7 Strategic Behavior

Competitive theory studies price-taking consumers and firms, that is, people who can't individually affect the transaction prices. The assumption that market participants take prices as given is justified only when there are many competing participants. We have also examined monopoly, precisely because a monopoly by definition doesn't have to worry about competitors. Strategic behavior involves the examination of the intermediate case, where there are few enough participants that they take each other into account and their actions individually matter, and where the behavior of any one participant influences choices of the other participants. That is, participants are *strategic* in their choice of action, recognizing that their choice will affect choices made by others.

The right tool for the job of examining strategic behavior in economic circumstances is *game theory*, the study of how people play games. Game theory was pioneered by the mathematical genius John von Neumann (1903-1957). Game theory has also been very influential in the study of military strategy, and indeed the strategy of the cold war between the United States and the U.S.S.R. was guided by game theoretic analyses.[84]

7.1 Games

The theory of games provides a description of games that fits common games like poker or the board game "Monopoly" but will cover many other situations as well. In any game, there is a list of players. Games generally unfold over time; at each moment in time, players have information, possibly incomplete, about the current state of play, and a set of actions they can take. Both information and actions may depend on the history of the game prior to that moment. Finally, players have payoffs, and are assumed to play in such a way as to maximize their expected payoff, taking into account their expectations for the play of others. When the players, their information and available actions, and payoffs have been specified, we have a game.

7.1.1 Matrix Games

The simplest game is called a matrix payoff game with two players. In a matrix payoff game, all actions are chosen simultaneously. It is conventional to describe a matrix payoff game as played by a row player and a column player. The row player chooses a row in a matrix; the column player simultaneously chooses a column. The outcome of the game is a pair of payoffs where the first entry is the payoff of the row player and the second is the payoff of the column player. Table 7-1 provides an example of a "2 × 2" matrix payoff game, the most famous game of all, which is known as the *prisoner's dilemma*.

[84] An important reference for game theory is John von Neumann (1903-1957) and Oskar Morgenstern (1902-1977), *Theory of Games and Economic Behavior*, Princeton: Princeton University Press, 1944. Important extensions were introduced by John Nash (1928 –), the mathematician made famous by Sylvia Nasar's delightful book *A Beautiful Mind* (Simon & Schuster, 1998). Finally, applications in the military arena were pioneered by Nobel Laureate Thomas Schelling (1921 –), *The Strategy of Conflict*, Cambridge: Cambridge University Press, 1960.

Table 7-1: The Prisoner's Dilemma

Row	Column: Confess	Column: Don't
Confess	(-10,-10)	(0,-20)
Don't	(-20,0)	(-1,-1)

In the prisoner's dilemma, two criminals named Row and Column have been apprehended by the police and are being questioned separately. They are jointly guilty of the crime. Each player can choose either to confess or not. If Row confesses, we are in the top row of the matrix (corresponding to the row labeled Confess). Similarly, if Column confesses, the payoff will be in the relevant column. In this case, if only one player confesses, that player goes free and the other serves twenty years in jail. (The entries correspond to the number of years lost to prison. The first entry is always Row's payoff, the second Column's payoff.) Thus, for example, if Column confesses and Row does not, the relevant payoff is the first column and the second row, in reverse color in Table 7-2.

Table 7-2: Solving the Prisoner's Dilemma

Row	Column: Confess	Column: Don't
Confess	(-10,-10)	(0,-20)
Don't	(-20,0)	(-1,-1)

If Column confesses and Row does not, Row loses twenty years, and Column loses no years, that is, goes free. This is the payoff (-20,0) in reverse color in Table 7-2. If both confess, they are both convicted and neither goes free, but they only serve ten years each. Finally, if neither confesses, there is a ten percent chance they are convicted anyway (using evidence other than the confession), in which case they average a year lost each.

The prisoner's dilemma is famous partly because it is readily solvable. First, Row has a strict advantage to confessing, no matter what Column is going to do. If Column confesses, Row gets -10 from confessing, -20 from not, and thus is better off from confessing. Similarly, if Column doesn't confess, Row gets 0 from confessing, -1 from not confessing, and is better off confessing. Either way, no matter what Column does, Row should choose to confess.[85] This is called a *dominant strategy*, a strategy that is optimal no matter what the other players do.

The logic is exactly similar for Column: no matter what Row does, Column should choose to confess. That is, Column also has a dominant strategy, to confess. To establish this, first consider what Column's best action is, when Column thinks Row will confess. Then consider Column's best action when Column thinks Row won't confess.

[85] If Row and Column are friends are care about each other, that should be included as part of the payoffs. Here, there is no honor or friendship among thieves, and Row and Column only care about what they themselves will get.

Either way, Column gets a higher payoff (lower number of years lost to prison) by confessing.

The presence of a dominant strategy makes the prisoner's dilemma particularly easy to solve. Both players should confess. Note that this gets them ten years each in prison, and thus isn't a very good outcome from their perspective, but there is nothing they can do about it in the context of the game, because for each, the alternative to serving ten years is to serve twenty years. This outcome is referred to as (Confess, Confess), where the first entry is the row player's choice, and the second entry is the column player's choice.

Consider an entry game, played by Microsoft (the row player) and Piuny (the column player), a small start-up company. Both Microsoft and Piuny are considering entering a new market for an online service. The payoff structure is

Table 7-3: An Entry Game

		Piuny	
		Enter	Don't
MS	Enter	(2,-2)	(5,0)
	Don't	(0,5)	(0,0)

In this case, if both companies enter, Microsoft ultimately wins the market, and earns 2, and Piuny loses 2. If either firm has the market to itself, they get 5 and the other firm gets zero. If neither enters, both get zero. Microsoft has a dominant strategy to enter: it gets 2 when Piuny enters, 5 when Piuny doesn't, and in both cases does better than when Microsoft doesn't enter. In contrast, Piuny does not have a dominant strategy: Piuny wants to enter when Microsoft doesn't, and vice-versa. That is, Piuny's optimal strategy depends on Microsoft's action, or, more accurately, Piuny's optimal strategy depends on what Piuny believes Microsoft will do.

Piuny can understand Microsoft's dominant strategy, if it knows the payoffs of Microsoft.[86] Thus, Piuny can conclude that Microsoft is going to enter, and this means that Piuny should not enter. Thus, the *equilibrium* of the game is for MS to enter and Piuny not to enter. This equilibrium is arrived at by the *iterated elimination of dominated strategies*, which sounds like jargon but is actually plain speaking. First, we eliminated Microsoft's *dominated strategy* in favor of its dominant strategy. Microsoft had a dominant strategy to enter, which means the strategy of not entering is dominated by the strategy of entering, so we eliminated the dominated strategy. That leaves a simplified game in which Microsoft enters:

[86] It isn't so obvious that one player will know the payoffs of another player, and that often causes players to try to signal that they are going to play a certain way, that is, to demonstrate commitment to a particular advantageous strategy. Such topics are taken up in business strategy and managerial economics.

Table 7-4; Eliminating a Dominated Strategy

		Piuny	
		Enter	Don't
MS	Enter	(2,-2)	(5,0)

In this simplified game, after the elimination of Microsoft's dominated strategy, Piuny also has a dominant strategy: not to enter. Thus, we *iterate* and eliminate dominated strategies again, this time eliminating Piuny's dominated strategies, and wind up with a single outcome: Microsoft enters, and Piuny doesn't. The *iterated elimination of dominated strategies* solves the game.[87]

Here is another game, with three strategies for each player.

Table 7-5: A 3 X 3 Game

		Column		
		Left	Center	Right
Row	Top	(-5,-1)	(2,2)	(3,3)
	Middle	(1,-3)	(1,2)	(1,1)
	Bottom	(0,10)	(0,0)	(0,-10)

The process of iterated elimination of dominated strategies is illustrated by actually eliminating the rows and columns, as follows. A reverse color (white writing on black background) indicates a dominated strategy.

Middle dominates bottom for Row, yielding:

Table 7-6: Eliminating a Dominated Strategy

		Column		
		Left	Center	Right
Row	Top	(-5,-1)	(2,2)	(3,3)
	Middle	(1,-3)	(1,2)	(1,1)
	Bottom	(0,10)	(0,0)	(0,-10)

With bottom eliminated, Left is now dominated for Column by either Center or Right, which eliminates the left column.

[87] A strategy may be dominated not by any particular alternate strategy but by a randomization over other strategies, which is an advanced topic not considered here.

Table 7-7: Eliminating Another Dominated Strategy

		Column		
		Left	Center	Right
Row	Top	(-5,-1)	(2,2)	(3,3)
	Middle	(1,-3)	(1,2)	(1,1)
	Bottom	(0,10)	(0,0)	(0,-10)

With Left and Bottom eliminated, Top now dominates Middle for Row.

Table 7-8: Eliminating a Third Dominated Strategy

		Column		
		Left	Center	Right
Row	Top	(-5,-1)	(2,2)	(3,3)
	Middle	(1,-3)	(1,2)	(1,1)
	Bottom	(0,10)	(0,0)	(0,-10)

Finally, Column chooses Right over Center, yielding a unique outcome after the iterated elimination of dominated strategies, which is (Top, Right).

Table 7-9: Game Solved

		Column		
		Left	Center	Right
Row	Top	(-5,-1)	(2,2)	(3,3)
	Middle	(1,-3)	(1,2)	(1,1)
	Bottom	(0,10)	(0,0)	(0,-10)

The iterated elimination of dominated strategies is a useful concept, and when it applies, the predicted outcome is usually quite reasonable. Certainly it has the property that no player has an incentive to change their behavior given the behavior of others. However, there are games where it doesn't apply, and these games require the machinery of a *Nash equilibrium*, named for Nobel laureate John Nash (1928 –).

7.1.2 Nash Equilibrium

In a *Nash equilibrium*, each player chooses the strategy that maximizes their expected payoff, given the strategies employed by others. For matrix payoff games with two players, a Nash equilibrium requires that the row chosen maximizes the row player's payoff, given the column chosen by the column player, and the column, in turn, maximizes the column player's payoff given the row selected by the row player. Let us consider first the prisoner's dilemma, which we have already seen.

Table 7-10: Prisoner's Dilemma Again

		Column	
		Confess	Don't
Row	Confess	(-10,-10)	(0,-20)
	Don't	(-20,0)	(-1,-1)

Given that the row player has chosen to confess, the column player also chooses confession because -10 is better than -20. Similarly, given that the column player chooses confession, the row player chooses confession, because -10 is better than -20. Thus, for both players to confess is a Nash equilibrium. Now let us consider whether any other outcome is a Nash equilibrium. In any outcome, at least one player is not confessing. But that player could get a higher payoff by confessing, so no other outcome could be a Nash equilibrium.

The logic of dominated strategies extends to Nash equilibrium, except possibly for ties. That is, if a strategy is strictly dominated, it can't be part of a Nash equilibrium. On the other hand, if it involves a tied value, a strategy may be dominated but still part of a Nash equilibrium.

The Nash equilibrium is justified as a solution concept for games as follows. First, if the players are playing a Nash equilibrium, no one has an incentive to change their play or re-think their strategy. Thus, the Nash equilibrium has a "steady state" aspect in that no one wants to change their own strategy given the play of others. Second, other potential outcomes don't have that property: if an outcome is not a Nash equilibrium, then at least one player does have an incentive to change what they are doing. Outcomes that aren't Nash equilibria involve mistakes for at least one player. Thus, sophisticated, intelligent players may be able to deduce each other's play, and play a Nash equilibrium

Do people actually play Nash equilibria? This is a controversial topic and mostly beyond the scope of this book, but we'll consider two well-known games: Tic-Tac-Toe (see, e.g. http://www.mcafee.cc/Bin/tictactoe/index.html) and Chess. Tic-Tac-Toe is a relatively simple game, and the equilibrium is a tie. This equilibrium arises because each player has a strategy that prevents the other player from winning, so the outcome is a tie. Young children play Tic-Tac-Toe and eventually learn how to play equilibrium strategies, at which point the game ceases to be very interesting since it just repeats the same outcome. In contrast, it is known that Chess has an equilibrium, but no one knows what it is. Thus, at this point we don't know if the first mover (White) always wins, or the second mover (Black) always wins, or if the outcome is a draw (neither is able to win). Chess is complicated because a strategy must specify what actions to take given the history of actions, and there are a very large number of potential histories of the game thirty or forty moves after the start. So we can be quite confident that people are not (yet) playing Nash equilibria to the game of Chess.

The second most famous game in game theory is *the battle of the sexes*. The battle of the sexes involves a married couple who are going to meet each other after work, but haven't decided where they are meeting. Their options are a baseball game or the ballet.

Both prefer to be with each other, but the man prefers the baseball game and the woman prefers the ballet. This gives payoffs something like this:

Table 7-11: The Battle of the Sexes

Man \ Woman	Baseball	Ballet
Baseball	(3,2)	(1,1)
Ballet	(0,0)	(2,3)

The man would rather that they both go to the baseball game, and the woman that they both go to the ballet. They each get 2 payoff points for being with each other, and an additional point for being at their preferred entertainment. In this game, iterated elimination of dominated strategies eliminates nothing. You can readily verify that there are two Nash equilibria: one in which they both go to the baseball game, and one in which they both go to ballet. The logic is: if the man is going to the baseball game, the woman prefers the 2 points she gets at the baseball game to the single point she would get at the ballet. Similarly, if the woman is going to the baseball game, the man gets three points going there, versus zero at the ballet. Thus, for both to go to the baseball game is a Nash equilibrium. It is straightforward to show that for both to go to the ballet is also a Nash equilibrium, and finally that neither of the other two possibilities, involving not going to the same place, is an equilibrium.

Now consider the game of *matching pennies*. In this game, both the row player and the column player choose heads or tails, and if they match, the row player gets the coins, while if they don't match, the column player gets the coins. The payoffs are provided in the next table.

Table 7-12: Matching Pennies

Row \ Column	Heads	Tails
Heads	(1,-1)	(-1,1)
Tails	(-1,1)	(1,-1)

You can readily verify that none of the four possibilities represents a Nash equilibrium. Any of the four involves one player getting -1; that player can convert -1 to 1 by changing his or her strategy. Thus, whatever the hypothesized equilibrium, one player can do strictly better, contradicting the hypothesis of a Nash equilibrium. In this game, as every child who plays it knows, it pays to be unpredictable, and consequently players need to *randomize*. Random strategies are known as mixed strategies, because the players mix across various actions.

7.1.3 Mixed Strategies

Let us consider the matching pennies game again.

Table 7-13: Matching Pennies Again

Row \ Column	Heads	Tails
Heads	(1,-1)	(-1,1)
Tails	(-1,1)	(1,-1)

Suppose that Row believes Column plays Heads with probability p. Then if Row plays Heads, Row gets 1 with probability p and -1 with probability (1-p), for an expected value of $2p - 1$. Similarly, if Row plays Tails, Row gets -1 with probability p (when Column plays Heads), and 1 with probability (1-p), for an expected value of $1 - 2p$. This is summarized in the next table.

Table 7-14: Mixed Strategy in Matching Pennies

Row \ Column	Heads	Tails	
Heads	(1,-1)	(-1,1)	$1p + -1(1-p)=2p-1$
Tails	(-1,1)	(1,-1)	$-1p + 1(1-p)=1-2p$

If $2p - 1 > 1 - 2p$, then Row is better off on average playing Heads than Tails. Similarly, if $2p - 1 < 1 - 2p$, Row is better off playing Tails than Heads. If, on the other hand, $2p - 1 = 1 - 2p$, then Row gets the same payoff no matter what Row does. In this case Row could play Heads, could play Tails, or could flip a coin and randomize Row's play.

A *mixed strategy Nash equilibrium* involves at least one player playing a randomized strategy, and no player being able to increase their expected payoff by playing an alternate strategy. A Nash equilibrium without randomization is called a *pure strategy Nash equilibrium*.

Note that that randomization requires equality of expected payoffs. If a player is supposed to randomize over strategy A or strategy B, then both of these strategies must produce the same expected payoff. Otherwise, the player would prefer one of them, and wouldn't play the other.

Computing a mixed strategy has one element that often appears confusing. Suppose Row is going to randomize. Then Row's payoffs must be equal, for all strategies Row plays with positive probability. But that equality in Row's payoffs doesn't determine the probabilities with which Row plays the various rows. Instead, that equality in Row's payoffs will determine the probabilities with which Column plays the various columns. The reason is that it is Column's probabilities that determine the expected payoff for Row; if Row is going to randomize, then Column's probabilities must be such that Row is willing to randomize.

Thus, for example, we computed the payoff to Row of playing Heads, which was $2p - 1$, where p was the probability Column played Heads. Similarly, the payoff to Row of playing Tails was $1 - 2p$. Row is willing to randomize if these are equal, which solves for $p = ½$.

7.1.3.1 (Exercise) Let q be the probability that Row plays Heads. Show that Column is willing to randomize if, and only if, $q = ½$. (Hint: First compute Column's expected payoff when Column plays Heads, and then Column's expected payoff when Column plays Tails. These must be equal for Column to randomize.)

Now let's try a somewhat more challenging example, and revisit the battle of the sexes.

Table 7-15: Mixed Strategy in Battle of the Sexes

		Woman	
		Baseball	Ballet
Man	Baseball	(3,2)	(1,1)
	Ballet	(0,0)	(2,3)

This game has two pure strategy Nash equilibria: (Baseball,Baseball) and (Ballet,Ballet). Is there a mixed strategy? To compute a mixed strategy, let the Woman go to the baseball game with probability p, and the Man go to the baseball game with probability q. Table 7-16 contains the computation of the mixed strategy payoffs for each player.

Table 7-16: Full Computation of the Mixed Strategy

		Woman		
		Baseball (p)	Ballet (1-p)	Man's E Payoff
Man	Baseball (prob q)	(3,2)	(1,1)	$3p + 1(1-p)=1+2p$
	Ballet (prob 1-q)	(0,0)	(2,3)	$0p + 2(1-p)=2-2p$
	Woman's E Payoff	$2q + 0(1-q)=2q$	$1q + 3(1-q)=3-2q$	

For example, if the Man (row player) goes to the baseball game, he gets 3 when the Woman goes to the baseball game (probability p) and otherwise gets 1, for an expected payoff of $3p + 1(1-p) = 1 + 2p$. The other calculations are similar but you should definitely run through the logic and verify each calculation.

