CM5: CONSUMER AND PRODUCER SURPLUS (3/26/21)

THE MATERIAL IN SECTIONS 3 AND 9 ARE, IN MY LESS THAN HUMBLE OPINION, VERY IMPORTANT. BUT THIS MATERIAL IS SOMEWHAT SOPHISTICATED. PLEASE READ IT CAREFULLY BUT I WILL NOT TO ASK ANY SOPHISTICATED EXAM QUESTIONS ON SECTIONS 3 AND 9.

SOME, BUT *NOT* ALL, OF WHAT YOU NEED TO KNOW

1. How do economists define MB and MC?

2. Why is the demand curve also the MB and the WTP curve?

3. What is your demand for $2m sports cars or $250,000 diamond rings?

4. How is consumer surplus defined?

5. How is consumer surplus shown on a supply and demand diagram?

6. Why is the supply curve also the MC and the WTA curve?

7. What is assumed about the slope of the supply/MC/WTA curve?

8. What is meant by the term "short run"?

9. How is the firm's operating profit defined?

10. How is producer surplus defined?

11. How is producer surplus shown on a supply and demand diagram?

12. Why do CS and PS exist?

13. What are the gains from trade?

14. How are the gains from trade shown on a supply and demand diagram?

15. Are the gains from trade distributed equally between the buyer and the seller?

16. What is a Pareto optimum?

17. Why is Pareto optimality a weak standard for providing policy advice?

18. What must be true if an equilibrium is efficient?

19. Does Pareto optimality take into account the distribution of income and wealth?

20. Are there many real-world situations that are Pareto optimal?

21. What is the so-called water-diamond paradox?

22. Why is the water-diamond paradox not a paradox?

1. MARGINAL BENEFIT AND MARGINAL COST

1. **Marginal benefit** *(MB) is the change in total benefit (TB) brought about by a unit change in the quantity consumed, that is,* $MB = \Delta TB/\Delta Q$, where the symbol Δ (the Greek capital d, pronounced delta) indicates a change in something (in our class we assume that the change in quantity is one unit).

2. **Marginal cost** *(MC) is the change in total cost (TC) brought about by a unit change in the quantity produced, that is,* $MC = \Delta TC/\Delta Q$.

3. In general a *marginal anything is the change in the dependent variable brought about by a unit change in an independent variable*.[1]

2. THE DEMAND CURVE IS ALSO THE MARGINAL BENEFIT AND WTP CURVE

1. As we saw in CM4 the demand curve can be read as a MB curve. For each Q (say the 100th Q, Q_{100}) there is a given height of the demand curve, which corresponds to the maximum price that we would be willing to pay to purchase and consume that unit. We interpret that price as measuring the dollar amount of the benefit (value/satisfaction) that we obtain from consuming that unit. If P_{100} = $4 then why would we be willing to sacrifice $4 of other goods and services to consume Q_{100}. *Each* P_X *is assumed to be equal to the MB of consuming the corresponding unit of X.* The demand curve is therefore also the MB curve.

2. Economists call the maximum price that you would pay for a specific unit of X

[1] Those of you who have done calculus will note that the marginal "whatever" is simply the instantaneous rate of change of the total "whatever" brought about by an infinitesimal change in "something else". [If y = f(x) then the *total* y is simply y, *average* y is y/x, and *marginal* y is equal to dy/dx.]

your willingness to pay (WTP) for that unit. The *height* of the demand curve/MB curve – the corresponding price for that unit of X – indicates how much you would be willing to pay to consume the marginal unit and so *the demand curve is also a WTP curve*.

