

Public Lecture  
Sabanci University, October 2008  
“Reference Points and the  
Theory of the Firm” (*Economica*,  
August 2008)

Oliver Hart – Harvard University

# Part 1: Coase (1937) and what followed

- Coase first raised the question of why we have firms at all in a modern market economy.
- If, as economists usually suggest, markets are so good at allocating resources, why do we need firms?
- Coase recognized that the other side of the question also has to be answered: firms cannot always be better at allocating resources; otherwise we wouldn't see markets.
- In D. H. Robertson's words, we find "islands of conscious power in this ocean of unconscious cooperation like lumps of butter coagulating in a pail of buttermilk."
- Breathtakingly simple and original questions. No precedent in Adam Smith.

- In February 2007, 41 companies in the world with market value of equity greater than \$100 billion
- Walmart, the largest U.S. employer, has 1.8 million employees (2007 figure)

Country	Employee-weighted average size of firms	
	Year	
	<u>1988</u>	<u>2001</u>
France	811	727
Germany	769	725
Italy	474	296
Spain	306	328
United Kingdom	859	935

- 2/3 of the growth in industries over the 1980s came from growth in size of existing firms (sample of 43 countries; Rajan-Zingales (1998))
- Of course, markets important too. Markets and firms coexist, boundaries keep changing. Worldwide value of M and A > \$4 trillion in 2006

# Coase's Answers

- Why are markets sometimes costlier than firms and sometimes less costly?
- Coase argued that the two most obvious costs of using the market/price mechanism are: (a) discovering what the relevant prices are, (b) negotiating a contract for each exchange transaction.
- Economists since Coase have referred to these as “haggling” costs (although I don’t believe that Coase uses this term). “Argument” costs might also be appropriate for (b).
- According to Coase, haggling costs are avoided inside the firm because bargaining is replaced by authority: an employer tells an employee what to do and (within limits) the employee obeys.
- What is the cost of using the firm? According to Coase, managers have limited capacity and so the manager of a large firm will make mistakes.

- Coase's questions are brilliant, but his answers are less so. First, it has been very difficult to formalize or operationalize haggling costs. Second, Coase's cost of using the firm is unconvincing. Why can't the overstretched manager do less, or hire another manager? Third, it seems optimistic to assume that no haggling or argument take place inside a firm.
- Coase has made life hard for his followers by never attempting to write down a formal model . . . .
- Why has it been difficult to operationalize haggling costs? The reason for this can be traced to a paper published in 1960 by an economist just as famous as Coase called . . . .

- Ronald Coase. Coase (1960), writing on the entirely different topic of externalities, made the following observation. Suppose that you can provide a good (or service) that is worth more to me than it costs you – say it's worth 20 to me and costs you 10. Then it would be silly for us not to trade the good at some price between 10 and 20 since we are both made better off. We will probably argue about what the price should be – I would like  $p = 10$  and you would like  $p = 20$ . But presumably we'll settle on something in between; e.g., we might split the difference at  $p = 15$ .

- Now, if we realize that this is what is going to happen, why don't we agree on  $p = 15$  right away and avoid the haggling costs? This argument, that rational parties will avoid, or bargain around, haggling costs "in the twinkling of an eye" has become known as the Coase theorem. (It requires some assumptions . . .)

- How has the literature dealt with this? By introducing a new feature: the idea that the buyer's value (here 20) and the seller's cost (here 10) depend on prior actions/investments undertaken by each party. Under the “split the difference” rule that we've assumed, it can be shown that these investments will be distorted. The literature has studied how allocating asset ownership can mitigate this distortion.



- This approach has yielded some useful insights, but it is far from what Coase had in mind when he talked about haggling costs. More seriously, an approach in which the Coase theorem holds, i.e., people can bargain to an efficient outcome in the twinkling of an eye, is unlikely to be useful for studying such important issues as the internal organization of large firms: in such circumstances why would authority, hierarchy, delegation matter?

- In summary, to make progress, I think that we must move away from Coase (1960) and back in the direction of Coase (1937). We need to bring back haggling costs!

## Part 2: Putting haggling costs back into the picture

- A recent paper with John Moore, “Contracts as Reference Points”, Quarterly Journal of Economics, Feb, 2008, tries to put haggling costs back into the picture. I will begin by describing it and then apply it to the theory of the firm.