A mixed strategy in the Battle of the Sexes game requires both parties to randomize (since a pure strategy by either party prevents randomization by the other). The Man's indifference between going to the baseball game and the ballet requires $1+2p = 2 – 2p$, which yields $p = ¼$. That is, the Man will be willing to randomize which event he attends if the Woman is going to the ballet ¾ of the time, and otherwise to the baseball game. This makes the Man indifferent between the two events, because he prefers to be with the Woman, but he also likes to be at the baseball game; to make up for the advantage that the game holds for him, the woman has to be at the ballet more often.

Similarly, in order for the Woman to randomize, the Woman must get equal payoffs from going to the game and going to the ballet, which requires $2q = 3 – 2q$, or $q = ¾$. Thus, the probability that the Man goes to the game is ¾, and he goes to the ballet ¼ of the time. These are independent probabilities, so to get the probability that both go to the game, we multiply the probabilities, which yields $3/16$. The next table fills in the probabilities for all four possible outcomes.

Table 7-17: Mixed Strategy Probabilities

Man \ Woman	Baseball	Ballet
Baseball	$3/16$	$9/16$
Ballet	$1/16$	$3/16$

Note that more than half the time, (Baseball, Ballet) is the outcome of the mixed strategy, and the two people are not together. This lack of coordination is a feature of mixed strategy equilibria generally. The expected payoffs for both players are readily computed as well. The Man's payoff was $1+2p = 2 - 2p$, and since $p = ¼$, the Man obtained 1 ½. A similar calculation shows the Woman's payoff is the same. Thus, both do worse than coordinating on their less preferred outcome. But this mixed strategy Nash equilibrium, undesirable as it may seem, is a Nash equilibrium in the sense that neither party can improve their payoff, given the behavior of the other party.

In the Battle of the sexes, the mixed strategy Nash equilibrium may seem unlikely, and we might expect the couple to coordinate more effectively. Indeed, a simple call on the telephone should rule out the mixed strategy. So let's consider another game related to the Battle of the Sexes, where a failure of coordination makes more sense. This is the game of "Chicken." Chicken is played by two drivers driving toward each other, trying to convince the other to yield, which involves swerving into a ditch. If both swerve into the ditch, we'll call the outcome a draw and both get zero. If one swerves and the other doesn't, the swerver loses and the other wins, and we'll give the winner one point.[88] The only remaining question is what happens when both don't yield, in which case a crash results. In this version, that has been set at four times the loss of swerving, but you can change the game and see what happens.

Table 7-18: Chicken

Row \ Column	Swerve	Don't
Swerve	(0,0)	(-1,1)
Don't	(1,-1)	(-4,-4)

This game has two pure strategy equilibria: (Swerve, Don't) and (Don't, Swerve). In addition, it has a mixed strategy. Suppose Column swerves with probability p. Then

[88] Note that adding a constant to a player's payoffs, or multiplying that player's payoffs by a positive constant, doesn't affect the Nash equilibria, pure or mixed. Therefore, we can always let one outcome for each player be zero, and another outcome be one.

Row gets $0p$ + $-1(1-p)$ from swerving, $1p + (-4)(1-p)$ from not swerving, and Row will randomize if these are equal, which requires $p = \frac{3}{4}$. That is, the probability that Column swerves, in a mixed strategy equilibrium is $\frac{3}{4}$. You can verify that the Row player has the same probability by setting the probability that Row swerves equal to q and computing Column's expected payoffs. Thus, the probability of a collision is $\frac{1}{16}$ in the mixed strategy equilibrium.

The mixed strategy equilibrium is more likely in some sense in this game; if the players already knew which player would yield, they wouldn't actually need to play the game. The whole point of the game is to find out who will yield, which means it isn't known in advance, which means the mixed strategy equilibrium is in some sense the more reasonable equilibrium.

Paper, Scissors, Rock is a child's game in which two children simultaneously choose paper (hand held flat), scissors (hand with two fingers protruding to look like scissors) or rock (hand in a fist). The nature of the payoffs is that paper beats rock, rock beats scissors, and scissors beat paper. This game has the structure

Table 7-19: Paper, Scissors, Rock

		Column		
		Paper	Scissors	Rock
Row	Paper	(0,0)	(-1,1)	(1,-1)
	Scissors	(1,-1)	(0,0)	(-1,1)
	Rock	(-1,1)	(1,-1)	(0,0)

7.1.3.2 (Exercise) Show that, in the Paper, Scissors, Rock game, there are no pure strategy equilibria. Show that playing all three actions with equal likelihood is a mixed strategy equilibrium.

7.1.3.3 (Exercise) Find all equilibria of the following games:

1

		Column	
		Left	Right
Row	Up	(3,2)	(11,1)
	Down	(4,5)	(8,0)

2

		Column	
		Left	Right
Row	Up	(3,3)	(0,0)
	Down	(4,5)	(8,0)

3

Row		Column Left	Right
	Up	(0,3)	(3,0)
	Down	(4,0)	(0,4)

4

Row		Column Left	Right
	Up	(7,2)	(0,9)
	Down	(8,7)	(8,8)

5

Row		Column Left	Right
	Up	(1,1)	(2,4)
	Down	(4,1)	(3,2)

6

Row		Column Left	Right
	Up	(4,2)	(2,3)
	Down	(3,8)	(1,5)

7.1.4 Examples

Our first example concerns public goods. In this game, each player can either contribute, or not. For example, two roommates can either clean their apartment, or not. If they both clean, the apartment is nice. If one cleans, that roommate does all the work and the other gets half of the benefits. Finally, if neither clean, neither is very happy. This suggests payoffs like:

Table 7-20: Cleaning the Apartment

Row		Column Clean	Don't
	Clean	(10,10)	(0,15)
	Don't	(15,0)	(2,2)

You can verify that this game is similar to the prisoner's dilemma, in that the only Nash equilibrium is the pure strategy in which neither player cleans. This is a game theoretic version of the tragedy of the commons – even though the roommates would both be better off if both cleaned, neither do. As a practical matter, roommates do solve this problem, using strategies that we will investigate when we consider dynamic games.

Table 7-21: Driving on the Right

Row \ Column	Left	Right
Left	(1,1)	(0,0)
Right	(0,0)	(1,1)

The important thing about the side of the road the cars drive on is not that it is the right side but that it is the *same* side. This is captured in the Driving on the Right game above. If both players drive on the same side, then they both get one point, otherwise they get zero. You can readily verify that there are two pure strategy equilibria, (Left,Left) and (Right,Right), and a mixed strategy equilibrium with equal probabilities. Is the mixed strategy reasonable? With automobiles, there is little randomization. On the other hand, people walking down hallways often seem to randomize whether they pass on the left or the right, and sometimes do that little dance where they try to get past each other, one going left and the other going right, then both simultaneously reversing, unable to get out of each other's way. That dance suggests that the mixed strategy equilibrium is not as unreasonable as it seems in the automobile application.[89]

Table 7-22: Bank Location Game

LA \ NYC	No Concession	Tax Rebate
No Concession	(30,10)	(10,20)
Tax Rebate	(20,10)	(20,0)

Consider a foreign bank that is looking to open a main office and a smaller office in the United States. The bank narrows its choice for main office to either New York (NYC) or Los Angeles (LA), and is leaning toward Los Angeles. If neither city does anything, LA will get $30 million in tax revenue and New York ten million. New York, however, could offer a $10 million rebate, which would swing the main office to New York, but now New York would only get a net of $20 M. The discussions are carried on privately with the bank. LA could also offer the concession, which would bring the bank back to LA.

[89] Continental Europe drove on the left until about the time of the French revolution. At that time, some individuals began driving on the right as a challenge to royalty who were on the left, essentially playing the game of chicken with the ruling class. Driving on the right became a symbol of disrespect for royalty. The challengers won out, forcing a shift to driving on the right. Besides which side one drives on, another coordination game involves whether one stops or goes on red. In some locales, the tendency for a few extra cars to go as a light changes from green to yellow to red forces those whose light changes to green to wait, and such a progression can lead to the opposite equilibrium, where one goes on red and stops on green. Under Mao Tse-tung, the Chinese considered changing the equilibrium to going on red and stopping on green (because 'red is the color of progress') but wiser heads prevailed and the plan was scrapped.

7.1.4.1 (Exercise) Verify that the bank location game has no pure strategy equilibria, and that there is a mixed strategy equilibrium where each city offers a rebate with probability ½.

Table 7-23: Political Mudslinging

		Republican	
		Clean	Mud
Dem	Clean	(3,1)	(1,2)
	Mud	(2,1)	(2,0)

On the night before the election, a Democrat is leading the Wisconsin senatorial race. Absent any new developments, the Democrat will win, and the Republican will lose. This is worth 3 to the Democrat, and the Republican, who loses honorably, values this outcome at one. The Republican could decide to run a series of negative advertisements ("throwing mud") against the Democrat, and if so, the Republican wins although loses his honor, which he values at 1, and so only gets 2. If the Democrat runs negative ads, again the Democrat wins, but loses his honor, so only gets 2. These outcomes are represented in the Mudslinging game above.

7.1.4.2 (Exercise) Show that the only Nash equilibrium is a mixed strategy with equal probabilities of throwing mud and not throwing mud.

7.1.4.3 (Exercise) Suppose that voters partially forgive a candidate for throwing mud when the rival throws mud, so that the (Mud, Mud) outcome has payoff (2.5,.5). How does the equilibrium change?

You have probably had the experience of trying to avoid encountering someone, who we will call Rocky. In this instance, Rocky is actually trying to find you. The situation is that it is Saturday night and you are choosing which party, of two possible parties, to attend. You like party 1 better, and if Rocky goes to the other party, you get 20. If Rocky attends party 1, you are going to be uncomfortable and get 5. Similarly, Party 2 is worth 15, unless Rocky attends, in which case it is worth 0. Rocky likes Party 2 better (these different preferences may be part of the reason you are avoiding him) but he is trying to see you. So he values Party 2 at 10, party 1 at 5 and your presence at the party he attends is worth 10. These values are reflected in the following table.

Table 7-24: Avoiding Rocky

		Rocky	
		Party 1	Party 2
You	Party 1	(5,15)	(20,10)
	Party 2	(15,5)	(0,20)

7.1.4.4 (Exercise) (i) Show there are no pure strategy Nash equilibria in this game. (ii) Find the mixed strategy Nash equilibria. (iii) Show that the probability you encounter Rocky is $\frac{7}{12}$.

Our final example involves two firms competing for customers. These firms can either price high or low. The most money is made if they both price high, but if one prices low, it can take most of the business away from the rival. If they both price low, they make modest profits. This description is reflected in the following table:

Table 7-25: Price Cutting Game

Firm 1 \ Firm 2	High	Low
High	(15,15)	(0,25)
Low	(25,0)	(5,5)

7.1.4.5 (Exercise) Show that the firms have a dominant strategy to price low, so that the only Nash equilibrium is (Low, Low).

7.1.5 Two Period Games

So far, we have considered only games that are played simultaneously. Several of these games, notably the price cutting and apartment cleaning games, are actually played over and over again. Other games, like the bank location game, may only be played once but nevertheless are played over time. Recall the bank location game:

Table 7-26; Bank Location Revisited

LA \ NYC	No Concession	Tax Rebate
No Concession	(30,10)	(10,20)
Tax Rebate	(20,10)	(20,0)

If neither city offered a rebate, then LA won the bidding. So suppose instead of the simultaneous move game, that first New York decided whether to offer a rebate, and then LA could decide to offer a rebate. This sequential structure leads to a game that looks like Figure 7-1:

In this game, NYC makes the first move, and chooses Rebate (to the left) or No Rebate (to the right). If NYC chooses Rebate, LA can then choose Rebate or None. Similarly, if NYC chooses No Rebate, LA can choose Rebate or None. The payoffs (using the standard of (LA, NYC) ordering) are written below the choices.

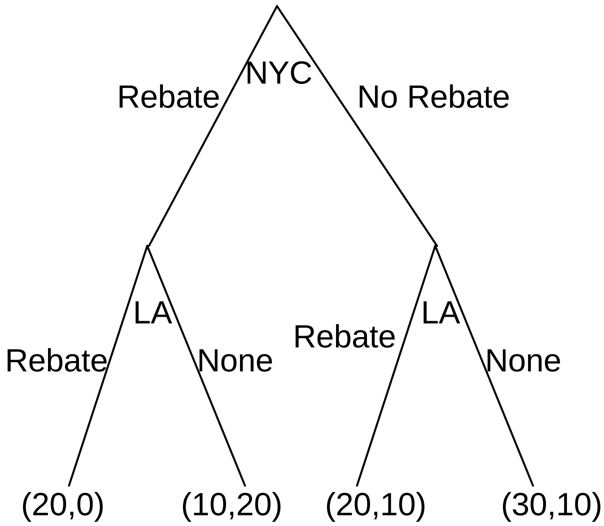

Figure 7-1 Sequential Bank Location (NYC payoff listed first)

What NYC would like to do depends on what NYC believes LA will do. What should NYC believe about LA? (Boy does that rhetorical question suggest a lot of facetious answers.) The natural belief is that LA will do what is in LA's best interest. This idea – that each stage of a game is played in a maximizing way – is called *subgame perfection*.

7.1.6 Subgame Perfection

Subgame perfection requires each player to act in its best interest, independent of the history of the game.[90] This seems very sensible and in most contexts it is sensible. In some settings, it may be implausible. Even if I see a player make a particular mistake three times in a row, subgame perfection requires that I must continue to believe that player will not make the mistake again. Subgame perfection may be implausible in some circumstances, especially when it pays to be considered somewhat crazy.

In the example, subgame perfection requires LA to offer a rebate when NYC does (since LA gets 20 by rebating versus 10), and not when NYC doesn't. This is illustrated in the game using arrows to indicate LA's choices. In addition, the actions that LA won't choose have been re-colored in a light grey in Figure 7-2.

Once LA's subgame perfect choices are taken into account, NYC is presented with the choice of offering a rebate, in which case it gets 0, or not offering a rebate, in which case it gets 10. Clearly the optimal choice for NYC is to offer no rebate, in which case LA doesn't either, and the result is 30 for LA, and 10 for NYC.

Dynamic games are generally "solved backward" in this way. That is, first establish what the last player does, then figure out based on the last player's expected behavior, what the penultimate player does, and so on.

[90] Subgame perfection was introduced by Nobel laureate Reinhart Selten (1930 –).

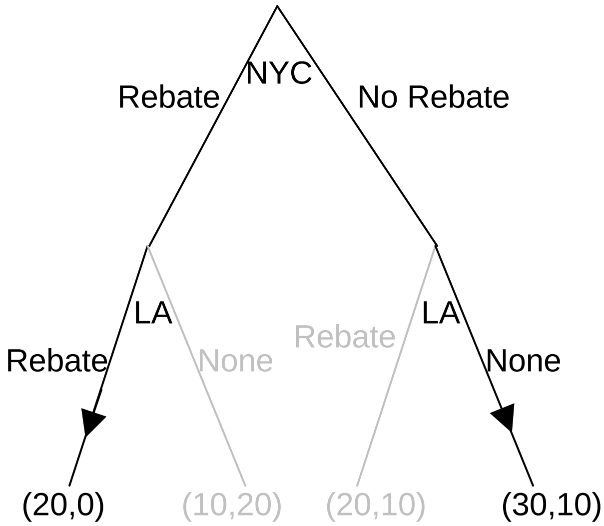

Figure 7-2: Subgame Perfection

We'll consider one more application of subgame perfection. Suppose, in the game "Avoiding Rocky," Rocky is actually stalking you, and can condition his choice on your choice. Then you might as well go to the party you like best, because Rocky is going to follow you wherever you go. This is represented in Figure 7-3.

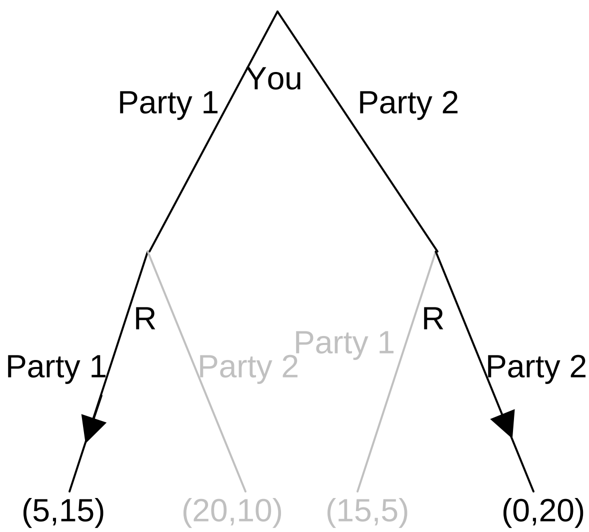

Figure 7-3: Can't Avoid Rocky

Since Rocky's optimal choice eliminates your best outcomes, you make the best of a bad situation by choosing Party 1. Here, Rocky has a *second mover advantage*: Rocky's ability to condition on your choice meant he does better than he would do in a simultaneous game.

7.1.6.1 (Exercise) Formulate the battle of the sexes as a sequential game, letting the woman choose first. (This situation could arise if the woman can leave a message for the man about where she has gone.) Show that there is only one subgame perfect equilibrium, and that the woman enjoys a first-mover advantage over the man, and she gets her most preferred outcome.

7.1.7 Supergames

Some situations, like the pricing game or the apartment-cleaning game, are played over and over. Such situations are best modeled as a *supergame*.[91] A supergame is a game played over and over again without end, where the players discount the future. The game played each time is known as a *stage game*. Generally supergames are played in times 1, 2, 3, …

Cooperation may be possible in supergames, if the future is important enough. Consider the pricing game introduced above.

Table 7-27: Price Cutting, Revisited

		Firm 2	
		High	Low
Firm 1	High	(15,15)	(0,25)
	Low	(25,0)	(5,5)

The dominant strategy equilibrium to this game is (Low, Low). It is clearly a subgame perfect equilibrium for the players to just play (Low, Low) over and over again, because if that is what Firm 1 thinks Firm 2 is doing, Firm 1 does best by pricing Low, and vice versa. But that is not the only equilibrium to the supergame.

Consider the following strategy, called a *grim trigger strategy*. Price high, until you see your rival price low. After your rival has priced low, price low forever. This is called a trigger strategy because an action of the other player (pricing low) triggers a change in behavior. It is a grim strategy because it punishes forever.