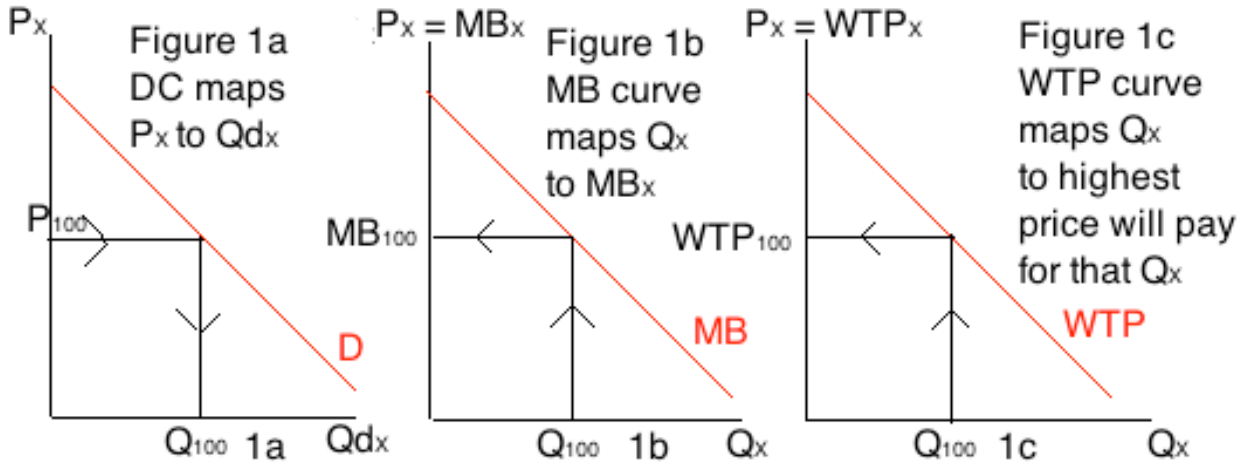
3. THE DEMAND CURVE CAN BE THOUGHT OF IN THREE WAYS

1. THE THREE WAYS

1) A DEMAND CURVE: Starting from the *vertical* axis at, say, P_{100} we can read off from the curve the maximum number of units of X that we would be willing to buy at that price, that is, the quantity demanded (Q^d_{100}). This is the familiar interpretation of the demand curve from CM4. (Figure 1a.)

2) A MARGINAL BENEFIT CURVE: As we have just seen the demand curve is the MB curve. (Figure 1b.) (Reading up from the *horizontal*, Q, axis.)

3) A WILLINGNESS TO PAY CURVE: We have just seen that we willingly give up $4 for the Q_{100} unit and so we say that the demand curve is also a willingness to pay (WTP) curve. We would not buy the 100th unit if the price was $4.01 and so the $4 price is the *maximum*, we will pay for the 100th unit. (Figure 1c.) (Again, reading up from the horizontal axis.) [There is a missing "we" in the legend to 1c.]



2. **Assumption**: *MB declines with increased consumption of X per unit of time*. That is, the first X gives us the *highest marginal gain* in benefit and subsequent increases in consumption yield smaller and smaller gains, MBs; the MB curve slopes down to the right. This is consistent with our everyday experience.

3. DEMAND VERSUS NEED

1. *Economists are only interested in your demand for X, not your need for it or how much you want it. Your demand for X is conditional on your ability to pay for it (ATP), which depends on your income, your wealth, and your ability to obtain credit.* This is very important when thinking about the welfare aspects of markets, what economists call their social optimality. How much is produced of a good or service depends on demand and supply, but demands and supplies are dependent on ATP. Therefore, the socially optimal resource allocations generated by the so called "invisible hand,"[2] are the result of the households' initial endowments of resources; a child born into poverty will have little influence on resource allocation, whereas a child born with a billion dollar trust fund will have much greater influence on what gets produced. Similarly, a child who is well educated and has excellent nutrition and health care will earn more than one who is poorly educated and is born underweight. A woman with the better endowment of human capital will have a larger say in resource allocation than a woman living in poverty because she received an inferior education and is raising three young children on her own. Bill Gates was born into an affluent family in Seattle and attended Lakeside Prep school in the early 1970s where he had access to a mainframe computer – you have probably never seen one. Gates is obviously very intelligent, has a great work ethic, and is very entrepreneurial. (My son emailed Gates about a parking ticket. Gates said fight it. My son hired a lawyer and the case was dismissed.) If Gates had been born to a poor family in a neighborhood he might have ended up as a drug czar!

https://www.thesun.co.uk/news/7404834/donald-trump-investigation-doubts-self-made-man/

2. The concept of a marginal benefit is tied to our ability to pay. I get no marginal benefit from flying first class because although I can afford to do so, I choose not to because the opportunity cost of a first-class seat is too high for me. A *market reflects only those desires that can potentially be consummated by a purchase.* I might get a very large marginal benefit from flying first class but I cannot realize that benefit unless I actually purchase the first-class ticket.

3. Of course, economists have no way of determining my potential marginal benefit if my willingness and ability to pay is less than the supplier's willingness to accept, because then there will be no market purchase. The economist is forced to use price as a measure of value because she cannot see into our brains. And, even if we could somehow determine an individual's subjective valuation of