- The best way to introduce the paper is to go back to the example where you can provide a good that costs you 10 and is worth 20 to me. To fix ideas imagine that we are talking about a musical evening that I am arranging at my house, and that I want you to sing. The musical evening is worth 20 to me and your effort cost of performing is 10.

- For the moment ignore the fact that I could engage other singers or that you could perform elsewhere on the night in question. Earlier I argued that we might agree to trade at a price of 15. That discussion implicitly assumed that, once we agreed, trade would proceed smoothly. But suppose that is not so. In particular, assume that each of us has some discretion about the quality of “performance” we provide, i.e., how pleasant we make the experience for the other party.

- The seller can perform within the letter rather than the spirit of the contract, or can stint on quality: In the music example you can be rude to my guests or refuse to give autographs. The buyer can quibble about the details of performance or be slow in paying.

- To use the language of Hart-Moore (2008), each party has the discretion to provide “perfunctory” or “consummate” performance. It is worth emphasizing that this is a significant departure from the literature. The literature usually assumes that trade is perfectly enforceable ex post by a court of law. Here we are assuming that only perfunctory performance can be enforced; consummate performance is always discretionary.

- What determines whether a party provides consummate performance? We appeal (quite loosely) to a number of ideas from the recent behavioral economics literature. We assume that a party is roughly indifferent between providing perfunctory and consummate performance – consummate performance costs only slightly more or may even be slightly pleasurable – and will provide consummate performance if he (or she) is “well treated” but not if he is “badly treated” (negative reciprocity).



- Importantly we suppose that a party feels “well treated” if he gets what he believes that he is entitled to; that a contract is a reference point for entitlements; and that in the absence of a reference point entitlements can diverge wildly.

- Let's apply these ideas to our 20-10 example. First, let's put in a time line.



The time line captures the idea that we will typically write a contract some months before the musical evening takes place (date 0) rather than the night before (date 1). One reason for this is that we each have more alternatives earlier on. In fact, I'm going to assume that there is a competitive market for singers at date 0.

- Let's suppose first that, although we sign a contract at date 0, we leave the determination of how much I will pay you until the night before the concert, date 1. This may seem odd and indeed I will show that it is a bad idea. If we don't specify price, then  $p$  can be anywhere between 10 and 20. What might each party feel entitled to?

- As mentioned, we take the view that entitlements can diverge wildly. You can convince yourself that you are hugely talented and that your presence is the entire reason the evening will be a success. You are entitled to  $p = 20$ . I have a dimmer view of your ability and contribution and think that you are worth much less:  $p$  should be 10. Even though we have these different views of what  $p$  should be, we are rational enough to come to some agreement; let's say we split the difference at  $p = 15$ . However, each of us feels shortchanged and aggrieved. I think that I have paid 5 too much; you think that you have been paid 5 too little. Neither of us is in the mood to provide consummate performance.

- The precise assumption that Hart-Moore (2008) makes is that each party “shades” on consummate performance in proportion to the amount he feels aggrieved. Since I’m aggrieved by 5, I shade to the point where your payoff falls by  $5\theta$ , where  $\theta$  is the constant of proportionality: it might be .2, say. And since you’re aggrieved by 5, you shade to the point where my payoff falls by  $5\theta$ .

- The bottom line is that if we leave the price open until the night before the concert, there will be a total deadweight loss of  $10\theta$  due to shading. This is money down the drain. It reduces the value of our relationship from 10 to  $10(1-\theta)$ . If  $\theta = .2$ , our relationship is worth 8 instead of 10.

- Economists don't like deadweight losses (and nor does anyone else). Can anything be done to avoid them here? The answer is yes. Given the assumptions of our model, there is a simple solution: put the price in the contract at date 0. Since I have supposed that there is a competitive market for singers at date 0, I will be able to hire you for  $p = 10$ . With  $p = 10$  specified in the contract, there is nothing for us to argue about at date 1. The fact that we may disagree about your talents as a singer does not matter any more. We have agreed that I will pay you 10, and neither you nor I will be disappointed or aggrieved when that happens.

- In short the role of the contract is to get us on the same page in terms of entitlements.
- Parenthetical note to devotees of the Coase theorem: You cannot bargain around the deadweight losses of shading at date 1 given that shading is noncontractible.