If your rival uses a grim trigger strategy, what should you do? Basically, your only choice is when to price low, because once you price low, your rival will price low, and then your best choice is to also price low from then on. Thus, your strategy is to price high up until some point $t-1$, and then price low from time t on. Your rival will price high through t, and price low from $t+1$ on. This gives a payoff to you of 15 from period 1 through $t-1$, 25 in period t, and then 5 in period $t+1$ on. We can compute the payoff for a discount factor δ.

$$V_t = 15(1+\delta+\delta^2+\ldots+\delta^{t-1}) + 25\delta^t + 5(\delta^{t+1}+\delta^{t+2}+\ldots)$$

$$= 15\frac{1-\delta^t}{1-\delta} + 25\delta^t + 5\frac{\delta^t}{1-\delta} = \frac{15}{1-\delta} - \frac{\delta^t}{1-\delta}(15-25(1-\delta)-5\delta) = \frac{15}{1-\delta} - \frac{\delta^t}{1-\delta}(-10+20\delta).$$

[91] The supergame was invented by Robert Aumann (1930 –) in 1959.

If -10 + 20δ < 0, it pays to price low immediately, at t=0, because it pays to price low and the earlier the higher the present value. If -10 + 20δ > 0, it pays to wait forever to price low, that is, $t = \infty$. Thus, in particular, the grim trigger strategy is an optimal strategy for a player when the rival is playing the grim trigger strategy if $\delta \geq ½$. In other words, cooperation in pricing is a subgame perfect equilibrium if the future is important enough, that is, the discount factor δ is high enough.

The logic of this example is that the promise of future cooperation is valuable when the future itself is valuable, and that promise of future cooperation can be used to induce cooperation today. Thus, firm 1 doesn't want to cut price today, because that would lead firm 2 to cut price for the indefinite future. The grim trigger strategy punishes price cutting today with future low profits.

Supergames offer more scope for cooperation than is illustrated in the pricing game. First, more complex behavior is possible. For example, consider the following game:

Table 7-28: A Variation of the Price Cutting Game

		Firm 2	
		High	Low
Firm 1	High	(10,10)	(0,25)
	Low	(25,0)	(5,5)

Here, again, the unique equilibrium in the stage game is (Low, Low). But the difference between this game and the previous game is that the total profits of firms 1 and 2 are higher in either (High, Low) or (Low, High) than in (High, High). One solution is to alternate between (High, Low) and (Low, High). Such alternation can also be supported as an equilibrium, using the grim trigger strategy – that is, if a firm does anything other than what is it supposed to in the alternating solution, the firms instead play (Low, Low) forever.

7.1.7.1 (Exercise) Consider the game in Table 7-28, and consider a strategy in which firm 1 prices high in odd numbered periods, and low in even numbered periods, while 2 prices high in even numbered periods, low in odd numbered periods. If either deviate from these strategies, both price low from then on. Let δ be the discount factor. Show that these firms have a payoff of $\frac{25}{1-\delta^2}$ or $\frac{25\delta}{1-\delta^2}$, depending on which period it is. Then show that the alternating strategy is sustainable if $10+\frac{5\delta}{1-\delta} \leq \frac{25\delta}{1-\delta^2}$. This, in turn, is equivalent to $\delta \geq \sqrt{6}-2$.

7.1.8 The Folk Theorem

The folk theorem says that if the value of the future is high enough, any outcome that is *individually rational* can be supported as an equilibrium to the supergame. Individual rationality for a player in this context means that the outcome offers a present value of profits at least as high as that offered in the worst equilibrium in the stage game from that player's perspective. Thus, in the pricing game, the worst equilibrium of the stage game offered each player 5, so an outcome can be supported if it offers each player at least a running average of 5.

The simple logic of the folk theorem is this. First, any infinite repetition of an equilibrium of the stage game is itself a subgame perfect equilibrium. If everyone expects this repetition of the stage game equilibrium, no one can do better than to play their role in the stage game equilibrium every period. Second, any other plan of action can be turned into a subgame perfect equilibrium merely by threatening any agent who deviates from that plan with an infinite repetition of the worst stage game equilibrium from that agent's perspective. That threat is credible because the repetition of the stage game equilibrium is itself a subgame perfect equilibrium. Given such a grim trigger type threat, no one wants to deviate from the intended plan.

The folk theorem is a powerful result, and shows that there are equilibria to supergames that achieve very good outcomes. The kinds of coordination failures we saw in the battle of the sexes, and the failure to cooperate in the prisoner's dilemma, need not arise, and cooperative solutions are possible if the future is sufficiently valuable.

However, it is worth noting some assumptions that have been made in our descriptions of these games, assumptions that matter and are unlikely to be true in practice. First, the players know their own payoffs. Second, they know their rival's payoffs. They possess a complete description of the available strategies and can calculate the consequences of these strategies, not just for themselves, but for their rivals. Third, each player maximizes his or her expected payoff, and they know that their rivals do the same, and they know that their rivals know that everyone maximizes, and so on. The economic language for this is the structure of the game and the player's preferences are *common knowledge*. Few real world games will satisfy these assumptions exactly. Since the success of the grim trigger strategy (and other strategies we haven't discussed) generally depends on such knowledge, informational considerations may cause cooperation to break down. Finally, the folk theorem shows us that there are lots of equilibria to supergames, and provides no guidance on which one will be played. These assumptions can be relaxed, although they may lead to wars on the equilibrium path "by accident," and a need to recover from such wars, so that the grim trigger strategy becomes sub-optimal.

7.2 Cournot Oligopoly

The Cournot[92] oligopoly model is the most popular model of imperfect competition. It is a model in which the number of firms matters, and represents one way of thinking about what happens when the world is neither perfectly competitive, nor a monopoly.

In the Cournot model, there are n firms, who choose quantities. We denote a typical firm as firm i and number the firms from $i = 1$ to $i = n$. Firm i chooses a quantity $q_i \geq 0$ to sell and this quantity costs $c_i(q_i)$. The sum of the quantities produced is denoted by Q. The price that emerges from the competition among the firms is $p(Q)$ and this is the same price for each firm. It is probably best to think of the quantity as really

[92] Augustus Cournot, 1801-1877.

representing a capacity, and competition in prices by the firms determining a market price given the market capacity.

The profit that a firm i obtains is

$$\pi_i = p(Q)q_i - c_i(q_i).$$

7.2.1 Equilibrium

Each firm chooses q_i to maximize profit. The first order conditions[93] give:

$$0 = \frac{\partial \pi_i}{\partial q_i} = p(Q) + p'(Q)q_i - c_i'(q_i).$$

This equation holds with equality provided $q_i > 0$. A simple thing that can be done with the first order conditions is to rewrite them to obtain the average value of the price-cost margin:

$$\frac{p(Q) - c_i'(q_i)}{p(Q)} = -\frac{p'(Q)q_i}{p(Q)} = -\frac{Qp'(Q)}{p(Q)}\frac{q_i}{Q} = \frac{s_i}{\varepsilon}.$$

Here $s_i = \frac{q_i}{Q}$ is firm i's market share. Multiplying this equation by the market share and summing over all firms $i = 1, \dots, n$ yields

$$\sum_{i=1}^{n} \frac{p(Q) - c_i'(q_i)}{p(Q)} s_i = \frac{1}{\varepsilon}\sum_{i=1}^{n} s_i^2 = \frac{HHI}{\varepsilon},$$

where $HHI = \sum_{i=1}^{n} s_i^2$ is the Hirschman-Herfindahl Index.[94] The HHI has the property that if the firms are identical, so that $s_i = 1/n$ for all i, then the HHI is also $1/n$. For this reason, antitrust economists will sometimes use 1/HHI as a proxy for the number of firms, and describe an industry with "2 ½ firms," meaning an HHI of 0.4.[95]

We can draw several inferences from these equations. First, larger firms, those with larger market shares, have a larger deviation from competitive behavior (price equal to marginal cost). Small firms are approximately competitive (price nearly equals marginal cost) while large firms reduce output to keep the price higher, and the amount of the reduction, in price/cost terms, is proportional to market share. Second, the HHI reflects the deviation from perfect competition on average, that is, it gives the average

[93] Bear in mind that Q is the sum of the firms' quantities, so that when firm i increases its output slightly, Q goes up by the same amount.
[94] Named for Albert Hirschman (1918 – 1972), who invented it in 1945, and Orris Herfindahl (1915 –), who invented it independently in 1950.
[95] To make matters more confusing, antitrust economists tend to state the HHI using shares in percent, so that the HHI is on a 0 to 10,000 scale.

proportion by which price equal to marginal cost is violated. Third, the equation generalizes the "inverse elasticity result" proved for monopoly, which showed that the price – cost margin was the inverse of the elasticity of demand. The generalization states that the weighted average of the price – cost margins is the HHI over the elasticity of demand.

Since the price – cost margin reflects the deviation from competition, the HHI provides a measure of how large a deviation from competition is present in an industry. A large HHI means the industry "looks like monopoly." In contrast, a small HHI looks like perfect competition, holding constant the elasticity of demand.

The case of a symmetric (identical cost functions) industry is especially enlightening. In this case, the equation for the first order condition can be restated as

$$0 = p(Q) + p'(Q) Q/n - c'\left(Q/n\right)$$

or

$$p(Q) = \frac{\varepsilon n}{\varepsilon n - 1} c'\left(Q/n\right).$$

Thus, in the symmetric model, competition leads to pricing as if demand were more elastic, and indeed is a substitute for elasticity as a determinant of price.

7.2.2 Industry Performance

How does the Cournot industry perform? Let us return to the more general model, that doesn't require identical cost functions. We already have one answer to this question: the average price – cost margin is the HHI divided by the elasticity of demand. Thus, if we have an estimate of the demand elasticity, we know how much the price deviates from the perfect competition benchmark.

The general Cournot industry actually has two sources of inefficiency. First, price is above marginal cost, so there is the dead weight loss associated with unexploited gains from trade. Second, there is the inefficiency associated with different marginal costs. This is inefficient because a re-arrangement of production, keeping total output the same, from the firm with high marginal cost to the firm with low marginal cost, would reduce the cost of production. That is, not only is too little output produced, but what output is produced is inefficiently produced, unless the firms are identical.

To assess the productive inefficiency, we let c_1' be the lowest marginal cost. The average deviation from the lowest marginal cost, then, is

$$\chi = \sum_{i=1}^{n} s_i (c_i' - c_1') = \sum_{i=1}^{n} s_i (p - c_1' - (p - c_i')) = p - c_1' - \sum_{i=1}^{n} s_i (p - c_i')$$

$$= p - c_1' - p\sum_{i=1}^{n} s_i \frac{(p - c_i')}{p} = p - c_1' - \frac{p}{\varepsilon}\sum_{i=1}^{n} s_i^2 = p - c_1' - \frac{p}{\varepsilon} HHI .$$

Thus, while a large HHI means a large deviation from price equal to marginal cost and hence a large level of monopoly power (holding constant the elasticity of demand), a large HHI also tends to indicate greater productive efficiency, that is, less output produced by high cost producers. Intuitively, a monopoly produces efficiently, even if it has a greater reduction in total output than other industry structures.

There are a number of caveats worth mentioning in the assessment of industry performance. First, the analysis has held constant the elasticity of demand, which could easily fail to be correct in an application. Second, fixed costs have not been considered. An industry with large economies of scale, relative to demand, must have very few firms to perform efficiently and small numbers should not necessarily indicate the market performs poorly even if price – cost margins are high. Third, it could be that entry determines the number of firms, and that the firms have no long-run market power, just short-run market power. Thus, entry and fixed costs could lead the firms to have approximately zero profits, in spite of price above marginal cost.

7.2.2.1 (Exercise) Suppose the inverse demand curve is $p(Q) = 1 - Q$, and that there are n Cournot firms, each with constant marginal cost c, selling in the market. (i) Show that the Cournot equilibrium quantity and price are $Q = \frac{n(1-c)}{n+1}$ and $p(Q) = \frac{1+nc}{n+1}$. (ii) Show each firm's gross profits are $\left(\frac{1-c}{n+1}\right)^2$.

Continuing with 7.2.2.1 (Exercise), suppose there is a fixed cost F to become a firm. The number of firms n should be such that firms are able to cover their fixed costs, but add one more and they can't. This gives us a condition determining the number of firms n:

$$\left(\frac{1-c}{n+1}\right)^2 \geq F \geq \left(\frac{1-c}{n+2}\right)^2 .$$

Thus, each firm's net profits are $\left(\frac{1-c}{n+1}\right)^2 - F \leq \left(\frac{1-c}{n+1}\right)^2 - \left(\frac{1-c}{n+2}\right)^2 = \frac{(2n+3)(1-c)^2}{(n+1)^2(n+2)^2}$.

Note that the monopoly profits π_m are ¼ $(1\text{-}c)^2$. Thus, with free entry, net profits are less than $\frac{(2n+3)4}{(n+1)^2(n+2)^2}\pi_m$, and industry net profits are less than $\frac{n(2n+3)4}{(n+1)^2(n+2)^2}\pi_m$.

Table 7-29 shows the performance of the constant cost, linear demand Cournot industry, when fixed costs are taken into account, and when they aren't. With two firms, gross industry profits are 8/9ths of the monopoly profits, not substantially different from monopoly. But when fixed costs sufficient to insure that only two firms enter are considered, the industry profits are at most 39% of the monopoly profits. This number – 39% -- is large because fixed costs could be "relatively" low, so that the third firm is just deterred from entering. That still leaves the two firms with significant profits, even though the third firm can't profitably enter. As the number of firms rises, gross industry profits fall slowly toward zero. The net industry profits, on the other hand, fall dramatically rapidly to zero. With ten firms, the gross profits are still about a third of the monopoly level, but the net profits are only at most 5% of the monopoly level.

Table 7-29: Industry Profits as a Fraction of Monopoly Profits

Number of Firms	Gross Industry Profits (%)	Net Industry Profits (%)
2	88.9	39.0
3	75.0	27.0
4	64.0	19.6
5	55.6	14.7
10	33.1	5.3
15	23.4	2.7
20	18.1	1.6

The Cournot model gives a useful model of imperfect competition, a model that readily permits assessing the deviation from perfect competition. The Cournot model embodies two kinds of inefficiency: the exercise of monopoly power, and technical inefficiency in production. In settings involving entry and fixed costs, care must be taken in applying the Cournot model.

7.3 Search and Price Dispersion

Decades ago, economists used to make a big deal about the "Law of One Price," which states that identical goods sell at the same price. The argument in favor of the law of one price is theoretical. Well-informed consumers will buy identical goods from the lowest price seller. Consequently, the only seller to make any sales is the low-price seller. This kind of consumer behavior forces all sellers to sell at the same price.

There are few markets where the law of one price is actually observed to hold. Organized exchanges, like stock, bond and commodity markets, will satisfy the law of one price. In addition, gas stations across the street from each other will often offer identical prices, but often is not always.

Many economists considered that the internet would force prices of standardized goods – DVD players, digital cameras, MP3 players – to a uniform, and uniformly low, price. However, this has not occurred. Moreover, it probably can't occur, in the sense that pure price competition would put the firms out of business, and hence can't represent equilibrium behavior.

There are many markets where prices appear unpredictable to consumers. The price of airline tickets is notorious for unpredictability. The price of milk, soft drinks, paper towels and canned tuna varies 50% or more depending on whether the store has an advertised sale of the item or not. Prices of goods sold on the internet varies substantially from day to day.[96] Such variation is known as *price dispersion* by economists. It is different from price discrimination, in that price dispersion entails a given store quoting the same price to all customers; the variation is across stores, while price discrimination is across customers.

Why are prices so unpredictable?

7.3.1 Simplest Theory

To understand price dispersion, we divide consumers into two types: shoppers and loyal customers. Loyal customers won't pay more than a price p_m for the good, but they only consult a particular store; if that store has the good for less than the price p_m, the loyal customer buys, and otherwise not. In contrast, the shoppers buy only from the store offering the lowest price; shoppers are informed about the prices offered by all stores. We let the proportion of shoppers be s. The loyal customers are allocated to the other stores equally, so that, if there are n stores, each store gets a fraction $(1 - s)/n$ of the customers. Let the marginal cost of the good be c, and assume $c < p_m$. Both kinds of customers buy only one unit.

For the purposes of this analysis, we will assume that prices can be chosen from the continuum. This makes the analysis more straightforward, but there is an alternate version of the analysis (not developed here) that makes the more reasonable assumption of prices that are an integer number of pennies.

First note that there is no pure strategy equilibrium. To see this, consider the lowest price p charged by any firm. If that price is c, the firm makes no money, so would do better by raising its price to p_m and selling only to the loyal customers. Thus, the lowest price p exceeds c. If there is a tie at p, it pays to break the tie by charging a billionth of a cent less than p, and thereby capturing all the shoppers rather than sharing them with the other firm charging p. So there can't be a tie.

But no tie at p means the next lowest firm is charging something strictly greater than p, which means the lowest price firm can increase price somewhat and not suffer any loss of sales. This contradicts profit maximization for that firm. The conclusion is that firms must randomize and no pure strategy equilibrium exists.

But how do they randomize? We are going to look for a distribution of prices. Each firm will choose a price from the continuous distribution F, where $F(x)$ is the probability the firm charges a price less than x. What must F look like? We use the logic of mixed strategies: the firm must get the same profits for all prices that might actually be charged under the mixed strategy, for otherwise it would not be willing to randomize.

A firm that charges price $p \le p_m$ always sells to its captive customers. In addition, it sells to the shoppers if the other firms have higher prices, which occurs with probability $(1-F(p))^{n-1}$. Thus, the firm's profits are

[96] It is often very challenging to assess internet prices because of variation in shipping charges.

$$\pi(p) = (p-c)\left(\frac{1-s}{n} + s(1-F(p))^{n-1}\right).$$

On each sale, the firm earns $p - c$. The firm always sells to its loyal customers, and in addition captures the shoppers if the other firms price higher. Since no firm will exceed p_m, the profits must be the same as the level arising from charging p_m, and this gives

$$\pi(p) = (p-c)\left(\frac{1-s}{n} + s(1-F(p))^{n-1}\right) = (p_m - c)\frac{1-s}{n}.$$

This equation is readily solved for F:

$$F(p) = \left(1 - \frac{(p_m - p)(1-s)}{s(p-c)n}\right)^{\frac{1}{n-1}}.$$

The lower bound of prices arises at the point L where $F(L)=0$, or

$$L = c + \frac{(p_m - c)\frac{1-s}{n}}{\frac{1-s}{n} + s}.$$

These two equations provide a continuous distribution of prices charged by each firm which is an equilibrium to the pricing game. That is, each firm randomizes over the interval $[L, p_m]$, according to the continuous distribution F. Any price in the interval $[L, p_m]$ produces the same profits for each firm, so the firms are willing to randomize over this interval.

