equation to 0, in other words i choses the price that maximizes his profit contingent on j's chosen production quantity.

---

[29]   We could do the same thing by differentiating with respect to p rather than q – this would mean substituting for q rather than p.

$$\frac{\partial \pi_i}{\partial q_i} = 1 - 2q_i - q_j - c_i = 0$$

Gives $\quad q_i = \dfrac{1 - q_j - c_i}{2}$ $\hspace{4cm}$ (3)

A symmetrical result for firm j, $\quad q_j = \dfrac{1 - q_i - c_j}{2}$ $\hspace{3cm}$ (4)

Substituting 3 into 4 to solve one firms production quantity gives

$$2q_i = 1 - \frac{1 - q_i - c_j}{2} - c_i$$

$$4q_i = 2 - 1 + q_i + c_j - 2c_i$$

$$q_i = \frac{1 + c_j - 2c_i}{3} \quad \text{and symmetrically} \quad q_j = \frac{1 + c_i - 2c_j}{3}$$

We have found the production quantities for firms i and j, contingent on their unit cost, c_i and c_j. If c_i and c_j are equal, then each produces the same quantity, $q = \dfrac{1 - c}{3}$.

Now compare this to the monopoly case, $q = \dfrac{1 - c}{2}$; if both firms produced half the profit maximizing quantity, each would produce $\dfrac{1 - c}{4}$, yet Cournot shows that each will end up producing more than this. The quantities predicted by Cournot are larger, and hence the market clearing price (given the larger amount of goods available for sale) will be lower.

If the two firms divide the optimal monopoly quantity in two, each makes a profit of $\dfrac{1 - 2c + c^2}{8}$, yet Cournot shows that in his model, each makes slightly less, $\dfrac{1 - 2c + c^2}{9}$.

The same analysis can be done for three or more firms; and as the number of firms increases the profit each makes goes down.

Extension to more than two firms

The figure below shows how the equilibrium profit-maximizing price, p·, varies with the number of firms for three different levels of variable cost. The horizontal dotted line shows the level to which the price line for c=0.5 is converging asymptotically. The dashed line shows the level to which the middle line (c=0.25) is converging and the lowest line (c=0) approaches its variable cost at zero.
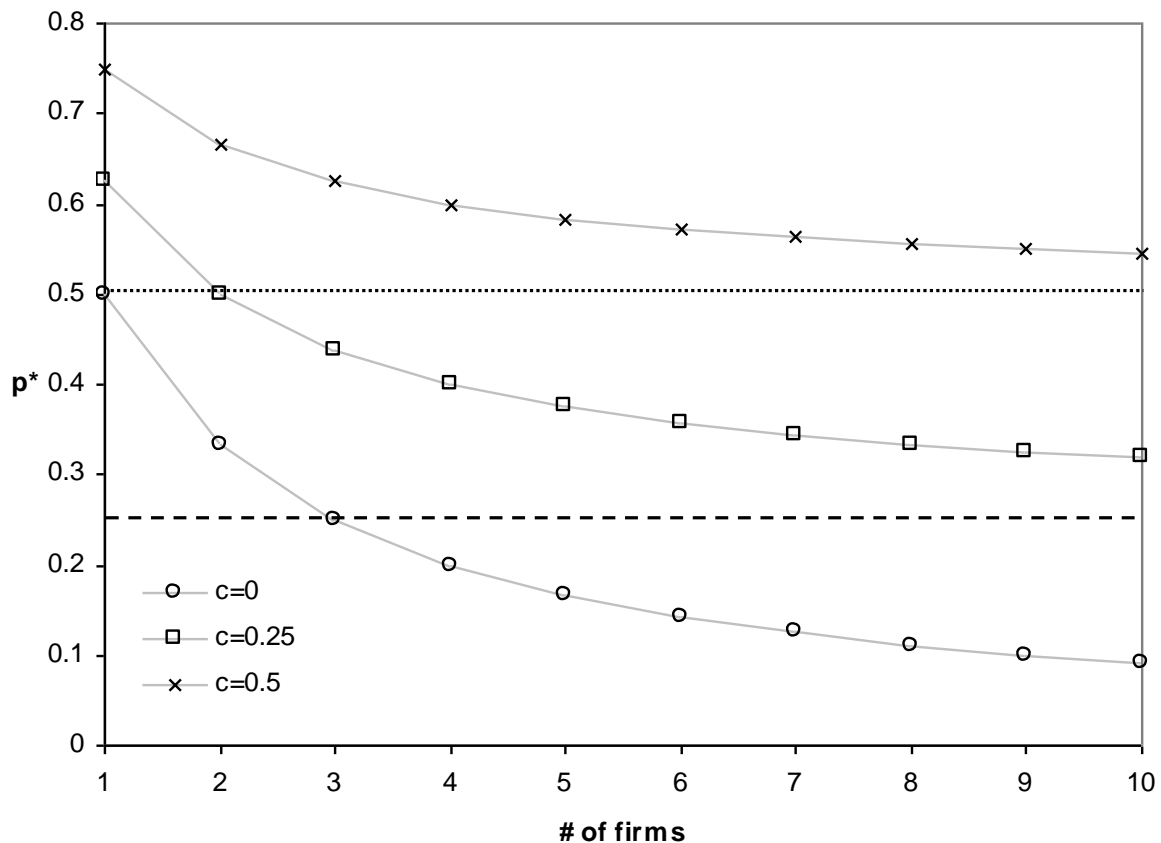
This looks quite different from the Bertrand model – Bertrand predicted that a single monopoly firm would make money but as soon as there were two firms in the market, neither would make above normal returns. Cournot suggested otherwise; although prices decline from the optimal monopoly price as the number of firms increases, there are still profits to be made in concentrated markets where the number firms is relatively small (which good news for firms like Intel, Hertz, BASF, SKB, Ford, Boeing, Bank of America to name but a few).