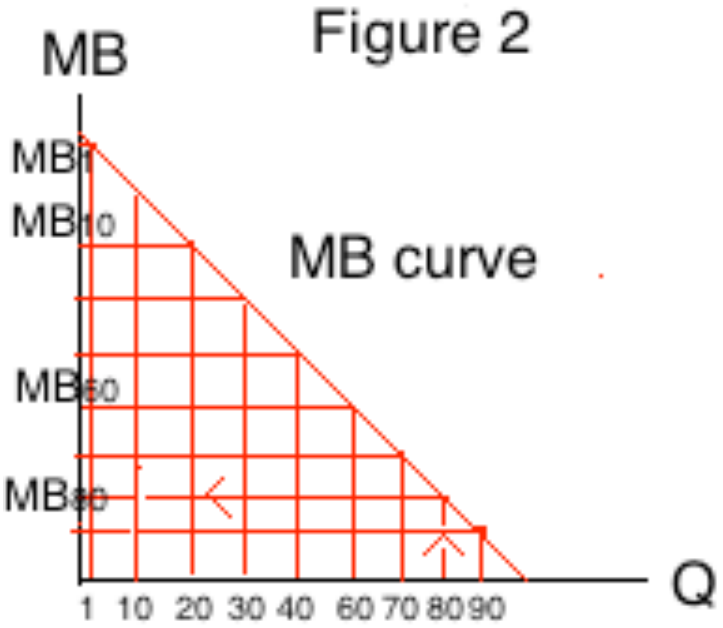
[2] The "invisible hand" is a term found in *The Wealth of Nations.* I will argue in CM13 that the standard interpretation of the "invisible hand" is a misreading of Smith.

something, we have no way of comparing that subjective valuation with that of another person.

4. CONSUMER SURPLUS

1. *The total benefit (TB)* obtained from consuming 100 pieces of dark chocolate, X, *is the sum of the MBs* of the individual pieces consumed: TB = MB_1 + MB_2 + ... + MB_{20} + ... + MB_{66} + ... + MB_{87} + ... + MB_{99} + MB_{100}. (The total cost is the price times 100 – plus the cost of throwing up!) If, we draw a vertical line from Q_1 on the horizontal axis until it reaches the demand curve then the height of that line is the MB of consuming the first square of chocolate. Similarly, the height of the demand curve above Q_{20} is the MB of consuming the 20th chocolate square. The vertical lines in Figure 2 show the MBs of various chocolate squares; the total length of all the lines would be equal to the TB of consuming the 100 squares of chocolate.

Figure 2



2. It is much more convenient to think of the pieces of chocolate (or whatever X stands for) as being infinitely divisible (if we melt the chocolate we can get down to the molecular level) and consumers as being infinitely discriminating. We can then think of the **TB** of consuming 100 units of dark chocolate as the ***area*** under the demand curve up to the Q_{100} unit consumed (Figure 3).

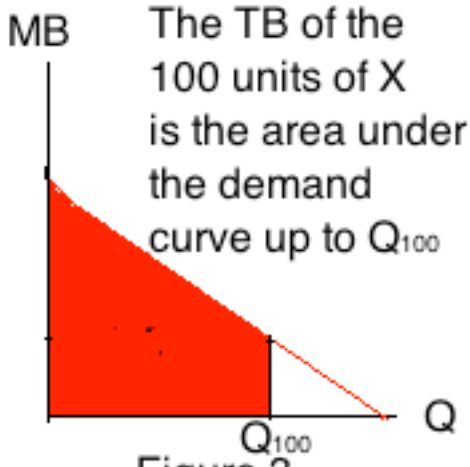


Figure 3

The **total expenditure (TE)** on the 100 pieces of chocolate *is the unit price times quantity (P x Q),* which is $4 x 100 = $400. TE is the black rectangle in Figure 4.

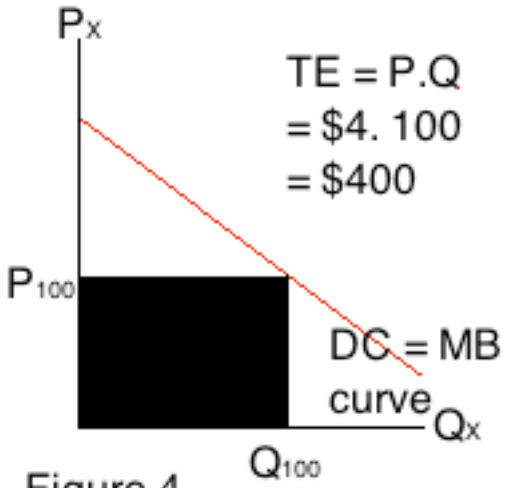


Figure 4

3. If we compare TE with TB then clearly TB is larger. It has to be because TE is determined by the price of the last/marginal, least beneficial unit bought and sold. ***Price is determined "at the margin"*** *– by the benefit of the last unit consumed and the cost to the firm of producing that last unit.* But, *by assumption, each unit of X sells for the same price*. The MB of the last unit purchased determines P^e but the last unit gives the lowest MB.

Consumer Surplus (CS): CS = TB – TE: The difference between TB and TE is *the triangular area* ***beneath*** *the demand curve and* ***above*** *the* P^e. (see Figure 5.)