The loyal customers get a price chosen randomly from F, so we immediately see that the shoppers make life better for the loyal customers, pushing average price down. (An increase in s directly increases F, which means prices fall – recall that F gives the probability that prices are below a given level, so an increase in F is an increase in the probability of low prices.)

Similarly loyal customers make life worse for shoppers, increasing prices on average to shoppers. The distribution of prices facing shoppers is actually the distribution of the minimum price. Since all firms charge a price exceeding p with probability $(1 - F(p))^n$, at least one charges a price less than p with probability $1 - (1 - F(p))^n$, and this is the distribution of prices facing shoppers. That is, the distribution of prices charged to shoppers is

$$1-(1-F(p))^n = 1 - \left(\frac{(p_m - p)(1-s)}{s(p-c)n}\right)^{\frac{n}{n-1}}.$$

7.3.2 Industry Performance

How does a price dispersed industry perform? First, average industry profits are

$$n\pi(p) = (p_m - c)(1 - s).$$

An interesting aspect of this equation is that it doesn't depend on the number of firms, only on the number of loyal customers. Essentially, the industry profits are the same that it would earn as if the shoppers paid marginal cost and the loyal customers paid the monopoly price, although that isn't what happens in the industry, except in the limit as the number of firms goes to infinity. Note that this formula for industry profits does not work for a monopoly. In order to capture monopoly, one must set s=0, because shoppers have no alternative under monopoly.

Figure 7-4: Expected Prices in Search Equilibrium

As the number of firms gets large, the price charged by any one firm converges to the monopoly price p_m. However, the lowest price offered by any firm actually converges to c, marginal cost. Thus, in the limit as the number of firms get large, shoppers obtain price equal to marginal cost and loyal firms pay the monopoly price.

The average price charged to shoppers and non-shoppers is a complicated object, so we consider the case where there are n firms, $s = ½$, p_m =1 and c = 0. Then the expected prices for shoppers and loyal customers are given in Figure 7-4, letting the number of firms vary. Thus, with many firms, most of the gains created by the shoppers flow to shoppers. In contrast, with few firms, a significant fraction of the gains created by shoppers goes instead to the loyal customers.

Similarly, we can examine the average prices for loyal customers and shoppers when the proportion of shoppers varies. Increasing the proportion of shoppers has two effects. First, it makes low prices more attractive, thus encouraging price competition, because capturing the shoppers is more valuable. Second, it lowers industry profits, because the set of loyal customers is reduced. Figure 7-5 plots the average price for loyal customers and shoppers, as the proportion of shoppers ranges from zero to one, when there are five firms, $p_m = 1$ and $c = 0$.

Figure 7-5: Expected Prices (*s*=Proportion of Shoppers)

People who are price-sensitive and shop around convey a positive externality on other buyers by encouraging price competition. Similarly, people who are less price sensitive and don't shop around convey a negative externality on the other buyers. In markets with dispersed information about the best prices, where some buyers are informed and some are not, randomized prices are a natural outcome. That is, randomization of prices, and the failure of the law of one price, is just a reflection of the different willingness or ability to *search* on the part of consumers.

This difference in the willingness to search could arise simply because search is itself costly. That is, the shoppers could be determined by their choice to shop, in such a way that the cost of shopping just balances the expected gains from searching. The proportion of shoppers may adjust endogenously to insure that the gains from searching exactly equal the costs of searching. In this way, a cost of shopping is translated into a randomized price equilibrium in which there is a benefit from shopping and all consumers get the same total cost of purchase on average.

7.4 Hotelling Model

Breakfast cereals range from indigestible, unprocessed whole grains to cereals that are almost entirely sugar with only the odd molecule or two of grain. Such cereals are hardly good substitutes for each other. Yet similar cereals are viewed by consumers as good substitutes, and the standard model of this kind of situation is the Hotelling model.[97] Hotelling was the first to use a line segment to represent both the product that is sold and the preferences of the consumers who are buying the products. In the Hotelling model, there is a line, and preferences of each consumer is represented by a point on this line. The same line is used to represent products. For example, movie customers are differentiated by age, and we can represent moviegoers by their ages. Movies, too, are designed to be enjoyed by particular ages. Thus a “pre-teen” movie is unlikely to appeal very much to a six year old or to a nineteen year old, while a Disney movie appeals to a six year old, but less to a fifteen year old. That is, movies have a target age, and customers have ages, and these are graphed on the same line.

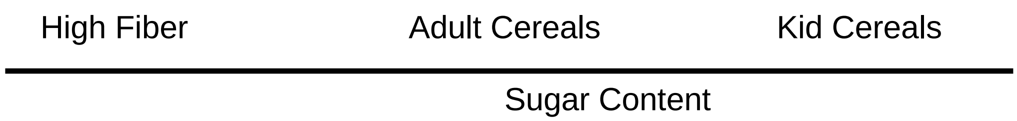

Figure 7-6: Hotelling Model for Breakfast Cereals

Breakfast cereal is a classic application of the Hotelling line, and this application is illustrated in Figure 7-6. Breakfast cereals are primarily distinguished by their sugar content, which ranges on the Hotelling line from low on the left to high on the right. Similarly, the preferences of consumers also fall on the same line. Each consumer has a “most desired point,” and prefers cereals closer to that point than those at more distant points.

7.4.1 Types of Differentiation

There are two main types of differentiation, each of which can be modeled using the Hotelling line. These types are quality and variety. Quality refers to a situation where consumers agree what product is better; the disagreement among consumers concerns whether higher quality is worth the cost. In automobiles, faster acceleration, better braking, higher gas mileage, more cargo capacity, more legroom, and greater durability are all good things. In computers, faster processing, brighter screens, higher resolution screens, lower heat, greater durability, more megabytes of RAM and more gigabytes of hard drive space are all good things. In contrast, varieties are the elements about which there is not widespread agreement. Colors and shapes are usually varietal rather than quality differentiators. Some people like almond colored appliances, others choose white, with blue a distant third. Food flavors are varieties, and while the quality of ingredients is a quality differentiator, the type of food is usually a varietal differentiator. Differences in music would primarily be varietal.

Quality is often called *vertical differentiation*, while variety is *horizontal differentiation*.

[97] Hotelling Theory is named for Harold Hotelling, 1895-1973.

7.4.2 The Standard Model

The standard Hotelling model fits two ice cream vendors on a beach. The vendors sell the identical product, and moreover they can choose to locate wherever they wish. For the time being, suppose the price they charge for ice cream is fixed at $1. Potential customers are also spread randomly along the beach.

We let the beach span an interval from 0 to 1. People desiring ice cream will walk to the closest vendor, since the price is the same. Thus, if one vendor locates at x and the other at y, and $x < y$, those located between 0 and ½ $(x + y)$ go to the left vendor, while the rest go to the right vendor. This is illustrated in Figure 7-7.

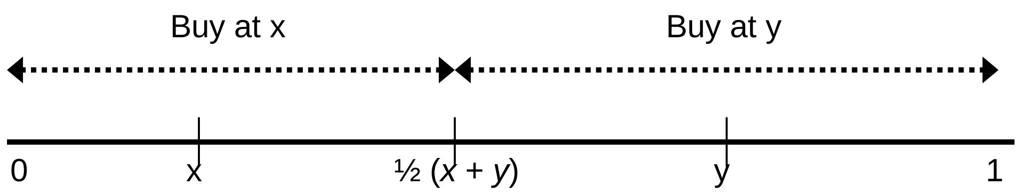

Figure 7-7: Sharing the Hotelling Market

Note that the vendor at x sells more by moving toward y, and vice versa. Such logic forces profit maximizing vendors to both locate in the middle! The one on the left sells to everyone left of ½, while the one on the right sells to the rest. Neither can capture more of the market, so equilibrium locations have been found. (To complete the description of an equilibrium, we need to let the two "share" a point and still have one on the right side, one on the left side of that point.)

This solution is commonly used as an explanation of why U.S. political parties often seem very similar to each other – they have met in the middle in the process of chasing the most voters. Political parties can't directly buy votes, so the "price" is fixed; the only thing parties can do is locate their platform close to voters' preferred platform, on a scale of "left" to "right." But the same logic that a party can grab the middle, without losing the ends, by moving closer to the other party will tend to force the parties to share the same "middle of the road" platform.

7.4.2.1 (Exercise) Suppose there are four ice cream vendors on the beach, and customers are distributed uniformly. Show that it is a Nash equilibrium for two to locate at ¼, and two at ¾.

The model with constant prices is unrealistic for the study of the behavior of firms. Moreover, the two-firm model on the beach is complicated to solve and has the undesirable property that it matters significantly whether the number of firms is odd or even. As a result, we will consider a Hotelling model on a circle, and let the firms choose their prices.

7.4.3 The Circle Model

In this model, there are n firms evenly spaced around the circle whose circumference is one. Thus, the distance between any firm and each of its closest neighbors is $1/n$.

Consumers care about two things: how distant the firm they buy from is, and how much they pay for the good, and they minimize the sum of the price paid and t times the distance between the consumer's location (also on the circle) and the firm. Each consumer's preference is uniformly distributed around the circle. The locations of firms are illustrated in Figure 7-8.

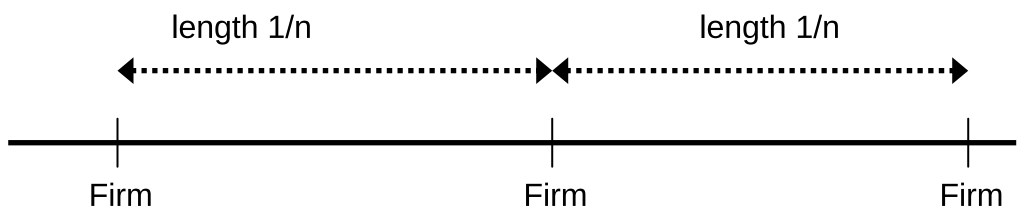

Figure 7-8: A Segment of the Circle Model

We conjecture a Nash equilibrium in which all firms charge the price p. To identify p, we look for what p must be to make any one firm choose to charge p, given that the others all charge p. So suppose the firm in the middle of Figure 7-8 charges an alternate price r, but every other firm charges p. A consumer who is x units away from the firm pays the price $r + tx$ from buying at the firm, or $p + t(1/n - x)$ from buying from the rival. The consumer is just indifferent between the nearby firms if these are equal, that is,

$$r + tx^* = p + t(1/n - x^*)$$

where x^* is the location of the consumer who is indifferent.

$$x^* = \frac{p + t/n - r}{2t} = \frac{1}{2n} + \frac{p-r}{2t}.$$

Thus, consumers who are closer than x^* to the firm charging r buy from that firm, and consumers who are further away than x^* buy from the alternative firm. Demand for the firm charging r is twice x^* (because the firm sells to both sides), so profits are price minus marginal cost times two x^*, that is,

$$(r-c)2x^* = (r-c)\left(\frac{1}{n} + \frac{p-r}{t}\right).$$

The first order condition[98] for profit maximization is

$$0 = \frac{\partial}{\partial r}(r-c)\left(\frac{1}{n} + \frac{p-r}{t}\right) = \left(\frac{1}{n} + \frac{p-r}{t}\right) - \frac{r-c}{t}.$$

[98] Since profit is quadratic in r, we will find a global maximum.

We could solve the first order condition for r. But remember that the concern is when is p a Nash equilibrium price? The price p is an equilibrium price if the firm wants to choose $r = p$. Thus, we can conclude that p is a Nash equilibrium price when

$$p = c + \frac{t}{n}.$$

This value of p insures that a firm facing rivals who charge p also chooses to charge p. Thus, in the Hotelling model, price exceeds marginal cost by an amount equal to the value of the average distance between the firms, since the average distance is $1/n$ and the value to a consumer of traveling that distance is t. The profit level of each firm is $\frac{t}{n^2}$, so industry profits are $\frac{t}{n}$.

How many firms will enter the market? Suppose the fixed cost is F. We are going to take a slightly unusual approach and assume that the number of firms can adjust in a continuous fashion, in which case the number of firms is determined by the zero profit condition $F = \frac{t}{n^2}$, or $n = \sqrt{t/F}$.

What is the socially efficient number of firms? The socially efficient number of firms minimizes the total costs, which are the sum of the transportation costs and the fixed costs. With n firms, the average distance a consumer travels is

$$n \int_{-1/2n}^{1/2n} |\, x \,|\, dx = 2n \int_0^{1/2n} x dx = n\left(\frac{1}{2n}\right)^2 = \frac{1}{4n}.$$

Thus, the socially efficient number of firms minimizes the transport costs plus the entry costs $\frac{t}{4n} + nF$. This occurs at $n = \frac{1}{2}\sqrt{t/F}$. The socially efficient number of firms is half the level that enter with free entry!

Too many firms enter in the Hotelling circle model. This extra entry arises because efficient entry is determined by the cost of entry and the average distance of consumers, while prices are determined by the marginal distance of consumers, or the distance of the marginal consumer. That is, competing firms' prices are determined by the most distant customer, and that leads to prices that are too high relative to the efficient level; free entry then drives net profits to zero only by excess entry.

The Hotelling model is sometimes used to justify an assertion that firms will advertise too much, or engage in too much R&D, as a means of differentiating themselves and creating profits.

7.5 Agency Theory

An *agent* is a person who works for, or on behalf of, another. Thus, an employee is an agent of a company. But agency extends beyond employee relationships. Independent contractors are also agents. Advertising firms, lawyers and accountants are agents of their clients. The CEO of a company is an agent of the board of directors of the company. A grocery store is an agent of the manufacturer of corn chips sold in the store. Thus, the agency relationship extends beyond the employee into many different economic relationships. The entity – person or corporation – on whose behalf an agent works is called a *principal*.

Agency theory is the study of incentives provided to agents. Incentives are an issue because agents need not have the same interests and goals as the principal. Employees spend billions of hours every year browsing the web, emailing friends, and playing computer games while they are supposedly working. Attorneys hired to defend a corporation in a lawsuit have an incentive not to settle, to keep the billing flowing. (Such behavior would violate the attorneys' ethics requirements.) Automobile repair shops have been known to use cheap or used replacement parts and bill for new, high quality parts. These are all examples of a conflict in the incentives of the agent and the goals of the principal.

Agency theory focuses on the cost of providing incentives. When you rent a car, an agency relationship is created. Even though a car rental company is called an agency, it is most useful to look at the renter as the agent, because it is the renter's behavior that is an issue. The company would like the agent to treat the car as if it were their own car. The renter, in contrast, knows it isn't their own car, and often drives accordingly.

> "[T]here's a lot of debate on this subject---about what kind of car handles best. Some say a front-engined car; some say a rear-engined car. I say a *rented* car. Nothing handles better than a rented car. You can go faster, turn corners sharper, and put the transmission into reverse while going forward at a higher rate of speed in a rented car than in any other kind." [99]

How can the car rental company insure that you don't put their car into reverse while going forward at a high rate of speed? They could monitor your behavior, perhaps by putting a company representative in the car with you. That would be a very expensive and unpleasant solution to the problem of incentives. Instead, the company uses outcomes – if damage is done, the driver has to pay for it. That is also an imperfect solution, because some drivers who abuse the cars get off scot-free and others who don't abuse the car still have cars that break down, and are then mired in paperwork while they try to prove their good behavior. That is, a rule that penalizes drivers based on outcomes imposes risk on the drivers. Modern technology is improving monitoring with GPS tracking.

---

[99] P. J. O'Rourke, Republican Party Reptile, Atlantic Monthly Press, 1987.

7.5.1 Simple Model

To model the cost of provision of incentives, we consider an agent like a door-to-door encyclopedia salesperson. The agent will visit houses, and sell encyclopedias to some proportion of the households; the more work the agent does, the more sales that are made. We let x represent the average dollar value of sales for a given level of effort; x is a choice the agent makes. However, x will come with risk to the agent, which we model using the variance σ^2.

The firm will pay the agent a share s of the money generated by the sales. In addition, the firm will pay the agent a salary y, which is fixed independently of sales. This scheme – a combination of salary and commission – covers many different situations. Real estate agents receive a mix of salary and commission. Authors receive an advance and a royalty, which works like a salary and commission.

The monetary compensation of the agent is $sx + y$. In addition, the agent has a cost of effort, which we take to be $\frac{x^2}{2a}$. Here a represents the ability of the agent: more able agents, who have a higher value of a, have a lower cost of effort. Finally, there is a cost of risk. The actual risk imposed on the agent is proportional to the degree they share in the proceeds; if s is small, the agent faces almost no monetary risk, while if s is high, most of the risk is imposed on the agent. We use the "linear cost of risk" model developed earlier, to impose a cost of risk which is $s\lambda\sigma^2$. Here, σ^2 is the variance of the monetary risk, λ defines the agent's attitude or cost of risk, and s is the share of the risk imposed on the agent. This results in a payoff to the agent of

$$u = sx + y - \frac{x^2}{2a} - s\lambda\sigma^2 .$$

The first two terms, $sx + y$, are the payments made to the agent. The next term is the cost of generating that level of x. The final term is the cost of risk imposed on the agent by the contract.

The agency game works as follows. First, the principal offers a contract, which involves a commission s and a salary y. The agent can either accept or reject the contract and accepts if he obtains at least u_0 units of utility, the value of his next best offer. Then the agent decides how much effort to expend, that is, the agent chooses x.

As with all subgame perfect equilibria, we work backwards to first figure out what x an agent would choose. Because our assumptions make u quadratic in x, this is a straightforward exercise, and we conclude $x=sa$. This can be embedded into u, and we obtain the agent's optimized utility, u^*, is

$$u^* = s^2 a + y - \frac{(sa)^2}{2a} - s\lambda\sigma^2 = y + ½\, s^2 a - s\lambda\sigma^2 .$$

Incentive Compensation: A Percentage of What?

Most companies compensate their sales force based on the revenue generated. However, maximizing revenue need not be the same thing as maximizing profit, which is generally the goal of the company. In what follows, Steve Bisset discusses the difference.

“Many years ago I was CEO of a company called Megatest, founded by Howard Marshall and myself in 1975. Around 1987 we were selling test systems in the $1M+ price range. Every Monday morning we would have a meeting with sales management and product marketing, mediated by myself. Individual salesmen came in to make their cases for how they just had to have a huge discount, else they would lose a particular sale. The meeting was part circus, with great performances, and part dogfight.