- So far so good, but we need one more step to get to the theory of the firm and Coase (1937). Let's now introduce the realistic notion that not all the details of the musical evening can be anticipated at date 0. To make it simple, imagine that the musical evening can be carried out in two ways, i.e., according to two methods (we might be talking about the exact songs, who are the other performers, the order of the program, etc.)

Method 1 yields value 20 and costs 10, as above. Method 2 yields value 14 to the buyer and costs the seller 8. Assume that the methods cannot be specified in the date 0 contract. However, the choice between them becomes clear at date 1. Note that with these numbers Method 1 is more efficient than Method 2 since  $10 > 6$ .

	<u>Method 1</u>	<u>Method 2</u>
Value	20	14
Cost	10	8

- Compare two different organizational forms. In the first we fix the price of the good at date 0 and make you an independent contractor. In other words, this is a market exchange in Coase's sense. I'm going to take this to mean that you get to decide on the choice between Method 1 and Method 2 at date 1. In other words, you have residual control rights in the sense of Grossman-Hart (1986).

- What will you do? Given that we have fixed the price, you will pick Method 2 since it is cheaper. This is inefficient. How will I feel about this? I will be aggrieved that you didn't choose Method 1 – I will feel entitled to this and will regard your choice as ungenerous – and will shade to the point where your payoff falls by  $6\theta$ . Total surplus =  $6(1 - \theta)$ .

- Now consider a second organizational form. We agree at date 0 that you are an employee: you will work for me at a fixed wage. I'm going to take this to mean that I get to decide on the choice between Method 1 and Method 2 (I have residual control rights) – and indeed this accords with common usage of the term “employment.” Given the fixed wage, I will, of course, choose Method 1 since it gives me more value. This is efficient. You will be aggrieved that I did not choose Method 2, but your aggrievement is only 2. Total surplus =  $10 - 2\theta$ .

- The conclusion is that in this example employment is the best arrangement. Employment is good because the production method matters more to me than to you and so it is efficient that I choose it. You will be aggrieved but not by much because you don't care that much.

- Note that bargaining about the method would be a bad idea . . .
- Now change the numbers. Keep Method 1 the same but suppose Method 2 yields value 14 and costs 2.

	<u>Method 1</u>	<u>Method 2</u>
Value	20	14
Cost	10	2

Method 2 is now more efficient. Under employment, however, the buyer will choose Method 1, yielding surplus  $10 - 8\theta$ . Independent contracting is superior here because the seller will select Method 2, yielding surplus  $12 - 6\theta$ .

- The conclusion is that, if the production method matters more to you than to me, then independent contracting is good: it is efficient that you choose the method. I will be aggrieved but not by much.



- Parenthetical note. I have implicitly assumed that, under independent contracting, the seller can choose the method without violating the original contract. But if the original contract is tight this may not be true. Switching methods may correspond to a “change order,” in which case both parties may have to approve. This raises interesting new possibilities.

- These are toy examples. However, they contain the ingredients of a theory of the choice between doing a transaction “in the market place” and “inside the firm.” The theory is in the spirit of Coase (1937) but perhaps is a bit more satisfactory in some respects. I’ve replaced haggling costs by aggrievement costs, but these are not so different since both have to do with not getting your way. I have emphasized that who controls or decides the production method is a key issue in choosing between the two organizational forms. I have also emphasized that aggrievement costs arise both inside the firm and in the market place.

- I have not had to suppose, as Coase did, that managers of large firms make mistakes, in order to explain why firms don't grow indefinitely.

- Can the simple theory I've presented throw light on the choice of organizational form in practice? For example, can it tell us when outsourcing of government services is a good idea? In applying the theory one should recognize that some important features have been left out. I have supposed that the cost of production (10 or 8 or 2) is always borne by the seller. In practice, if S becomes an employee, many of the production costs will be transferred to the buyer (and it is not hard to see why, given that B has residual control rights). This means that S's incentive to reduce costs will be lower under employment, something that is missing from our model.

- Even with this qualification, our model suggests that outsourcing is likely to be efficient when a detailed contract can be written about the nature of the good to be delivered; in which case B's value will be pretty insensitive to the choice of production method. In contrast, if a detailed contract is hard to write and B's value is very sensitive to the details of production, then in-house production may be better. Municipal garbage collection probably falls into the first category, and fighting wars into the second. The provision of prison services may be somewhere in between (see Hart, Shleifer, Vishny (1997)).