Collusion, Cournot and Bertrand compared

The following figure shows the profit that *might* have been made if producers could have held the line and together, as an industry, produced no more than the monopoly profit maximizing quantity. This would be the outcome one might expect if firms colluded to fix prices and quantities. The second line shows the profit Cournot predicts when firms make optimizing choices that are contingent on their expectation of the other firms' choices. Cournot's model show profits as lower than that expected when firms collude. Finally we have Bertrand's rather gloomier prediction.
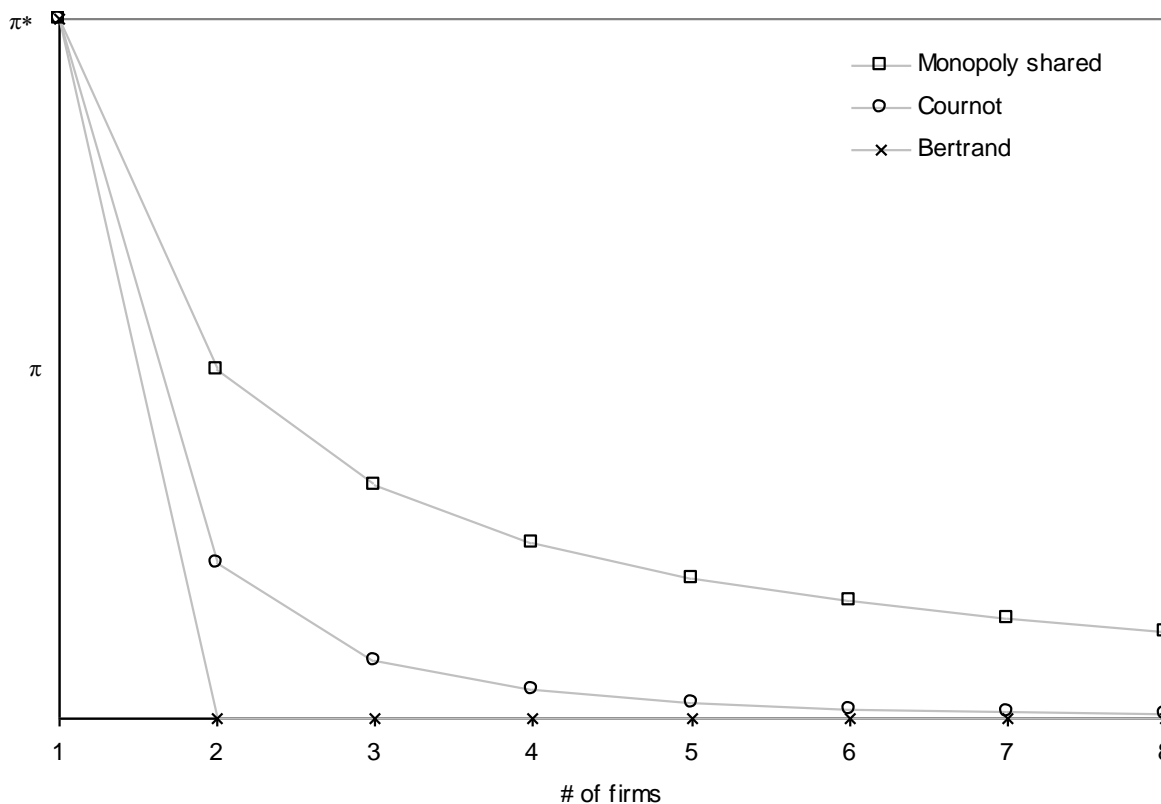
Profit is shown as a proportion of the single firm monopoly profit.

It's also worth noting that we haven't talked about things like customer loyalty and switching costs. In Cournot's model, oligopoly profits are made even without assuming any friction in customer switching from one supplier to another. Add these to the mix and producer profits are higher still.

Summary

The two models make different predictions for concentrated industries or oligopolies. Why the difference? It boils down to sunk costs. When firms are required to make an ex-ante commitment which will alter their ex-post cost (long run average costs) they must choose a price and stick to it. This involves in essence a game-theoretic calculus of the competitor's intentions and they in turn will be making the same calculations about you.

Where business is mostly about trading and little in the way of investment in fixed costs is needed, Bertrand's model may apply. I say may because collusion, tacit or otherwise, may allow prices to remains above the p = LRAC of Bertrand's model. But when firms have to make big bets and plan ahead and anticipate others doing the same, Cournot's is a more appropriate model.

Bibliography

Coase, R. H. (1937). "The Nature of the Firm." Economica 4: 386-405.

Mintzberg, H. (1987). "The Strategy Concept 1: Five Ps for strategy." California Management Review 30(1): 11-23.

Porter, M. E. (1980). Competitive Strategy, Free Press.

Rumelt, R. P., Ed. (1984). Towards a Strategic Theory of the Firm. Competitive Strategic Management. Englewood Cliffs, New Jersey, Prentice Hall.

Rumelt, R. P. (1991). "How much does industry matter?" Strategic Management Journal 12: 167-185.

Williamson, O. E. (1981). "The Economics of Organizations: The Transaction Cost Approach." American Journal of Sociology 87(3): 584-577.

Williamson, O. E. (1996). Transaction Cost Economics and Organization Theory. The Mechanisms of Governance. Oxford, Oxford University Press: 219-249.

Williamson, O. E. (2006). "Transaction Cost Economics and Business Administration." Scandinavian Journal of Management 21: 19-40.