“I could visualize the sales guys spending their time in the car or the shower using their substantial creative powers to dream up good justifications for their next plea for a huge discount. They knew that if we were willing to bleed enough we could usually win the sale. I wanted to solve both the resultant profitability problem and the problem of the unpleasant meeting.

“Commissions were traditionally a percentage of bookings (net of discount), with part held back until cash was received. The percentage increased after a salesman met his individual quota for the quarter (the performances at quota-setting time to sandbag the quota were even more impressive). The fact that a discount reduced commission did not affect a salesman's behavior, because the difference in commission was small. Better to buy the order by giving a big discount and then move on quickly to the next sale.

“Salesmen are "coin operated", and will figure out how to maximize their total commission. Most salesmen I have met are quite good at math. Further, they have learned to “watch the hips, not the lips” – in other words, they judge management intentions by actions rather than words. They reason – and I agree with them – that if management really wanted them to maximize margins rather than revenues, then they would pay them more if they maximize margins.

“We decided to try a new scheme, against some howling from the sales department. We set a base "cost" for each product, approximately representing the incremental cost to manufacture and support the product. Then we offered commission on the amount that the net sales price exceeded this base cost. The base cost may have been as much as 80% of the list price (it was a very competitive market at the time). Then we increased the commission rate by a factor of about six, so that if the salesman brought in an order at a price near list, then his commission was somewhat higher than before. If he started discounting, his commission dropped sharply. We gave broad discretion to sales management to approve discounts.

“We still had sales guys claiming that a sale was massively strategic and had to be sold at or below cost, and that they needed their commission anyway to justify the effort. In some cases we approved this, but mostly we said that if it's strategic then you'll get your commission on the follow-on sales. While salesmen have a strong preference for immediate cash, they will act so as to maximize income over time, and will think and act strategically if financially rewarded for such.

“The results were that our margins increased significantly. Revenues were increasing too. It's hard to attribute the revenue gain to the new commission plan, given the number of other variables, but I like to think that it was a factor. Now our salesmen spent more of their creative energies devising ways to sell our customers on the value of our products and company, instead of conspiring with sales management as to the best tale of woe to present to marketing at the Monday pricing meeting.

“The Monday meetings became shorter and more pleasant, focused on truly creative ways to make each sale. There certainly were steep discounts given in some cases, but they reflected the competitive situation and future sales potential at each account much more accurately.”

(Source: private correspondence, quotation permission received)

The agent won't accept employment unless $u^*\geq u_0$, the reservation utility. The principal can minimize the cost of employing the agent by setting the salary such that $u^*=u_0$, which results in

$$y = u_0 - \tfrac{1}{2}s^2 a + s\lambda\sigma^2 .$$

Observe that the salary has to be higher, the greater is the risk σ^2. That is, the principal has to cover the cost of risk in the salary term.

7.5.2 Cost of Providing Incentives

The principal obtains profits which are the remainder of the value after paying the agent, and minus the salary:

$$\begin{aligned}\pi &= (1-s)x - y \\ &= (1-s)sa - (u_0 - \tfrac{1}{2}s^2 a + s\lambda\sigma^2) \\ &= sa - u_0 - \tfrac{1}{2}s^2 a - s\lambda\sigma^2.\end{aligned}$$

Note that the principal gets the entire output $x = sa$ minus all the costs – the reservation utility of the agent u_0, the cost of providing effort, and the risk cost on the agent. That is, the principal obtains the full gains from trade – the value of production minus the total cost of production. However, the fact that the principal obtains the full gains from trade doesn't mean the principal induces the agent to work extremely hard, because there is no mechanism for the principal to induce the agent to work hard without imposing more risk on the agent, and this risk is costly to the principal. Agents are induced to work hard by tying their pay to their performance, and such a link necessarily imposes risk on the agent, and risk is costly.[100]

We take the principal to be risk neutral. This is reasonable when the principal is "economically large" relative to the agent, so that the risks faced by the agent are small to the principal. For example, the risks associated with any one car are small to a car rental company. The principal who maximizes expected profits chooses s to maximize π, which yields

$$s = 1 - \frac{\lambda}{a}\sigma^2 .$$

This formula is interesting for several reasons. First, if the agent is neutral to risk, which means $\lambda=0$, then s is 1. That is, the agent gets 100% of the marginal return to effort, and the principal just collects a lump sum. This is reminiscent of some tenancy contracts used by landlords and peasants; the peasant paid a lump sum for the right to farm the land and then kept all of the crops grown. Since these peasants were unlikely to be risk neutral, while the landlord was relatively neutral to risk, such a contract was

[100] There is a technical requirement that the principal's return π must be positive, for otherwise the principal would rather not contract at all. This amounts to an assumption that u_0 is not too large. Moreover, if s comes out less than zero, the model falls apart, and in this case, the actual solution is s=0.

unlikely to be optimal. The contract with s=1 is known as "selling the agency" since the principal sells the agency to the agent for a lump sum payment. (Here, y will generally be negative – the principal gets a payment rather than paying a salary.) The more common contract, however, had the landowner and the tenant farmer share the proceeds of farming, which gives rise to the name *sharecropper*.

Second, more risk or more risk aversion on the part of the agent decreases the share of the proceeds accruing to the agent. Thus, when the cost of risk or the amount of risk is high, the best contract imposes less risk on the agent. Total output sa falls as the risk costs rise.

Third, more able agents (higher a) get higher commissions. That is, the principal imposes more risk on the more able agent because the returns to imposition of risk – in the form of higher output – are greater, and thus worth the cost in terms of added risk.

Most real estate agencies operate on a mix of salary and commission, with commissions paid to agents averaging about 50%. The agency RE/MAX, however, pays commissions close to 100%, collecting a fixed monthly fee that covers agency expenses from the agents. RE/MAX claims that their formula is appropriate for better agents. The theory developed suggests that more able agents should obtain higher commissions. But in addition, RE/MAX's formula also tends to attract more able agents, since able agents earn a higher wage under the high commission formula. (There is a potential downside to the RE/MAX formula, that it discourages agency-wide cooperation.)

7.5.3 Selection of Agent

Consider what contracts attract what kinds of agents. For a fixed salary y and commission s, the agent's utility, optimizing over x, is

$$u^* = y + ½ s^2 a - s\lambda\sigma^2 .$$

The agent's utility is increasing in a and decreasing in λ. Thus, more able agents get higher utility, and less risk averse agents get higher utility.

How do the terms of the contract affect the pool of applicants? Let us suppose two contracts are offered, one with a salary y_1 and commission s_1, the other with salary y_2 and commission s_2. We suppose $y_2 < y_1$ and $s_2 > s_1$. What kind of agent prefers contract 2, the high commission, low salary contract, over contract 1?

$$y_2 + ½ s_2^2 a - s_2\lambda\sigma^2 \geq y_1 + ½ s_1^2 a - s_1\lambda\sigma^2 ,$$

or, equivalently,

$$½ a\left(s_2^2 - s_1^2\right) - (s_2 - s_1)\lambda\sigma^2 \geq y_1 - y_2 .$$

Thus, agents with high ability a or low level of risk aversion λ prefers the high commission, low salary contract. A company that puts more of the compensation in the form of commission tends to attract more able agents, and agents less averse to risk.

The former is a desirable feature of the incentive scheme, since more able agents produce more. The latter, the attraction of less risk averse agents, may or may not be desirable but is probably neutral overall.

One important consideration is that agents who overestimate their ability will react the same as people who have higher ability. Thus, the contract equally attracts those with high ability and those who overestimate their ability.

Agency theory provides a characterization of the cost of providing incentives. The source of the cost is the link between incentives and risk. Incentives link pay and performance; when performance is subject to random fluctuations, linking pay and performance also links pay and the random fluctuations. Thus the provision of incentives necessarily imposes risk on the agent, and if the agent is risk averse, this is costly.

In addition, the extent to which pay is linked to performance will tend to affect the type of agent who is willing to work for the principal. Thus, a principal must not only consider the incentive to work hard created by the commission and salary structure, but also the type of agent who would choose to accept such a contract.

7.5.4 Multi-tasking

Multi-tasking refers to performing several activities simultaneously. All of us multi-task. We study while drinking a caffeinated beverage; we think about things in the shower; we talk all too much on cell phones and eat French fries while driving. In the context of employees, an individual employee will be assigned a variety of tasks and responsibilities, and the employee must divide their time and efforts among the tasks. Incentives provided to the employee must direct not only the total efforts of the employee, but also the allocation of time and effort across activities. An important aspect of multi-tasking is the interaction of incentives provided to employees, and the effects of changes in one incentive on the behavior of the employee over many different dimensions. In this section, we will establish conditions under which the problem of an employer disaggregates, that is to say, the incentives on each individual task can be set independently of the incentives applied to the others.

This section is relatively challenging and involves a number of pieces. To simplify the presentation, some of the pieces are set aside as claims.

To begin the analysis, we consider a person who has n tasks or jobs. For convenience we will index these activities with the natural numbers 1, 2, … , n. The level of activity, which may also be thought of as an action, in task i will be denoted by x_i. It will prove convenient to denote the vector of actions by $\mathbf{x} = (x_1, \dots, x_n)$. We suppose the agent bears a cost $c(\mathbf{x})$ of undertaking the vector of actions $\mathbf{x}$. We make four assumptions on c:

- c is increasing in each x_i,
- c has a continuous second derivative
- c is strictly convex, and

- c is homogeneous[101] of degree r.

For example, if there are two tasks (n=2), then all four of these assumptions are met by the cost function $c(x_1, x_2) = x_1^2 + x_2^2 + ½ x_1 x_2$. This function is increasing in x_1 and x_2, has continuous derivatives, is strictly convex (more about this below) and is homogeneous of degree 2.

It is assumed that c is increasing to identify the activities as costly. Continuity of derivatives is used for convenience. Convexity of c will insure that a solution to the first order conditions is actually an optimum for the employee. Formally, it means that for any vectors $\mathbf{x} \neq \mathbf{y}$ and scalar α between zero and one ($0 \leq \alpha \leq 1$),

$$\alpha c(\mathbf{x}) + (1-\alpha)c(\mathbf{y}) \geq c(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}).$$

One way of interpreting this requirement is that it is less costly to do the average of two things than the average of the costs of the things. Intuitively, convexity requires that doing a medium thing is less costly than the average of two extremes. This is plausible when extremes tend to be very costly. It also means the set of vectors which cost less than a fixed amount, $\{\mathbf{x} \mid c(\mathbf{x}) \leq b\}$, is a convex set. Thus, if two points cost less than a given budget, the line segment connecting them does, too. Convexity of the cost function insures that the agent's optimization problem is concave, and thus that the first order-conditions describe a maximum. When the inequality is strict for α satisfying $0 < \alpha < 1$, we refer to convexity as *strict convexity*.

The assumption of homogeneity dictates that scale works in a particularly simple manner. Scaling up activities increases costs at a fixed rate r. Homogeneity has very strong implications that are probably unreasonable in many settings. Nevertheless, homogeneity leads to an elegant and useful theory, as we shall see. Recall the definition of a homogeneous function: c is homogeneous of degree r means that for any $\lambda > 0$,

$$c(\lambda\mathbf{x}) = \lambda^r c(\mathbf{x}).$$

Claim: strict convexity implies that $r > 1$.

> Proof of Claim: Fix any $\mathbf{x}$ and consider the two points $\mathbf{x}$ and $\lambda\mathbf{x}$. By convexity, for $0 < \alpha < 1$,
>
> $(\alpha + (1-\alpha)\lambda^r)c(\mathbf{x}) = \alpha c(\mathbf{x}) + (1-\alpha)c(\lambda\mathbf{x})$
>
> $> c(\alpha\mathbf{x} + (1-\alpha)\lambda\mathbf{x})) = (\alpha + (1-\alpha)\lambda)^r c(\mathbf{x})$
>
> which implies $(\alpha + (1-\alpha)\lambda^r) > (\alpha + (1-\alpha)\lambda)^r$.
>
> Define a function k which is the left hand side minus the right hand side:

[101] Homogeneous functions were defined in 4.1.8.3 (Exercise).

$k(\alpha) = \alpha + (1-\alpha)\lambda^{\mathrm{r}} - (\alpha + (1-\alpha)\lambda)^{\mathrm{r}}$. Note that $k(0)=k(1)=0$. Moreover, $k''(\alpha) = -r(r-1)(\alpha + (1-\alpha)\lambda)^{\mathrm{r}-2}(1-\lambda)^2$. It is readily checked that if a convex function of one variable is twice differentiable, then the second derivative is greater than zero. If $r \leq 1$, $k''(\alpha) \geq 0$, implying that k is convex, and hence, if $0 < \alpha < 1$,

$$k(\alpha) = k((1-\alpha)0 + \alpha 1) \leq (1-\alpha)k(0) + \alpha k(1) = 0.$$

Similarly, if $r > 1$, k is concave and $k(\alpha) > 0$. This completes the proof, showing that $r \leq 1$ is not compatible with the strict convexity of c.

How should our person behave? Consider linear incentives, which are also known as piece rates. With piece rates, the employee gets a payment p_i for each unit of x_i produced. The person then chooses $\mathbf{x}$ to maximize

$$u = \sum_{i=1}^{n} p_i x_i - c(\mathbf{x}) = \mathbf{p} \bullet \mathbf{x} - c(\mathbf{x}).$$

Here $\bullet$ is the dot product, which is the sum of the products of the components.

The agent chooses $\mathbf{x}$ to maximize u, resulting in n first order conditions

$$\frac{\partial u}{\partial x_i} = p_i - \frac{\partial c(\mathbf{x})}{\partial x_i} = p_i - c_i(\mathbf{x}),$$

where c_i is the partial derivative of c with respect to the i^{th} argument x_i. This first order condition can be expressed more compactly as

$$0 = \mathbf{p} - c'(\mathbf{x})$$

where $c'(\mathbf{x})$ is the vector of partial derivatives of c. Convexity of c insures that any solution to this problem is a global utility maximum, since the function u is concave, and strict convexity insures that there is at most one solution to the first order conditions.[102]

One very useful implication of homogeneity is that incentives scale. Homogeneity has the effect of turning a very complicated optimization problem into a problem that is readily solved, thanks to this very scaling.

Claim: If all incentives rise by a scalar factor α, then $\mathbf{x}$ rises by $\alpha^{\frac{1}{r-1}}$.

[102] This description is slightly inadequate, because we haven't considered boundary conditions. Often a requirement like $x_i \geq 0$ is also needed. In this case, the first order conditions may not hold with equality for those choices where $x_i = 0$ is optimal.

Proof of Claim: Note that differentiating $c(\lambda\mathbf{x}) = \lambda^r c(\mathbf{x})$ with respect to x_i yields $\lambda c_i(\lambda\mathbf{x}) = \lambda^r c_i(\mathbf{x})$, and thus $c'(\lambda\mathbf{x}) = \lambda^{r-1} c'(\mathbf{x})$. That is, if c is homogeneous of degree r, c' is homogeneous of degree $r - 1$. Consequently, if $0 = \mathbf{p} - c'(\mathbf{x})$, $0 = \alpha\mathbf{p} - c'(\alpha^{\frac{1}{r-1}}\mathbf{x})$. Thus, if the incentives are scaled up by α, the efforts rise by the scalar factor $\alpha^{\frac{1}{r-1}}$.

Now consider an employer with an agent engaging in n activities. The employer values the ith activity at v_i, and thus wishes to maximize

$$\pi = \sum_{i=1}^{n} (v_i - p_i) x_i = \sum_{i=1}^{n} (v_i - c_i(\mathbf{x})) x_i .$$

This equation embodies a standard trick in agency theory. Think of the principal (employer) not as choosing the incentives $\mathbf{p}$, but instead as choosing the effort levels $\mathbf{x}$, with the incentives as a constraint. That is, the principal can be thought of choosing $\mathbf{x}$ and then choosing the $\mathbf{p}$ that implements this $\mathbf{x}$. The principal's expected profit is readily differentiated with respect to each x_j, yielding

$$0 = v_j - c_j(\mathbf{x}) - \sum_{i=1}^{n} c_{ij}(\mathbf{x})) x_i .$$

However, since $c_j(\mathbf{x})$ is homogeneous of degree $r - 1$,

$$\sum_{i=1}^{n} c_{ij}(\mathbf{x})) x_i = \frac{d}{d\lambda} c_j(\lambda\mathbf{x}) \Big|_{\lambda=1} = \frac{d}{d\lambda} \lambda^{r-1} c_j(\mathbf{x}) \Big|_{\lambda=1} = (r-1) c_j(\mathbf{x}),$$

and thus $0 = v_j - c_j(\mathbf{x}) - \sum_{i=1}^{n} c_{ij}(\mathbf{x})) x_i = v_j - r c_j(\mathbf{x})$.

This expression proves the main result of this section. Under the maintained hypotheses (convexity and homogeneity), an employer of a multi-tasking agent uses incentives which are a constant proportion of value, that is,

$$p_j = \frac{v_j}{r},$$

where r is the degree of homogeneity of the agent's costs. Recalling that r>1, the principal uses a *sharing rule*, sharing a fixed proportion of value with the agent.

When agents have a homogeneous cost function, the principal has a very simple optimal incentive scheme, requiring quite limited knowledge of the agent's cost function (just the degree of homogeneity). Moreover, the incentive scheme works through a somewhat surprising mechanism. Note that if the value of one activity, say activity 1, rises, p_1 rises and all the other payment rates stay constant. The agent responds by increasing x_1, but the other activities may rise or fall depending on how complementary they are to activity 1. Overall, the agent's substitution across activities given the new incentive level on activity 1 implements the desired effort levels on other activities. The remarkable implication of homogeneity is that, although the principal desires different effort levels for all activities, only the incentive on activity 1 must change!

7.5.5 Multi-tasking without Homogeneity

In the previous section we saw, for example, that if the agent has quadratic costs, the principal pays the agent half the value of each activity. Moreover, the more rapidly costs rise in scale, the lower the payments to the agent.