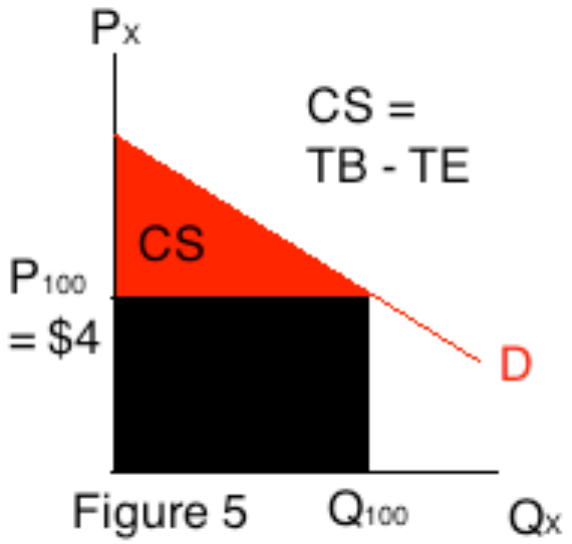


Figure 5

4. *Consumer surplus is the difference between the TB and the TE associated with consuming some given number of units of X. Consumer surplus is measured as the area* ***under*** *the demand curve and* ***above*** *the equilibrium price.* The existence of a consumer surplus is the consequence of the fact that *in a perfectly competitive market all units sell for the same price* (they are identical to one another) and *the price at which they are sold is the price set at the margin*; the price of the last and *least valuable* unit. The consumer receives a marginal benefit that is greater than the price paid for all of the other 99 units. The gain is largest on the first unit purchased and much less on the 99th unit, but each unit except the last one gives the consumer a surplus.

6. THE SUPPLY CURVE CAN BE THOUGHT OF IN THREE WAYS

1. THE THREE WAYS

1) A SUPPLY CURVE: Starting from the *vertical* axis at, say, P_{100}, we can read off from the curve the maximum number of units of X that a profit maximizing firm would be willing to sell at that price, that is, the quantity supplied (Q^s_{100}). This is the familiar interpretation of the supply curve from CM4.

2) A MARGINAL COST CURVE: Starting on the *horizontal* axis at, say, Q_{100}, we can move vertically to the supply curve. The firm will only produce the Q_{100} unit if the price it receives covers its marginal cost of producing the Q_{100} unit (the height of the curve at Q_{100}). The marginal cost of producing the Q_{100} unit is $4. The $4 is the value of the other goods and services that society must give up to obtain the Q_{100} unit. There is nothing special about the 100th unit. The argument

is perfectly general; therefore, the supply curve for X is also the MC curve for producing X.

3) A WILLINGNESS TO ACCCEPT CURVE: A profit-maximizing firm will only produce the Q_{100} unit if the price covers its MC, which is $4. So, the height of the supply curve is the minimum payment that the firm must receive if it is to produce the Q_{100} unit. We call that price the firm's willingness to accept (WTA). The supply curve is also the firm's WTA curve.

Assumption: *The MC curve has a positive slope in the short run.*[3]

7. PRODUCER SURPLUS

1. We know that the supply curve is the firm's marginal cost (MC) curve and we have assumed that MC is positively sloped; MC increases as Q increases.[4] At each Q (say Q_{100}) we can use the supply curve to determine the minimum number of dollars the producer must receive if she is to produce that Q_{100} unit, which is P_{100} = $4. In order to persuade the producer to increase her output by another unit we must pay her at least the MC of producing that unit (and competition ensures that we do not have to pay more than the MC), see Figure 6.

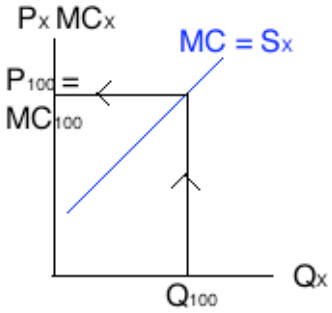


Figure 6

2. We are doing short run analysis, which means that we are assuming that at least one of the inputs (usually capital) in the production process is fixed. *Fixed means: does not vary with output, even if that output is zero.*

3. Variable costs (VC) are the costs of labor, energy and raw materials all of which usually vary as output varies. In the short run the firm maximizes its

[3] The short run is a period of time in which the firm is unable to increase its capital stock.
[4] The MC curve's slope increases at an increasing rate (because of diminishing returns in the short run), but, *for convenience*, we will draw the supply curves/MC curves as straight lines.

operating profit (π_o) by subtracting its variable costs (VC) from its total revenues (TR): π_o = TR – VC. The VC of producing 100 units of X is the sum of the MCs of producing each of the 100 units, *adding* the MCs from the first unit to the one hundredth unit. The MC of the last unit produced (MC_{100}) is the cost that the firm must bear if it increases output from 99 units to 100 units. MC = $\Delta TC/\Delta Q$ but, by assumption, FC does not change when output changes and so MC = $\Delta VC/\Delta Q$; MC does not depend on fixed costs; MC depends only on variable costs.