This remarkable theorem has several limitations. The requirement of homogeneity is itself an important limitation, although this assumption is reasonable in some settings. More serious is the assumption that *all* of the incentives are set optimally *for the employer*. Suppose, instead, that one of the incentives is set "too high," at least from the employer's perspective. This might arise if, for example, the agent acquired all the benefits of one of the activities. An increase in the power of one incentive will then tend to "spill over" to the other actions, increasing for complements and decreasing for substitutes. When the efforts are substitutes, an increase in the power of one incentive will cause others to optimally rise, to compensate for the reduced supply of efforts of that type.[103]

We can illustrate the effects of cost functions that aren't homogeneous in a relatively straightforward way. Suppose the cost depends on the sum of the squared activity levels:

$$c(\mathbf{x}) = g\left(\sum_{i=1}^{n} x_i^2\right) = g(\mathbf{x} \bullet \mathbf{x}) .$$

This is a situation where vector notation (dot-products) dramatically simplifies the expressions. You may find it useful to work through the notation on a separate sheet, or in the margin, using summation notation to verify each step. At the moment, we won't be concerned with the exact specification of g, but instead use the first order conditions to characterize the solution.

The agent maximizes

$$u = \mathbf{p} \bullet \mathbf{x} - g(\mathbf{x} \bullet \mathbf{x}) .$$

---

[103] Multi-tasking, and agency theory more generally, is a rich theory with many implications not discussed here. For a challenging and important analysis, see Bengt Holmstrom and Paul Milgrom, "The Firm as an Incentive System," *American Economic Review*, Vol. 84, No. 4 (Sep., 1994), pp. 972-991.

This gives a first order condition

$$0 = \mathbf{p} - 2g'(\mathbf{x} \bullet \mathbf{x})\mathbf{x}$$

It turns out that a sufficient condition for this equation to characterize the agent's utility maximization is that *g* is both increasing and convex (increasing second derivative).

This is a particularly simple expression, because the vector of efforts, **x**, points in the same direction as the incentive payments **p**. The scalar that gives the overall effort levels, however, is not necessarily a constant, as occurs with homogeneous cost functions. Indeed, we can readily see that **x•x** is the solution to

$$\mathbf{p} \bullet \mathbf{p} = (2g'(\mathbf{x} \bullet \mathbf{x}))^2 (\mathbf{x} \bullet \mathbf{x}) .$$

Because **x•x** is a number, it is worth introducing notation for it: S = **x•x**. Then S is the solution to

$$\mathbf{p} \bullet \mathbf{p} = 4S(g'(S))^2 .$$

The principal or employer chooses **p** to maximize

$$\pi = \mathbf{v} \bullet \mathbf{x} - \mathbf{p} \bullet \mathbf{x} = \mathbf{v} \bullet \mathbf{x} - 2g'(\mathbf{x} \bullet \mathbf{x})(\mathbf{x} \bullet \mathbf{x}) .$$

This gives the first order condition

$$\mathbf{0} = \mathbf{v} - 4\left(g'(\mathbf{x} \bullet \mathbf{x}) + (\mathbf{x} \bullet \mathbf{x})g''(\mathbf{x} \bullet \mathbf{x})\right)\mathbf{x} .$$

Thus, the principal's choice of **p** is such that **x** is proportional to **v**, with constant of proportionality $g'(\mathbf{x} \bullet \mathbf{x}) + \mathbf{x} \bullet \mathbf{x} g''(\mathbf{x} \bullet \mathbf{x})$. Using the same trick (dotting each side of the first order condition $\mathbf{v} = 4\left(g'(\mathbf{x} \bullet \mathbf{x}) + \mathbf{x} \bullet \mathbf{x} g''(\mathbf{x} \bullet \mathbf{x})\right)\mathbf{x}$ with itself), we obtain:

$$\mathbf{v} \bullet \mathbf{v} = 16\left(g'(S^*) + S^* g''(S^*)\right)^2 S^* ,$$

which gives the level of **x•x**=S^* induced by the principal. Given S^*, **p** is given by

$$\mathbf{p} = 2g'(\mathbf{x} \bullet \mathbf{x})\mathbf{x} = 2g'(S^*)\frac{\mathbf{v}}{4(g'(S^*) + S^* g''(S^*))} = \frac{1}{2}\left(\frac{1}{1 + \frac{S^* g''(S^*)}{g'(S^*)}}\right)\mathbf{v} .$$

Note that this expression gives the right answer when costs are homogeneous. In this case, $g(S)$ must be in the form $S^{r/2}$, and the formula gives

$$\mathbf{p} = \frac{1}{2}\left(\frac{1}{1+r-1}\right)\mathbf{v} = \frac{\mathbf{v}}{r}$$

as we already established.

The natural assumption to impose on the function g is that $\left(g'(S) + Sg''(S)\right)^2 S$ is an increasing function of S. This assumption implies that as the value of effort rises, the total effort also rises.

Suppose $\frac{Sg''(S)}{g'(S)}$ is increasing in S. Then an increase in v_i increases S, decreasing p_j for $j \neq i$. That is, when one item becomes more valuable, the incentives on the others are reduced. Moreover, since $\mathbf{p} \bullet \mathbf{p} = 4S(g'(S))^2$, an increase in S only occurs if $\mathbf{p} \bullet \mathbf{p}$ increases.

These equations together imply that an increase in any one v_i increases the total effort (as measured by $S^* = \mathbf{x} \bullet \mathbf{x}$), increases the total incentives as measured by $\mathbf{p} \bullet \mathbf{p}$, and decreases the incentives on all activities other than activity i. In contrast, if $\frac{Sg''(S)}{g'(S)}$ is a decreasing function of S, then an increase in any one v_i causes *all* the incentives to rise. Intuitively, the increase in v_i directly causes p_i to rise, since x_i is more valuable. This causes the agent to substitute toward activity i. This causes the relative cost of total activity to fall (since $\frac{Sg''(S)}{g'(S)}$ decreases), which induces a desire to increase the other activity levels, which is accomplished by increase in the incentives on the other activities.

This conclusion generalizes readily and powerfully. Suppose that $c(\mathbf{x}) = g(h(\mathbf{x}))$, where h is homogeneous of degree r and g is increasing. In the case just considered, $h(\mathbf{x}) = \mathbf{x} \bullet \mathbf{x}$. Then the same conclusion, that the sign of $\frac{dp_i}{dv_j}$ is determined by the derivative of $\frac{Sg''(S)}{g'(S)}$, holds. In the generalization, S now stands for $h(\mathbf{x})$.

7.6 *Auctions*

When we think of auctions, we tend to think of movies where people scratch their ear and accidentally purchase a Faberge egg, like the one pictured at left.[104] However, stock exchanges, bond markets and commodities markets are organized as auctions, too, and because of such exchanges, auctions are the most common means of establishing prices. Auctions are one of the oldest transactions means recorded in human history, and were used by the Babylonians. The word *auction* comes from the Latin *auctio*, meaning *to increase*.

Auctions have been used to sell a large variety of things. Internet auction house eBay is most famous for weird items that have been auctioned (e.g. one person's attempt to sell their soul), but in addition, many of the purchases of the U.S. government are made by auction. The U.S. purchases everything from fighter aircraft to French fries by auction, and the U.S. government is the world's largest purchaser of French fries. In addition, corporations are occasionally sold by auction.

Items that are usually sold by auction include prize bulls, tobacco, used cars, race horses, coins, stamps, antiques, and fine art.

Information plays a significant role in bidding in auctions. The two major extremes in information, which lead to distinct theories, are *private values*, which means bidders know their own value, and *common values*, in which bidders don't know their own value, but have some indication or signal about the value. In the private values situation, a bidder may be outbid by another bidder, but doesn't learn anything from another bidder's willingness to pay. The case of private values arises when the good being sold has a quality apparent to all bidders, no hidden attributes, and no possibility of resale. In contrast, the case of common values arises when bidders don't know the value of the item for sale, but that value is common to all. The quintessential example is an off-shore oil lease. No one knows for sure how much oil can be extracted from an off-shore lease, and companies have estimates of the amount of oil. The estimates are closely guarded because rivals could learn from them. Similarly, when antiques dealers bid on an antique, the value they place on it is primarily the resale value. Knowing rivals' estimates of the resale value would influence the value each bidder placed on the item.

The private values environment is unrealistic in most settings, because items for sale usually have some element of common value. However, some situations approximate the private values environment and these are the most readily understood.

7.6.1 English Auction

An English auction is the common auction form used for selling antiques, art, used cars and cattle. The auctioneer starts low, and calls out prices until no bidder is willing to bid higher than the current high price. The most common procedure is for a low price to

[104] Photo courtesy of Paris Jewelers, 107 East Ridgewood Ave. Ridgewood, New Jersey 07450.

be called out, and a bidder accept it. A higher price is called out, and a different bidder accepts it. When several accept simultaneously, the auctioneer accepts the first one spotted. This process continues until a price is called out that no one accepts. At that point the auction ends, and the highest bidder wins.

In a private values setting, a very simple bidding strategy is optimal for bidders: a bidder should keep bidding until the price exceeds the value a bidder places on it, at which point the bidder should stop. That is, bidders should drop out of the bidding when the price exceeds their value, because at that point, winning entails a loss. Every bidder should be willing to continue to bid and not let the item sell to someone else if the price is less than their value. If you have a value v and another bidder is about to win at a price $p_a < v$, you might as well accept a price p_b between the two, $p_a < p_b < v$, since a purchase at this price would provide profits. This strategy is a dominant strategy for each private values bidder, because no matter what strategy other bidders adopt, bidding up to value is the strategy that maximizes the profits of a bidder.

The presence of a dominant strategy makes it straightforward to bid in the private values environment. In addition, it makes an analysis of the outcome of the English auction relatively simple.

Most auctioneers use a somewhat flexible system of *bid increments*. A bid increment is the difference between successive price requests. The theory is simplest when the bid increment, which we will denote as Δ, is very small. In this case, the bidder with the highest value will win, and the price will be no more than the second-highest value, but at least the second-highest value minus Δ, since if the price was less than this level, the bidder with the second-highest value would submit another bid. If we denote the second-highest value with the somewhat obscure (but standard) notation $v_{(2)}$, the final price p satisfies

$$v_{(2)} - \Delta \le p \le v_{(2)}.$$

As the bid increment gets small, this nails down the price. The conclusion is that, when bid increments are small and bidders have private values, the bidder with the highest value wins the bidding at a price equal to the second-highest value. The notation for the highest value is $v_{(1)}$, and thus the seller obtains $v_{(2)}$, and the winning bidder obtains profits of $v_{(1)} - v_{(2)}$.

7.6.2 Sealed-bid Auction

In a sealed-bid auction, each bidder submits a bid in an envelope. These are opened simultaneously, and the highest bidder wins the item and pays his or her bid. Sealed-bid auctions are used to sell off-shore oil leases, and used by governments to purchase a wide variety of items. In a purchase situation, known often as a *tender*, the lowest bidder wins and is paid the bid.

The analysis of the sealed-bid auction is more challenging because the bidders don't have a dominant strategy. Indeed, the best bid depends on what the other bidders are bidding. The bidder with the highest value would like to bid a penny more than the next highest bidder's bid, whatever that might be.

To pursue an analysis of the sealed-bid auction, we are going to make a variety of simplifying assumptions. These assumptions aren't necessary to the analysis but are made to simplify the mathematical presentation.

We suppose there are n bidders, and label the bidders $1,\ldots,n$. Bidder i has a private value v_i which is a draw from the uniform distribution on the interval [0,1]. That is, if $0 \le a \le b \le 1$, the probability that bidder i's value is in the interval $[a, b]$ is $b - a$. An important attribute of this assumption is *symmetry* – the bidders all have the same distribution. In addition, the formulation has assumed *independence* – the value one bidder places on the object for sale is statistically independent from the value placed by others. In addition, each bidder knows their own value, but doesn't know the other bidders' values. Each bidder is assumed to bid in such a way as to maximize expected profit, and we will look for a Nash equilibrium of the bidding game. Bidders are permitted to submit any bid not less than zero.

To find an equilibrium, it is helpful to restrict attention to linear rules, in which a bidder bids a proportion of their value. Thus, we suppose each bidder bids λv when their value is v, and examine under what conditions this is in fact Nash equilibrium behavior. We have an equilibrium if, when all other bidders bid λv when their value is v, the remaining bidder will, too.

So fix a bidder and suppose that bidder's value is v_i. What bid should the bidder choose? A bid of b wins the bidding if all other bidders bid less than b. Since the other bidders, by hypothesis, bid λv when their value is v, our bidder wins when $b \ge \lambda v_j$ for each other bidder j. This occurs when $b/\lambda \ge v_j$ for each other bidder j, and this in turn occurs with probability b/λ.[105] Thus, our bidder with value v_i who bids b wins with probability $\left(b/\lambda\right)^{n-1}$, since the bidder must beat all n-1 other bidders. That creates expected profits for the bidder of

$$\pi = (v_i - b)\left(b/\lambda\right)^{n-1}.$$

The bidder chooses b to maximize expected profits. The first order condition requires

$$0 = -\left(\frac{b}{\lambda}\right)^{n-1} + (v_i - b)(n-1)\frac{b^{n-2}}{\lambda^{n-1}}.$$

The first order condition solves for

$$b = \frac{n-1}{n} v.$$

[105] If $b > \lambda$, then in fact the probability is one. You can show that no bidder would ever bid more than λ.

But this is a linear rule! Thus, if $\lambda = \frac{n-1}{n}$, we have a Nash equilibrium.

The nature of this equilibrium is that each bidder bids a fraction $\lambda = \frac{n-1}{n}$ of their value, and the highest value bidder wins at a price equal to that fraction of their value.

In some cases, the sealed-bid auction produces *regret*. Regret means that a bidder wishes she had bid differently. Recall our notation for values: $v_{(1)}$ is the highest value and $v_{(2)}$ is the second-highest value. Since the price in a sealed-bid auction is $\frac{n-1}{n} v_{(1)}$, the second-highest bidder will regret her bid when $v_{(2)} > \frac{n-1}{n} v_{(1)}$. In this case, the bidder with the second-highest value could have bid higher and won, if the bidder had known the winning bidder's bid. In contrast, the English auction is regret-free, in that the price rises to the point that the bidder with the second-highest value won't pay.

How do the two auctions compare in prices? It turns out that statistical independence of private values implies *revenue equivalence*, which means the two auctions produce the same prices on average. Given the highest value $v_{(1)}$, the second-highest value has distribution $\left(\frac{v_{(2)}}{v_{(1)}}\right)^{n-1}$, since this is the probability that all n-1 other bidders have values less than $v_{(2)}$. But this gives an expected value of $v_{(2)}$ of

$$Ev_{(2)} = \int_0^{v_{(1)}} v_{(2)}(n-1)\frac{v_{(2)}^{n-2}}{v_{(1)}^{n-1}} dv_{(2)} = \frac{n-1}{n} v_{(1)}.$$

Thus, the average price paid in the sealed-bid auction is the same as the average price in the English auction.

7.6.3 Dutch Auction

The Dutch auction is like an English auction, except that prices start high and are successively dropped until a bidder accepts the going price, at which point the auction ends. The Dutch auction is so named because it is used to sell cut flowers in Holland, in the enormous flower auctions.

A strategy in a Dutch auction is a price at which the bidder bids. Each bidder watches the price decline, until such a point that either the bidder bids, or a rival bids, and the auction ends. Note that a bidder could revise their bid in the course of the auction, but there isn't any point. For example, suppose the price starts at $1,000, and a bidder decides to bid when the price reaches $400. Once the price gets to $450, the bidder could decide to revise and wait until $350. However, no new information has become available and there is no reason to revise. In order for the price to reach the original planned bid of $400, it had to reach $450, meaning that no one bid prior to a price of $450. In order for a bid of $400 to wins, the price had to reach $450; if the price reaching $450 means that a bid of $350 is optimal, than the original bid of $400 wasn't optimal.[106]

What is interesting about the Dutch auction is that it has exactly the same possible strategies and outcomes as the sealed-bid auction. In both cases, a strategy for a bidder is a bid, no bidder sees the others' bids until after their own bid is formulated, and the winning bidder is the one with the highest bid. This is called *strategic equivalence*. Both games – the Dutch auction and the sealed-bid auction – offer identical strategies to the bidders, and given the strategies chosen by all bidders, produce the same payoff. Such games should produce the same outcomes.

The strategic equivalence of the Dutch auction and the sealed-bid auction is a very general result, which doesn't depend on the nature of the values of the bidders (private versus common) or the distribution of information (independent versus correlated). Indeed, the prediction that the two games should produce the same outcome doesn't even depend on risk aversion, although that is more challenging to demonstrate.

7.6.4 Vickrey Auction

The strategic equivalence of the Dutch and sealed-bid auction suggests another fact: there may be more than one way of implementing a given kind of auction. Such logic led Nobel laureate William Vickrey (1914-1996) to design what has become known as the *Vickrey* auction, which is a "second-price sealed-bid" auction. This auction is most familiar because it is the foundation of eBay's auction design. The Vickrey auction is a sealed-bid auction, but with a twist: the high bidder wins, but pays the second-highest bid. This is why the Vickrey auction is called a second-price auction: the price is not the highest bid, but the second-highest bid.

The Vickrey auction underlies the eBay outcome because when a bidder submits a bid in the eBay auction, the current "going" price is not the highest bid, but the second-highest bid, plus a bid increment. Thus, up to the granularity of the bid increment, the basic eBay auction is a Vickrey auction run over time.

As in the English auction, bidders with private values in a Vickrey auction have a dominant strategy. Fix a bidder, with value v, and let p be the highest bid of the other bidders. If the bidder bids b, the bidder earns profits of

$$\begin{cases} 0 & \text{if } b < p \\ v - p & \text{if } b > p \end{cases}.$$

[106] Of course, a bidder who thinks losing is likely may wait for a lower price to formulate the bid, a consideration ignored here. In addition, because the Dutch auction unfolds over time, bidders who discount the future will bid slightly higher in a Dutch auction as a way of speeding it along, another small effect that is ignored for simplicity.

It is profitable for the bidder to win if $v > p$, and lose if $v < p$. To win when $v > p$, and lose if $v < p$, can be assured by bidding v. Essentially, there is no gain to bidding less than your value, because your bid doesn't affect the price, only the likelihood of winning. Bidding less than value causes the bidder to lose when the highest rival bid falls between the bid and the value, which is a circumstance that the bidder would like to win. Similarly, bidding more than value only creates a chance of winning when the price is higher than the bidder's value, in which case the bidder would prefer to lose.

Thus, bidders in a Vickrey auction have a dominant strategy to bid their value. This produces the same outcome as the English auction, however, because the payment made is the second-highest value, which was the price in the English auction. Thus, the Vickrey auction is a sealed-bid implementation of the English auction when bidders have private values, producing the same outcome, which is that the highest value bidder wins, but pays the second-highest value.

Because the Vickrey auction induces bidders to bid their value, it is said to be *demand revealing*. Unlike the English auction, in which the bidding stops when the price reaches the second-highest value and thus doesn't reveal the highest value, the Vickrey auction reveals the highest value. In a controlled, laboratory setting, demand revelation is useful, especially when the goal is to identify buyer values. Despite its theoretical niceties, the Vickrey auction can be politically disastrous. Indeed, New Zealand sold radio spectrum with the Vickrey auction on the basis of advice by a naïve economist, and the Vickrey auction created a political nightmare when a nationwide cellular license received a high bid of $110 million, and a second-highest bid of $11 million. The political problem was that the demand revelation showed that the government received only about 10% of the value of the license, making the public quite irate and dominating news coverage at the time.[107] Some smaller licenses sold for tenths of a percent of the highest bid.

In a private values setting, the Vickrey auction, or the English auction, are much easier on bidders than a regular sealed-bid auction, because of the dominant strategy. The sealed-bid auction requires bidders to forecast their rivals' likely bids, and produces the risks of either bidding more than necessary, or losing the bidding. Thus, the regular sealed-bid auction has undesirable properties. Moreover, bidders in the sealed-bid auction have an incentive to bribe the auctioneer to reveal the best bid by rivals, because that is useful information in formulating a bid. Such (illegal) bribery occurs from time to time in government contracting.

On the other hand, the regular sealed-bid auction has an advantage over the other two that it makes price-fixing more difficult. A bidder can cheat on a conspiracy and not be detected until after the current auction is complete.

Another disadvantage of the sealed-bid auction is that it is easier to make certain kinds of bidding errors. In the U.S. PCS auctions, in which rights to use the radio spectrum

[107] The Vickrey auction generally produces higher prices than regular sealed-bid auctions if bidders are symmetric (share the same distribution of values), but is a poor choice of auction format when bidders are not symmetric. Since the incumbent telephone company was expected to have a higher value than others, the Vickrey auction was a poor choice for that reason as well.

for cellular phones was sold for around $20 billion, one bidder, intending to bid $200,000, inadvertently bid $200,000,000. Such an error isn't possible in an English auction, because prices rise at a measured pace. Such errors have little consequence in a Vickrey auction, since getting the price wrong by an order of magnitude requires two bidders to make such errors.

7.6.5 Winner's Curse

"I paid too much for it, but it's worth it."
-Sam Goldwyn

The analysis so far has been conducted under the restrictive assumption of private values. In most contexts, bidders are not sure of the actual value of the item being sold, and information held by others is relevant to the valuation of the item. If I estimate an antique to be worth $5,000, but no one else is willing to bid more than $1,000, I might revise my estimate of the value down. This revision leads bidders to learn from the auction itself what the item is worth.

The early bidders in the sale of oil lease rights in the Gulf of Mexico (the outer continental shelf) were often observed to pay more than the rights were worth. This phenomenon came to be known as the *winner's curse.* The winner's curse is the fact that *the bidder who most overestimates the value of the object wins the bidding*. Naïve bidders, who don't adjust for the winner's curse, will tend to lose money because they only win the bidding when they've bid too high.

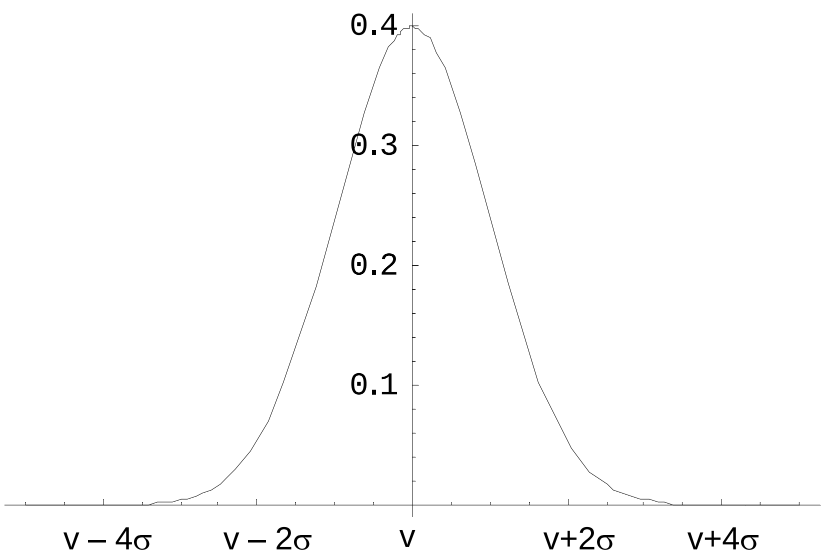

Figure 7-9: Normally Distributed Estimates

Auctions, by their nature, select optimistic bidders. Consider the case of an oil lease (right to drill for and pump oil) that has an unknown value v. Different bidders will obtain different estimates of the value, and we may view these estimates as draws from a normal distribution illustrated in Figure 7-9. The estimates are correct on average, which is represented by the fact that the distribution is centered on the true value v. Thus a randomly chosen bidder will have an estimate that is too high as often as it is too low, and the average estimate of a randomly selected bidder will be correct. However, the winner of an auction will tend to be bidder with the highest estimate, not a randomly chosen bidder. The highest of five bidders will have an estimate that is too large 97% of the time. The only way the highest estimate is not too large is if all the estimates are below the true value. With ten bidders, the highest estimate is larger than the true value with probability 99.9%, because the odds that all the estimates are less than the true value is $(\frac{1}{2})^{10} = 0.1\%$. This phenomenon – that auctions tend to select the bidder with the highest estimate, and the highest estimate is larger than the true value most of the time – is characteristic of the winner's curse.

A savvy bidder corrects for the winner's curse. Such a correction is actually quite straightforward when a few facts are available, and here a simplified presentation is given. Suppose there are *n* bidders for a common value good, and the bidders receive normally distributed estimates that are correct on average. Let σ be the standard deviation of the estimates.[108] Finally, suppose that no prior information is given about the likely value of the good.

In this case, it is a straightforward matter to compute a correction for the winner's curse. Because the winning bidder will generally be the bidder with the highest estimate of value, the winner's curse correction should be the expected amount by which the highest value exceeds the average value. This can be looked up in a table for the normal distribution. The values are given for selected numbers *n* in Table 7-30. This shows, as a function of the number of bidders, how much each bidder should reduce their estimate of value to correct for the fact that auctions select optimistic bidders. The units are standard deviations.

Table 7-30: Winner's Curse Correction

n	1	2	3	4	5	10	15
WCC (σ)	0	.56	.85	1.03	1.16	1.54	1.74
n	20	25	50	100	500	1000	10,000
WCC (σ)	1.87	1.97	2.25	2.51	3.04	3.24	3.85

For example, with one bidder, there is no correction, since it was supposed that the estimates are right on average. With two bidders, the winner's curse correction is the amount that the higher of two will be above the mean, which turns out to be 0.56σ, a little more than half a standard deviation. This is the amount which should be subtracted from the estimate to insure that, when the bidder wins, the estimated value is on average correct. With four bidders, the highest is a bit over a whole standard deviation. As is apparent from the table, the winner's curse correction increases relatively slowly after ten or fifteen bidders. With a million bidders, it is 4.86σ.

---

[108] The standard deviation is a measure of the dispersion of the distribution, and is the square root of the average of the square of the difference of the random value and its mean. The estimates are also assumed to be independently distributed around the true value. Note that estimating the mean adds an additional layer of complexity.

The standard deviation σ measures how much randomness or noise there is in the estimates. It is a measure of the average difference between the true value and the estimated value, and thus the average level of error. Oil companies know from their history of estimation how much error arises in the company estimates. Thus, they can correct their estimates to account for the winner's curse using their historical inaccuracies.

Bidders who are imperfectly informed about the value of an item for sale are subject to losses arising from the way auctions select the winning bidder. The winning bidder is usually the bidder with the highest estimate, and that estimate is too high on average. The difference between the highest estimate and the average estimate is known as the winner's curse correction. The size of the winner's curse correction is larger the more bidders there are but tends to grow slowly beyond a dozen or so bidders.

If the bidders have the same information on a common value item, they will generally not earn profits on it. Indeed, there is a general principle that it is the privacy of information, rather than the accuracy of information, that leads to profits. Bidders earn profits on the information that they hold that is not available to others. Information held by others will be built into the bid price and therefore not lead to profits.

7.6.6 Linkage

The U.S. Department of the Interior, when selling off-shore oil leases, not only takes an upfront payment (the winning bid) but also takes 1/6 of the oil that is eventually pumped. Such a royalty scheme links the payment made to the outcome, and in a way, shares risk, since the payment is higher when there is more oil. Similarly, a book contract provides an author with an upfront payment and a royalty. Many U.S. Department of Defense purchases of major weapons systems involve cost-sharing, where the payments made pick up a portion of the cost. Purchases of ships, for example, generally involve 50% to 70% cost sharing, which means the DOD pays a portion of cost overruns. The contract for U.S. television broadcast rights for the summer Olympics in Seoul, South Korea, involved payments that depended on the size of the U.S. audience.

Royalties, cost-sharing and contingent payments generally link the actual payment to the actual value, which is unknown at the time of the auction. Linkage shares risk, which is a topic already considered in Section 7.5. But linkage does something else, too. Linkage reduces the importance of estimates in the auction, replacing the estimates with actual values. That is, the price a bidder pays for an object, when fully linked to the true value, is just the true value. Thus, linkage reduces the importance of estimation in the auction by taking the price out of the bidder's hands, at least partially.

The *linkage principle*[109] states that, in auctions where bidders are buyers, the expected price rises the more the price is linked to the actual value. (In a parallel fashion, the expected price in an auction where bidders are selling falls.) Thus, linking price to value generally improves the performance of auctions. While this is a mathematically deep result, an extreme case is straightforward to understand. Suppose the government is purchasing by auction a contract for delivery of 10,000 gallons of gasoline each week for

[109] The linkage principle, and much of modern auction theory, was developed by Paul Milgrom (1948 –).

the next year. Suppliers face risk in the form of gasoline prices; if the government buys at a fixed price, the suppliers' bids will build in a cushion to compensate for the risk, and for the winner's curse. In addition, because their estimates of future oil prices will generally vary, they will earn profits based on their private information about the value. In contrast, if the government buys only delivery and then pays for the cost of the gasoline, whatever it might be, any profits that the bidders earned based on their ability to estimate gasoline prices evaporates. The overall profit level of bidders falls, and the overall cost of the gasoline supply can fall. Of course, paying the cost of the gasoline reduces the incentive of the supplier to shop around for the best price, and that agency incentive effect must be balanced against the reduction in bidder profits from the auction to select a supplier.

7.6.7 Auction Design

We saw above that the English auction tends to reduce regret relative to sealed-bid auctions, and that the linkage principle suggests tying payments to value where possible. These are examples of auction design, in which auctions are designed to satisfy objectives of the auction designer. Proper auction design should match the rules of the auction to the circumstances of the bidders and the goal of the seller. Some of the principles of auction design include:

- Impose an appropriate reserve price or minimum bid
- Use ascending price (English) auctions rather than sealed-bid
- Reveal information about the value of the item
- Conceal information about the extent of competition
- Handicap bidders with a known advantage

However, many of these precepts change if the seller is facing a cartel. For example, it is easier for bidders to collude in a sealed-bid auction than in an English auction; and reserve prices should be made relatively high.

Reserve prices (minimum bid) have several effects. They tend to force marginal bidders to bid a bit higher, which increases bids of all bidders, reducing bidder profits. However, reserve prices also lead to a failure to sell on occasion, and the optimal reserve trades off this failure to sell against the higher prices. In addition, reserve prices may reduce the incentive of bidders to investigate the sale, reducing participation, which is an additional negative consideration for a high reserve price.

Ascending price auctions like the English auction have several advantages. Such auctions reduce the complexity of the bidder's problem, because bidder's can stretch their calculations out over time, and because bidders can react to the behavior of others and not plan for every contingency in advance. In addition, because bidders in an English auction can see the behavior of others, there is a linkage created – the price paid by the winning bidder is influenced not just by that bidder's information but also by the information held by others, tending to drive up the price, which is an advantage for the seller.

One caveat to the selection of the English auction is that risk aversion doesn't affect the outcome in the private values case. In contrast, in a sealed-bid auction, risk aversion works to the advantage of the seller, because bidders bid a little bit higher than they would have otherwise, to reduce the risk of losing. Thus, in the private values case, risk averse bidders will bid higher in the sealed-bid auction than in the English auction.

When considering the revelation of information, there is always an issue of lying and misleading. In the long-run, lying and misleading is found out, and thus the standard approach is to ignore the possibility of lying. Making misleading statements is, in the long-run, the same thing as silence, since those who repeatedly lie or mislead are eventually discovered, and then not believed. Thus, in the long-run, a repeat seller has a choice of being truthful or silent. Because of the linkage principle, the policy of revealing truthful information about the value of the good for sale dominates the policy of concealing information, because the revelation of information links the payment to the actual outcome.

In contrast, revealing information about the extent of competition may not increase the prices. Consider the case where occasionally there are three bidders, and sometimes only one. If the extent of competition is concealed, bidders will bid without knowing the extent of competition. If the bidders are risk neutral, it turns out that the revelation doesn't matter and the outcomes are the same on average. If, in contrast, bidders are risk averse, the concealment of information tends to increase the bid prices, because the risk created by the uncertainty about the extent of competition works to the advantage of the seller. Of course, it may be difficult to conceal the extent of competition in the English auction, suggesting a sealed-bid auction instead.

Bidders with a large, known advantage have several deleterious effects. For example, incumbent telephone companies generally are willing to pay more for spectrum in their areas than outsiders are. Advantaged bidders discourage participation of others, since the others are likely to lose. This can result in a bidder with an advantage facing no competition and picking up the good cheaply. Second, rivals don't present much competition to the advantaged bidder, even if the rivals do participate. Consequently, when a bidder has a large advantage over rivals, it is advantageous to handicap the advantaged bidder, favoring the rivals. This handicapping encourages participation and levels the playing field, forcing the advantaged bidder to bid more competitively to win.

A common means of favoring disadvantaged bidders is by the use of bidder credits. For example, with a 20% bidder credit for disadvantaged bidders, a disadvantaged bidder only has to pay 80% of the face amount of the bid. This lets such a bidder bid 25% more (since a $100 payment corresponds to a $125 bid) than they would have otherwise, which makes the bidder a more formidable competitor. Generally, the ideal bidder credit is less than the actual advantage of the advantaged bidder.

Auction design is an exciting development in applied industrial organization, in which economic theory and experience is used to improve the performance of markets. The U.S. Federal Communications Commissions auctions of spectrum, were the first major instance of auction design in an important practical setting, and the auction design was credited with increasing the revenue raised by the government substantially.

7.7 Antitrust

In somewhat archaic language, a trust was a group of firms acting in concert, which is now known as a cartel. The antitrust laws made such trusts illegal, and were intended to protect competition. In the United States, these laws are enforced by the Department of Justice's Antitrust Division, and by the Federal Trade Commission. The United States began passing laws during a time when some European nations were actually passing laws forcing firms to join industry cartels. By and large, however, the rest of the world has since copied the U.S. antitrust laws in one version or another.

7.7.1 Sherman Act

The Sherman Act, passed in 1890, is the first significant piece of antitrust legislation. It has two main requirements:

> *Section 1. Trusts, etc., in restraint of trade illegal; penalty*
> Every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, or with foreign nations, is declared to be illegal. Every person who shall make any contract or engage in any combination or conspiracy hereby declared to be illegal shall be deemed guilty of a felony, and, on conviction thereof, shall be punished by fine not exceeding $10,000,000 if a corporation, or, if any other person, $350,000, or by imprisonment not exceeding three years, or by both said punishments, in the discretion of the court.
>
> *Section 2. Monopolizing trade a felony; penalty*
> Every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of the trade or commerce among the several States, or with foreign nations, shall be deemed guilty of a felony, and, on conviction thereof, shall be punished by fine not exceeding $10,000,000 if a corporation, or, if any other person, $350,000, or by imprisonment not exceeding three years, or by both said punishments, in the discretion of the court.[110]

The phrase "in restraint of trade" is challenging to interpret. Early enforcement of the Sherman Act followed the "Peckham Rule," named for noted Justice Rufus Peckham, which interpreted the Sherman Act to prohibit contracts that reduced output or raised prices, while permitting contracts that would increase output or lower prices.

In one of the most famous antitrust cases ever, the United States sued Standard Oil, which had monopolized the transportation of oil from Pennsylvania to the east coast cities of the United States, in 1911.

The exact meaning of the Sherman Act had not been settled at the time of the Standard Oil case. Indeed, Supreme Court Justice Edward White suggested that, because contracts by their nature set the terms of trade and thus restrain trade to those terms and Section 1 makes contracts restraining trade illegal, one could read the Sherman Act to imply all contracts were illegal. Chief Justice White concluded that, since Congress couldn't have intended to make all contracts illegal, the intent must have been to make unreasonable contracts illegal, and therefore concluded that judicial discretion is necessary in applying the antitrust laws. In addition, Chief Justice White noted that the

[110] The current fines were instituted in 1974; the original fines were $5,000, with a maximum imprisonment of one year. The Sherman Act is 15 U.S.C. § 1.

act makes *monopolizing* illegal, but doesn't make having a monopoly illegal. Thus, Chief Justice White interpreted the act to prohibit certain acts leading to monopoly, but not monopoly itself.

The legality of monopoly was further clarified through a series of cases, starting with the 1945 Alcoa case, in which the United States sued to break up the aluminum monopoly Alcoa. The modern approach involves a *two-part test*. First, does the firm have monopoly power in a market? If not, no monopolization has occurred and there is no issue for the court. Second, if so, did the firm use illegal tactics to extend or maintain that monopoly power? In the language of a later decision, did the firm engage in "the willful acquisition or maintenance of that power as distinguished from growth or development as a consequence of superior product, business acumen or historic accident?" (U.S. v. Grinnell, 1966.)

There are several important points that are widely misunderstood and even misreported in the press. First, the Sherman Act does *not* make having a monopoly illegal. Indeed, three legal ways of obtaining a monopoly – a better product, running a better business, or luck – are spelled out in one decision. It is illegal to leverage that existing monopoly into new products or services, or to engage in anticompetitive tactics to maintain the monopoly. Moreover, you must have monopoly power currently to be found guilty of illegal tactics.