4. Because the MC curve is upward sloping, *the cost of all units prior to the 100th must be lower than the cost of the 100th unit*. The VC (also referred to as the operating cost) of producing the 100 units is the sum of the MCs of each of those units, i.e. VC = $MC_1 + MC_2 + ... + MC_{99} + MC_{100}$ (see Figure 7).

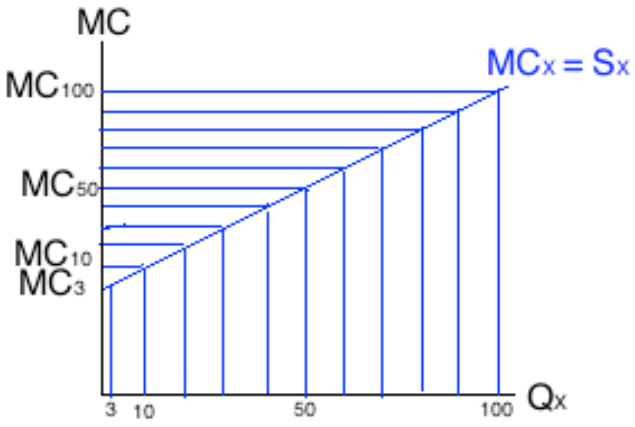


Figure 7

The VC of the 100 units is best thought of as the *area* beneath the MC curve up to Q = 100. (See Figure 8.)

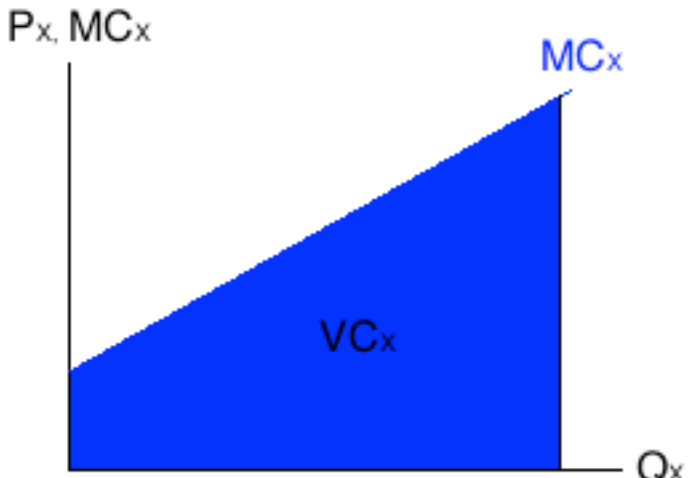


Figure 8

5. *The firm sells each unit for the same price* (P_{100}) and so its TR is P x Q = \$4 x 100 = \$400. In a legal exchange, as opposed to robbery, what is spent on the good or service, Total Expenditure (TE), is the same as what is received from selling it, TE = TR. In Figure 9 TR is equal to the area of the blue plus purple rectangle with height P_{100} = \$4 and width Q = 100, that is the rectangle $0P_{100}AQ_{100}$. If we subtract the VC from TR, we get the firm's operating profit, which corresponds to the purple triangular area *above* the MC curve and *below* P_e. Producer surplus (PS) is equal to the firm's operating profit (see Figure 9).

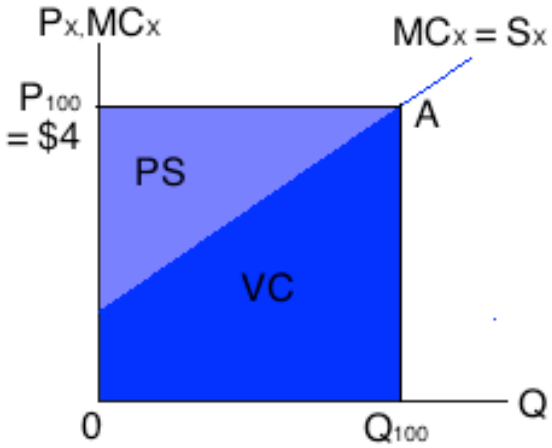


Figure 9

6. *Producer surplus is the difference between the firm's total revenue and its variable cost. Producer surplus is the area* ***above*** *the supply (MC) curve and*

below the equilibrium price. The producer surplus arises because *each unit sells for* the same price, *the price of the marginal* (100th) *unit*. But the last unit is the most expensive unit to produce because, by assumption, marginal costs increase with output. Because the market price has to be just sufficient to cover the cost of the last unit produced, the market price will be greater than MC for all of the first 99 units, with the first unit generating the largest surplus and the 99th unit producing the smallest surplus.

8. THE GAINS FROM TRADE

1. In Figure 10 we put everything together in a single diagram. Equilibrium occurs when $Q_s = Q_d$, which is achieved when $P = P_{100} = P_e = \$4$ and when $Q = Q_e = 100$. *The triangle above P_e is the CS and the triangle below P_e is the PS.* If we add them together, we get **the gains from trade (GFT)**, that is: GFT = CS + PS.

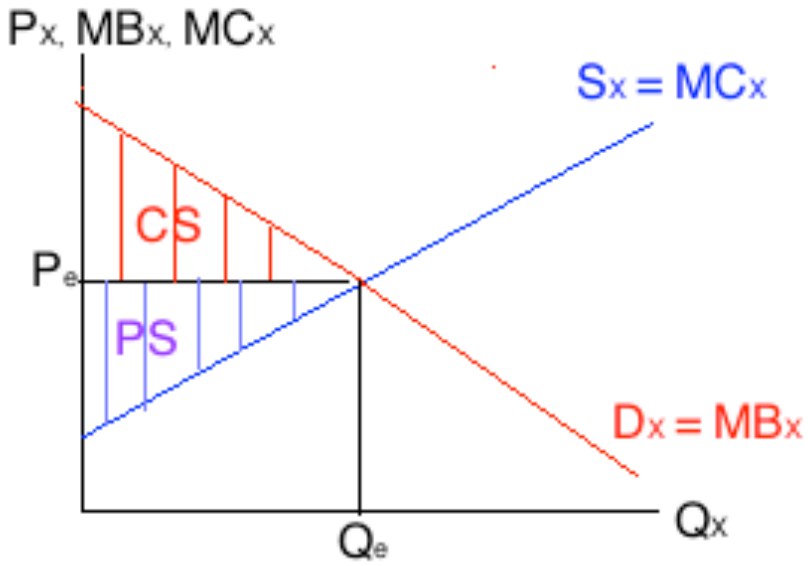


Figure 10

2. Figure 11 illustrates the fact that there is no reason to expect that CS is equal to PS; **trade is mutually beneficial but not necessarily equally beneficial.**

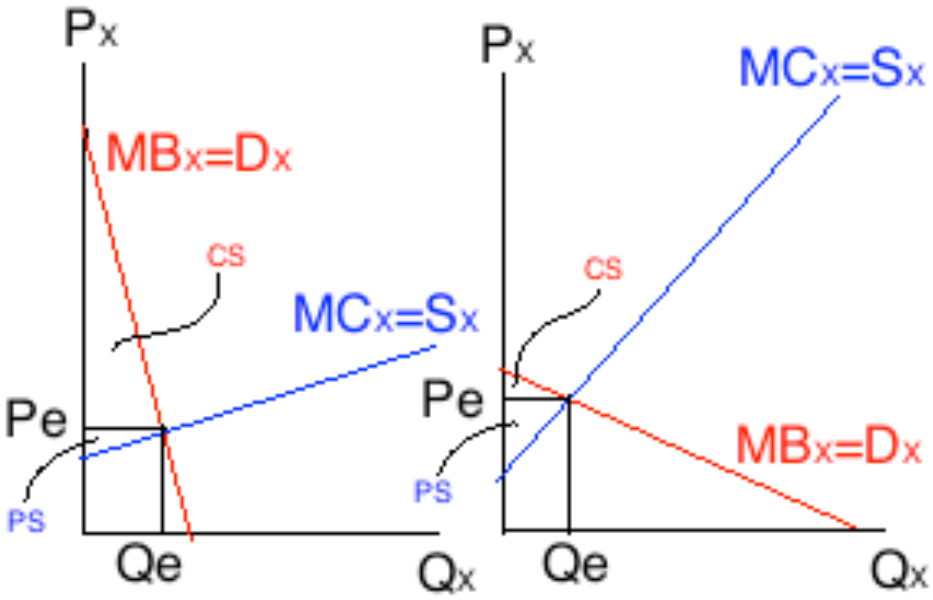


Figure 11

3. The crucial idea is that **price is determined "at the margin"; it is the marginal benefit (MB) and the marginal cost (MC) of the *last* unit bought or sold that determines the price of X, *not* the total benefit (TB) derived from consuming the total number of units consumed of X *nor* the total cost (TC) of producing them.** In a *perfectly competitive market,* all units of X are identical and perfectly interchangeable and therefore *each unit of X sells for the same price*. If we pay $4 for the last unit of X we buy, then we pay $4 for every other unit of X, including the first unit. We are not forced to buy the first unit of X for a higher price than the second unit of X, etc., even though the first unit of X gives us more satisfaction than the second does, and this applies equally to the 100^{th} unit and to the 1000^{th} unit.