When the Department of Justice sued Microsoft over the incorporation of the browser into the operating system and other acts (including contracts with manufacturers prohibiting the installation of Netscape), the allegation was not that Windows was an illegal monopoly. The DOJ alleged Microsoft was trying to use its Windows monopoly to monopolize another market, the internet browser market. Microsoft's defense was two-fold. First, it claimed not to be a monopoly, citing the 5% share of Apple. (Linux had a negligible share at the time.) Second, it alleged a browser was not a separate market but an integrated product necessary for the functioning of the operating system. This defense follows the standard "two-part test."

Microsoft's defense brings up the question of "what is a monopoly?" The simple answer to this question depends on whether there are good substitutes in the minds of consumers – will they substitute to an alternate product in the event of some bad behavior by the seller? By this test, Microsoft had an operating system monopoly in spite of the fact that there was a rival, because for most consumers, Microsoft could increase the price, tie the browser and MP3 player to the operating system, or even disable Word Perfect, and the consumers would not switch to the competing operating system. However, Microsoft's second defense, that the browser wasn't a separate market, is a much more challenging defense to assess.

The Sherman Act provides criminal penalties, which are commonly applied in price-fixing cases, that is, when groups of firms join together and collude to raise prices. Seven executives of General Electric and Westinghouse, who colluded in the late 1950s to set the prices of electrical turbines, spent several years in jail each, and there was over $100 million in fines. In addition, Archer Daniels Midland executives went to jail after their 1996 conviction for fixing the price of lysine, which approximately doubled the price of this common additive to animal feed. When highway contractors are convicted of bid-rigging, generally the conviction is under the Sherman Act, for monopolizing their market.

7.7.2 Clayton Act

Critics of the Sherman Act, including famous "trust-buster" and President Teddy Roosevelt, felt the ambiguity of the Sherman Act was an impediment to its use and that the United States needed a more detailed law setting out a list of illegal activities. The Clayton Act, 15 U.S.C. §§ 12-27, was passed in 1914 and it adds detail to the Sherman Act. The same year, the FTC Act was passed, creating the Federal Trade Commission, which has authority to enforce the Clayton Act, as well as engage in other consumer protection activities.

The Clayton Act does not have criminal penalties, but does allow for monetary penalties that are three times as large as the damage created by the illegal behavior. Consequently, a firm, motivated by the possibility of obtaining a large damage award, may sue another firm for infringement of the Clayton Act. A plaintiff must be directly harmed to bring such a suit. Thus, customers who paid higher prices, or firms driven out of business by exclusionary practices are permitted to sue under the Clayton Act. When Archer Daniels Midland raised the price of lysine, pork producers who bought lysine would have standing to sue, while final pork consumers who paid higher prices for pork, but who didn't directly buy lysine, would not.

Highlights of the Clayton Act include:

- Section 2, which prohibits price discrimination that would lessen competition,
- Section 3, which prohibits exclusionary practices that lessen competition, such as tying, exclusive dealing and predatory pricing,
- Section 7, which prohibits share acquisition or merger that would lessen competition or create a monopoly

The language "lessen competition" is generally understood to mean that a significant price increase becomes possible – that is, competition has been harmed if the firms in the industry can successfully increase prices.

Section 2 is also known as 'Robinson-Patman' because of a 1936 amendment by that name. It prohibits price discrimination that lessens competition. Thus, price discrimination to final consumers is legal under the Clayton Act; the only way price discrimination can lessen competition is if one charges different prices to different businesses. The logic of the law was articulated in the 1948 Morton Salt decision, which concluded that lower prices to large chain stores gave an advantage to those stores, thus injuring competition in the grocery market. The discounts in that case were not cost-based, and it is permissible to charge different prices based on costs.

Section 3 rules out practices that lessen competition. A manufacturer who also offers service for the goods it sells may be prohibited from favoring its own service organization. Generally manufacturers may not require the use of the manufacturer's own service. Moreover, an automobile manufacturer can't require the use of replacement parts made by the manufacturer, and many car manufacturers have lost lawsuits on this basis. In an entertaining example, Mercedes prohibited Mercedes dealers from buying Bosch parts directly from Bosch, even though Mercedes itself was selling Bosch parts to the dealers. This practice was ruled illegal because the quality of the parts was the same as Mercedes (indeed, identical), so Mercedes' action lessened competition.

Predatory pricing involves pricing below cost in order to drive a rival out of business. It is relatively difficult for a firm to engage in predation, simply because it only makes sense if, once the rival is eliminated, the predatory firm can then increase its prices and recoup the losses incurred. The problem is that once the prices go up, entry becomes attractive; what keeps other potential entrants away? One answer is reputation: a reputation for a willingness to lose money in order to dominate market could deter potential entrants. Like various rare diseases that happen more often on TV than in the real world (e.g. Tourette's syndrome), predatory pricing probably happens more often in textbooks than in the real world.[111]

The Federal Trade Commission also has authority to regulate mergers that would lessen competition. As a practical matter, the Department of Justice and the Federal Trade Commission divide responsibility for evaluating mergers. In addition, other agencies may also have jurisdiction over mergers and business tactics. The Department of Defense has oversight of defense contractors, using a threat of "we're your only customer." The Federal Communications Commission has statutory authority over telephone and television companies. The Federal Reserve Bank has authority over national and most other banks.

Most states have antitrust laws as well, and can challenge mergers that would affect commerce in the state. In addition, attorneys general of many states may join the Department of Justice or the Federal Trade Commission is suing to block a merger or in other antitrust actions, or sue independently. For example, many states joined the Department of Justice in its lawsuit against Microsoft. Forty-two states jointly sued the major record companies over their "minimum advertised prices" policies, which the states argued resulted in higher compact disc prices. The "MAP" case settlement resulted in a modest payment to compact disc purchasers. The Federal Trade Commission had earlier extracted an agreement to stop the practice.

7.7.3 Price-Fixing

Price-fixing, which is called bid-rigging in a bidding context, involves a group of firms agreeing to increase the prices they charge and restrict competition against each other. The most famous example of price-fixing is probably the "Great Electrical Conspiracy" in which GE and Westinghouse (and a smaller firm, Allis-Chalmers and many others) fixed the prices of turbines used for electricity generation. Generally these turbines were the subject of competitive (or in this case not-so-competitive) bidding, and one way that the companies set the prices was to have a designated winner for each bidding situation, and using a price book to provide identical bids by all companies. An amusing

---

[111] Economists have argued that American Tobacco, Standard Oil and A.T. & T. each engaged in predation in their respective industries.

element of the price-fixing scheme was the means by which the companies identified the winner in any given competition: it used the phase of the moon. The phase of the moon determined the winner and each company knew what to bid based on the phase of the moon. Executives from the companies met often to discuss terms of the price-fixing arrangement, and the Department of Justice acquired a great deal of physical evidence in the process of preparing its 1960 case. Seven executives went to jail and hundreds of millions of dollars in fines were paid.

Most convicted price-fixers are from small firms. The turbine conspiracy and the Archer Daniels Midland lysine conspiracy are unusual. (There is evidence that large vitamins manufacturers conspired in fixing the price of vitamins in many nations of the world.) Far more common conspiracies involve highway and street construction firms, electricians, water and sewer construction companies or other "owner operated" businesses. Price-fixing seems most common when owners are also managers and there are a small number of competitors in a given region.

As a theoretical matter, it should be difficult for a large firm to motivate a manager to engage in price-fixing. The problem is that the firm can't write a contract promising the manager extraordinary returns for successfully fixing prices because such a contract itself would be evidence and moreover implicate higher management. Indeed, Archer Daniels Midland executives paid personal fines of $350,000 as well as each serving two years in jail. Thus, it is difficult to offer a substantial portion of the rewards of price-fixing to managers, in exchange for the personal risks the managers would face from engaging in price-fixing. Most of the gains of price-fixing accrue to shareholders of large companies, while large risks and costs fall on executives. In contrast, for smaller businesses in which the owner is the manager, the risks and rewards are borne by the same person, and thus the personal risk more likely to be justified by the personal return.

We developed earlier a simple theory of cooperation, in which the grim trigger strategy was used to induce cooperation. Let us apply that theory to price-fixing. Suppose that there are n firms, and that they share the monopoly profits π_m equally if they collude. If one firm cheats, that firm can obtain the entire monopoly profits until the others react. This is clearly the most the firm could get from cheating. Once the others react, the collusion breaks down and the firms earn zero profits (the competitive level) from then on. The cartel is feasible if $1/n$ of the monopoly profits forever is better than the whole monopoly profits for a short period of time. Thus, cooperation is sustainable if:

$$\frac{\pi_m}{n(1-\delta)} = \frac{\pi_m}{n}(1+\delta+\delta^2+\ldots) \geq \pi_m \,.$$

The left hand side gives the profits from cooperating – the present value of the $1/n$ share of the monopoly profits. In contrast, if a firm chooses to cheat, it can take at most the monopoly profits, but only temporarily. How many firms will this sustain? The inequality simplifies to $n \leq \frac{1}{1-\delta}$. Suppose the annual interest rate is 5% and the reaction time is 1 week – that is, a firm that cheats on the cooperative agreement sustains profits for a week, after which time prices fall to the competitive level. In this case, 1-δ is a week's worth of interest (δ is the value of money received in a week) and therefore $\delta = 0.95^{1/52} = .999014$. According to standard theory, the industry with a week-long reaction time should be able to support cooperation with up to a thousand firms!

There are a large variety of reasons why this theory fails to work very well empirically, including that some people are actually honest and don't break the law, but we will focus on one game-theoretic reason here. The cooperative equilibrium is not the only equilibrium, and there are good reasons to think that full cooperation is unlikely to persist. The problem is the prisoner's dilemma itself: generally the first participant to turn in the conspiracy can avoid jail. Thus, if one member of a cartel is uncertain whether the other members of a price-fixing conspiracy are contacting the Department of Justice, that member may race to the DOJ – the threat of one confession may cause them all to confess in a hurry. A majority of the conspiracies that are prosecuted arise because someone – a member who feels guilty, a disgruntled ex-spouse of a member, or perhaps a member who thinks another member is suffering pangs of conscience – turns them in. Lack of confidence in the other members creates a self-fulfilling prophecy. Moreover, cartel members should lack confidence in the other cartel members who are, after all, criminals.

On average, prosecuted conspiracies were about seven years old when they were caught. Thus, there is about a 15% chance annually of a breakdown of a conspiracy, at least among those that are eventually caught.

7.7.4 Mergers

The U.S. Department of Justice and the Federal Trade Commission share responsibility for evaluating mergers. Firms with more than $50 million in assets are required under the Hart-Scott-Rodino Act to file an intention to merge with the government. The government then has a limited amount of time to either approve the merger or request more information (called a *second request*). Once the firms have complied with the second request, the government again has a limited amount of time before it either approves the merger or sues to block it. The agencies themselves don't stop the merger, but instead sue to block the merger, asking a federal judge to prevent the merger as a violation of one of the antitrust laws. Mergers are distinct from other violations, because they have not yet occurred at the time the lawsuit is brought, so there is no threat of damages or criminal penalties; the only potential penalty imposed on the merging parties is that the proposed merger may be blocked.

Many proposed mergers result in settlements. As part of the settlement associated with GE's purchase of RCA in 1986, a small appliance division was sold to Black & Decker, thereby maintaining competition in the small kitchen appliance market. In the 1999 merger of oil companies Exxon and Mobil, a California refinery, shares in oil pipelines connecting the gulf with the northeast, and thousands of gas stations were sold to other companies. The 1996 merger of Kimberley-Clark and Scott Paper would have resulted in a single company with over 50% of the facial tissue and baby wipes markets, and in both cases divestitures of production capacity and the "Scotties" brand name preserved competition in the markets. Large bank mergers, oil company mergers and other large companies usually present some competitive concerns, and the majority of these cases are solved by divestiture of business units to preserve competition.

A *horizontal merger* is a merger of competitors, such as Exxon and Mobil or two banks located in the same city. In contrast, a *vertical merger* is a merger between an input supplier and input buyer. The attempt by book retailer Barnes and Noble to purchase the intermediary Ingram, a company that buys books from publishers and sells to retailers but didn't directly sell to the public, would have resulted in a vertical merger. Similarly, Disney is a company that sells programs to television stations (among other activities), so its purchase of TV network ABC was a vertical merger. The AOL--Time Warner merger involved several vertical relationships. For example, Time Warner is a large cable company, and cable represents a way for AOL to offer broadband services. In addition, Time Warner is a content provider, and AOL delivers content to internet subscribers.

Vertical mergers raise two related problems: *foreclosure* and *raising rivals' costs*. *Foreclosure* refers to denying access to necessary inputs. Thus, the AOL--Time Warner merger threatened rivals to AOL internet service (like EarthLink) with an inability to offer broadband services to consumers with Time Warner cable. This potentially injures competition in the internet service market, forcing Time Warner customers to use AOL. In addition, by bundling Time Warner content and AOL internet service, users could be forced to purchase AOL internet service in order to have access to Time Warner content. Both of these threaten foreclosure of rivals, and both were resolved to the government's satisfaction by promises that the merged firm would offer equal access to rivals.

Raising rivals' costs is a softer version of foreclosure. Rather than deny access to content, AOL--Time Warner could instead make the content available under disadvantageous terms. For example, American Airlines developed the Sabre computerized reservation system, which was used by about 40% of travel agents. This system charged airlines, rather than travel agents, for bookings. Consequently, American Airlines had a mechanism for increasing the costs of its rivals, by increasing the price of bookings on the Sabre system. The advantage to American was not just increased revenue of the Sabre system but also the hobbling of airline rivals. Similarly, banks offer free use of their own automated teller machines (ATMs), but charge the customers of other banks. Such charges raise the costs of customers of other banks, thus making other banks less attractive, and hence providing an advantage in the competition for bank customers.

The Department of Justice and the Federal Trade Commission periodically issue horizontal merger guidelines, which set out how mergers will be evaluated. This is a three step procedure for each product that the merging companies have in common.

The procedure starts by identifying product markets. To identify a product market, start with a product or products produced by both companies. Then ask if the merged parties can profitably raise price by a *small but significant and non-transitory increase in price*, also known as a "SSNIP," pronounced 'snip.' A SSNIP is often taken to be a 5% price increase, which must prevail for several years. If the companies can profitably increase price by a SSNIP, then they are judged to have monopoly power and consumers will be directly harmed by the merger. (This is known as a *unilateral* effect, because the merging parties can increase price unilaterally after the merger is consummated.) If they can't increase prices, then an additional product has to be added to the group; generally the best substitute is added. Ask whether a hypothetical monopoly seller of these three products can profitably raise price. If so, an antitrust market has been identified; if not, yet another substitute product must be added. The process stops adding products when enough substitutes have been identified which, if controlled by a hypothetical monopoly, would have their prices significantly increased.

The logic of product market definition is that, if a monopoly wouldn't increase price in a meaningful way, that there is no threat to consumers – any price increase won't be large or won't last. The market is defined by the smallest set of products for which consumers can be harmed. The test is also known as the hypothetical monopoly test.

The second step is to identify a geographic market. The process starts with an area in which both companies sell, and asks if the merged company has an incentive to increase price by a SSNIP. If so, that geographic area is a geographic market. If not, it is because of buyers substituting outside the area to buy cheaply, and the area must be expanded. For example, owning all the gas stations on a corner doesn't let one increase price profitably because an increase in price leads to substitution to stations a few blocks away. If one company owned all the stations in a half mile radius, would it be profitable to increase price? Probably not, as there would still be significant substitution to more distant stations. Suppose, instead, that one owned all the stations for a 15 mile radius. Then an increase in price in the center of the area is not going to be thwarted by too much substitution outside the area, and the likely outcome is that prices would be increased by such a hypothetical monopoly. In this case, a geographic market has been identified. Again, parallel to the product market definition, a geographic market is the smallest area in which competitive concerns would be raised by a hypothetical monopoly. In any smaller area, attempts to increase price are defeated by substitution to sellers outside the area.

The product and geographic markets together are known as a *relevant antitrust market*, relevant for the purposes of analyzing the merger.

The third and last step of the procedure is to identify the level of concentration in each relevant antitrust market. The Hirschman-Herfindahl index, or HHI, is used for this purpose. The HHI is the sum of the squared market shares of the firms in the relevant antitrust market, and is justified because it measures the price – cost margin in the Cournot model. Generally in practice the shares in percent are used, which makes the scale range from 0 to 10,000. For example, if one firm has 40%, one 30%, one 20% and the remaining firm 10%, the HHI is

$$40^2 + 30^2 + 20^2 + 10^2 = 3{,}000.$$

Usually, anything over 1800 is considered "very concentrated," and anything over 1000 is "concentrated."

Suppose firms with shares x and y merge, and nothing in the industry changes besides the combining of those shares. Then the HHI goes up by $(x + y)^2 - x^2 - y^2 = 2xy$. This is referred to as the change in the HHI. The merger guidelines suggest the government will likely challenge mergers with (i) a change of 100 and a concentrated post-merger HHI, or (ii) a change of 50 and a very concentrated post-merger HHI. It is more accurate to understand the merger guidelines to say that the government likely won't challenge unless either (i) or (ii) is met. Even if the post-merger HHI suggests a very concentrated industry, the government is unlikely to challenge is the change in the HHI is less than 50.

Several additional factors affect the government's decision. First, if the firms are already engaging in price discrimination, the government may define quite small geographic markets, and possibly as small as a single customer. Second, if one firm is very small (less than a percent) and the other not too large (less than 35%) the merger may escape scrutiny because the effect on competition is likely small. Third, if one firm is going out of business, the merger may be allowed as a means of keeping the assets in the industry. Such was the case with Greyhound's takeover of Trailways, a merger to monopoly of the only intercity bus companies in the United States.

Antitrust originated in the United States and the United States remains the most vigorous enforcer of antitrust laws. However, the European Union has recently taken a more aggressive antitrust stance and in fact blocked mergers that obtained tentative U.S. approval, such as General Electric and Honeywell.

Antitrust is, in some sense, the applied arm of oligopoly theory. Because real situations are so complex, the application of oligopoly theory to antitrust analysis is often challenging, and we have only scratched the surface of many of the more subtle issues of law and economics in this text. For example, intellectual property, patents and standards all have their own distinct antitrust issues.