9. PARETO OPTIMA

1. The market equilibrium is what economists call a Pareto optimum. ***A Pareto optimum is defined to be an economic situation in which no one can be made better off without making at least one person worse off.*** If we produce less than 100 units then we lose part of the potential CS and part of the potential PS, and if we produce more than 100 units then we have a situation in which MCs are greater than P^e, and MBs are less then P^e and so the GFT are reduced.

In Figure 12 if the market produces 30 units rather than the 100 units that would be produced in equilibrium, then consumers lose all the surpluses they would

have gained on the 31st through the 99th units (the lost CS is equal to the area of the triangle labeled A). Similarly, producers lose the surpluses they would have made on the 31st through the 99th units (the lost PS is equal to the area of the triangle labeled G). If, on the other hand, the market over produces at Q=150, then consumers lose because they are paying the price P_e for units 101 through 150 but those units have MBs that are less then P_e (this is the area of the triangle labeled F). At an output of 150 units firms are selling their product at a price P_e that does not cover the MC of producing those units from the 101st to 150th. (The lost PS is equal to the area labeled H.)

Remember that *the price is given even if the market is not producing at the equilibrium level*.

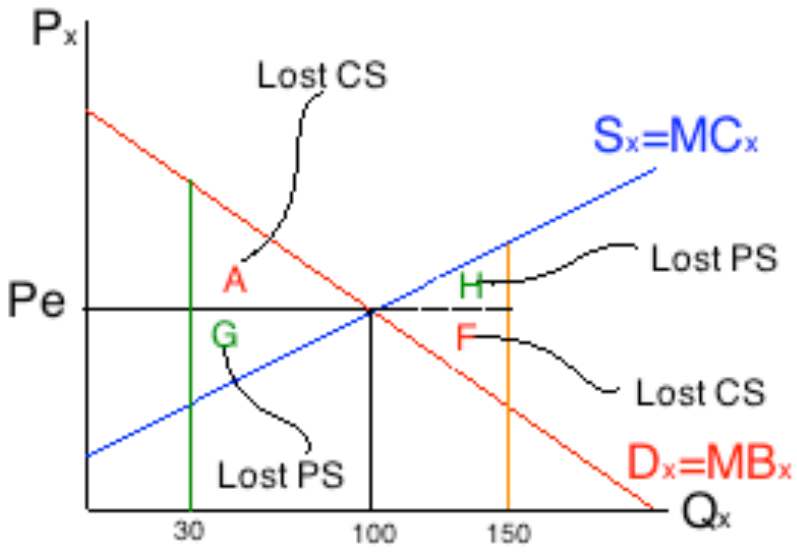


Figure 12

2. Economists say that the *market equilibrium is efficient in the sense that the benefit to society from producing the last unit of the good or service (MB) is just equal to the cost to society of producing that last unit of output (MC) – assuming* that the marginal benefits and marginal costs are also marginal *social* benefits and marginal *social* costs, which will not be the case if there are external effects such as pollution, something that we will analyze in CM14.

3. Pareto efficiency is a very weak policy criterion. *Almost all real-world allocations of resources will be Pareto optimal*, even though the allocation is inefficient in the sense that MB does not equal MC. This is because **even if an**

allocation is inefficient, those households and firms who gain from the inefficient allocation will be made worse off if we remove the inefficiency. Monopolists will lose if their monopoly is taken away, and tenants who gain from rent control will lose if rent control is removed.

5. Further, Pareto optimality has a well-known deficiency, although standard courses usually mention it and then rapidly go on to discuss other issues. **Pareto optimality ignores the distribution of income and wealth.** An equilibrium in which one person has ninety-nine percent of income and everyone else (millions of people) share the remaining one percent can be Pareto optimal. Because economists will not make value judgments about people's welfare economics implicitly supports the existing distribution of income and wealth.

10. CONSUMER'S OR CONSUMERS'?

Note that without pointing this out I have switched from statements about individual economic agents (consumer's surplus and producer's surplus), to statements about market behavior (consumers' surplus and producers' surplus). Dropping the apostrophes partly hides what is going on. There are problems with consumer's and producer's surplus even at the individual level, and we can prove that aggregate consumers' and producers' surpluses are not simply scaled up versions of the individual consumer's and producer's surpluses. However, once again, we will follow hallowed tradition, and having looked the problem squarely in the eye, we will continue to use CS and PS to evaluate market allocations because no one has come up with a better alternative.

11. THE WATER-DIAMOND PARADOX

1. The so-called Water-Diamond "Paradox". Adam Smith observed that although water is necessary to keep us alive and therefore has great intrinsic value to us, it has a low price, whereas diamonds, which are just pretty bits of carbon have very high prices, although their "value" in some ultimate sense is very low because they are not vital to our existence. If you were dying of thirst then you would choose a paper cup of water over a barrel full of diamonds. Smith's analysis of price determination was largely concerned with costs – the so-called labor theory of value. Because prices are set at the margin diamonds, which are artificially scarce, have a large MB whereas water, which in our society is currently in abundant supply, has a low MB. Prices reflect relative demands and supplies *at the margin* not the relative "values" of the goods and services.

2. Because prices are set at the margin GDP is an underestimate of the welfare gains from economic activity.

3. An overlooked example of the so-called paradox is the high incomes earned by people in finance who tweak assets or shave micro seconds off transaction speeds, neither of which seem to be socially advantageous (my value judgment), as opposed to the low incomes of. (Think about the current pandemic and how little hospital staff are paid.) In Figure 13 the gains from trade in finance are much smaller than the gains from trade in nursing but the nurses receive much smaller wages (prices for their labor) than do the finance workers. This is a commonplace outcome in our society. Workers' pay is determined *at the margin* not by the total benefits they generate, occupations that generate very large benefits to society may be poorly paid because the supply of labor for that occupation is relatively large compared to the demand for that type of labor (think of professors of economics and day care workers.)
https://www.washingtonpost.com/business/2020/04/06/why-do-so-many-essential-workers-get-paid-so-little-heres-what-economists-have-say/?utm_campaign=wp_post_most&utm_medium=email&utm_source=newsletter&wpisrc=nl_most

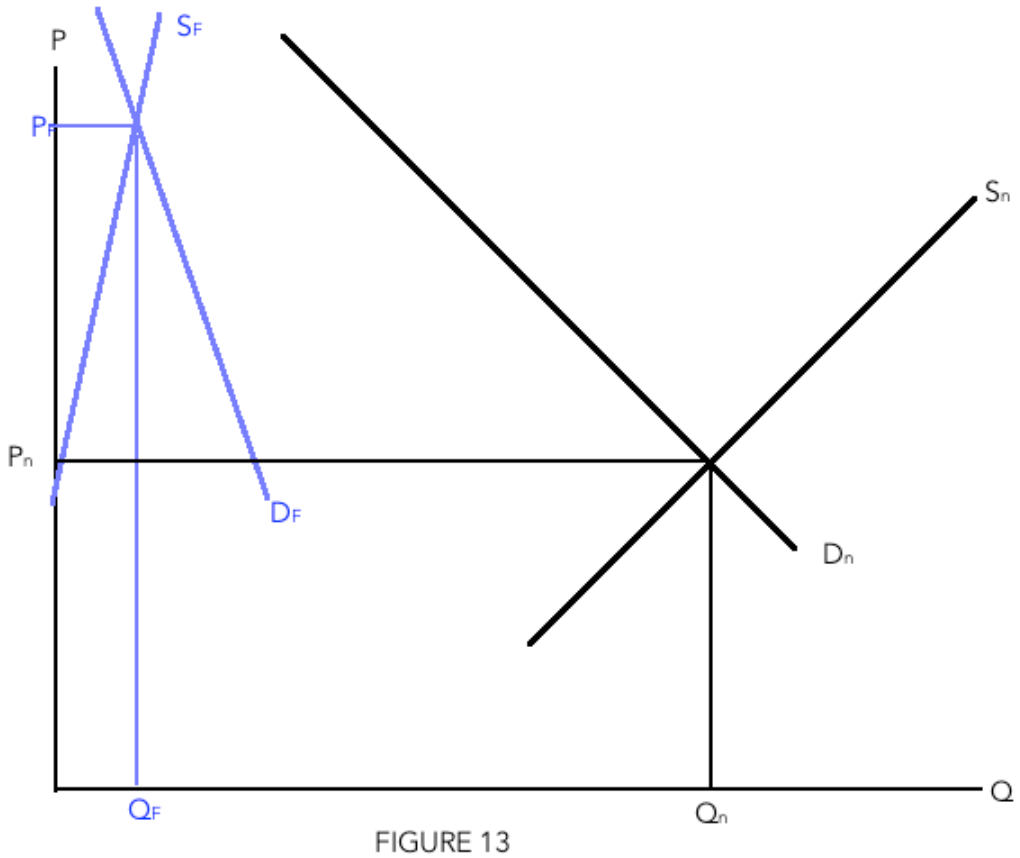


FIGURE 13

(4,037)