Externalities and the Limits of Markets

6

This chapter covers . . .

- the implicit assumptions underlying the assertion that competitive markets are efficient.
- the concepts of interdependency and externality and how they contribute towards understanding the problem of how to organize economic activities and the role of markets.
- the concept of transaction cost and why it is important to not only understand limitations of markets but also the firms and the state as alternative means to organizing economic activities.
- how to apply the concept of transaction costs to understanding how specific markets have to be regulated.
- the relationship between externalities, common goods and public goods, and why these types of goods may justify state interventions beyond property-rights enforcement, contract law, and market regulation.
- a lot about climate change, why status concerns make one unhappy, and the social responsibilities of firms.

6.1 Introduction

> It is not possible to add pesticides to water anywhere without threatening the purity of water everywhere. Seldom if ever does Nature operate in closed and separate compartments, and she has not done so in distributing the earth's water supply. (Rachel Carson, 1962)

The last chapter showed that competitive markets are a very effective way to organize economic activity, because they are able to coordinate the behavior of economic actors in a Pareto-efficient manner, as long as one sticks to the assumption implicit in mainstream economics that true and revealed preferences coincide (see Sect. 5.3). This finding has potentially far-reaching consequences

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Kolmar, *Principles of Microeconomics*, Classroom Companion: Economics,
https://doi.org/10.1007/978-3-030-78167-5_6

for one's perception of the economic role of institutions and, especially, of the state: if competitive markets can alleviate scarcity efficiently, and if efficiency is a convincing normative ideal, then the role of the state is restricted to that of a night watchman. The *night-watchman state* is a metaphor from libertarian political philosophy that refers to a state whose only legitimate function is the enforcement of property rights and contracts and whose only legitimate institutions are, therefore, the military, police, and courts.

This concept of a state, whose monopoly on violence is restricted to the enforcement of property rights and contracts, is sometimes also called a *minimum state*. According to this view, a state that extends its functions beyond this role needs a different normative legitimization by, for example, including distributive objectives. However, even in this case, the Second Theorem of Welfare Economics guides directions of government intervention: distributive objectives can best be achieved by redistributing exogenous endowments and, if they are not subject to redistribution, one should look for the closest substitutes.

The purpose of this chapter is to scrutinize this narrative by extracting the implicit assumptions underlying the claims about competitive markets. The idea is to put the conclusions into perspective in order to allow one to better understand the reasons for the efficiency of competitive markets, as well as their potential and their limitations in the organization of economic activities. In summary, there are three qualitatively different lines of reasoning suggesting that the First Theorem of Welfare Economics cannot be the final say in the debate about the best way to organize economic activities:

- The first line of reasoning has already been mentioned in the last chapter: irrespective of the functioning of the price mechanism, it is unclear whether the revealed-preference paradigm is adequate for all types of goods and services.
- The second point that one has been able to tackle is the relationship between the mode of production in a given industry, summarized by the production technology, and the viable market structures. Not all market structures are compatible with all technological modes of production, but there is a close link between the two. For example, perfect competition requires a specific production technology to be sustainable. In the absence of such production technologies, other market forms might emerge. I only briefly mention this point in this chapter, for completeness, but will dive into the details in Chap. 12.
- Last, but not least, there can be contractual limits to the establishment of markets, and these limits will be the focus for now. As I argued in the last chapter, any exchange of goods and services has two dimensions: a goods- or service-oriented one that focuses on the physical aspects and a legal one that focuses on the transfer of rights. This is why the majority of arguments, which will be developed below, are also relevant in legal contexts and, in fact, *law and economics* as an interdisciplinary field of research evolved along some of the lines demonstrated here.

I will start with some observations that should puzzle one if one looks at them from the perspective of the First Theorem of Welfare Economics.

First, the welfare theorems provide only necessary, but not sufficient, arguments for a night-watchman state, because one has not looked for the efficiency properties of alternate solutions yet. It may be that, under the conditions of the welfare theorems, other organizational structures would also turn out to be efficient. As I will show in later chapters, both monopolistic and oligopolistic markets can turn out to be Pareto-efficient as well, and one does not have any *a priori* reason to assume that centralized planning is not efficient, even though the big historical experiment in centralized planning called socialism can, in all fairness, be called a failure. However, maybe a comparison between "capitalism" and "socialism" is too bold and ideologically charged to allow for a constructive view on institutions. A strong and complete argument in favor of markets has to close this gap.

Second, there is a big theory-reality gap that one has to approach. Assume that the First Theorem of Welfare Economics is a correct characterization of competitive markets, for all types of goods and services. What one should expect, in this case, is a strong tendency of real economies to evolve into the direction of perfect competition, because such an organization would outcompete others. What would such an economy look like? Every transaction would take place in markets and, for example, corporations and other institutional entities would not exist. The assembly of, for example, cars would be organized by a complex chain of bilateral contracts between all the persons who contribute to the manufacturing of the car and the customers. Everyone would act as "You Inc.'s" on atomistic markets without any firms as hierarchical organizations, which replace the market place by a system of hierarchical command and control mechanisms. However, this is not what we observe.

A lot of economic activities are revoked from markets and are, instead, organized according to the different logic of corporations. Basically, what happens if a firm hires a worker is that the worker accepts, within a certain scope, to comply with the instructions of one's principal, which is a hierarchical and not a market interaction. The first step into the corporation is of course a market transaction (signing the job contract), but it is exactly with this contract that one agrees to simply follow the orders of one's boss without further negotiations about prices and so on. A firm can be interpreted as an institution that replaces markets with hierarchies. Yet how could this ever be beneficial, if a market is a reliable instrument for achieving Pareto efficiency? Taking the First Theorem of Welfare Economics at face value, firms should not exist. But they do. Here are two potential reasons why: first, because people are not sufficiently smart to figure out how efficient markets are, so they make mistakes by withdrawing so many transactions from the market place; second, there is something missing in the theory.

One can also turn the question on its head: if one infers, from the existence of firms, that there must be good reasons (in efficiency terms) for their existence, why does one not organize all economic activities within a firm? Why does one organize some transactions with the use of markets? This question has been baptized the "Williamson puzzle" after one of the founding fathers of contract theory and the

theory of the firm, Oliver Williamson. Here is the idea: if one has a set of transactions that are organized on markets, one could just as well organize them under the roof of one big firm. If markets are efficient, the manager leaves everything as it is, so the performance of the firm must be equal to the performance of the market. However, if the market is, for some reason, inefficient, then the manager can correct this inefficiency by a centralized, selective intervention. Hence, the firm should be able to outperform the market. However, if one thinks about it, this one big firm, which is organizing all economic activities under its roof, comes close to a system of centralized planning. Again, this is not what we observe in reality. Thus, again, the puzzle shows that there must be something missing in the theory thus far.

Additionally, to further increase the confusion, why do some firms replace the market mechanism for a set of transactions and then hurry to imitate its functioning in mimicking its mechanisms internally by, for example, the introduction of cost and profit centers, where the inter-center transfer of goods and services is organized by centrally administered transfer prices? The reason for the this is the topic of the next chapter.

6.2 Transaction Costs

There is a plethora of different institutions in modern economies: markets, profit-oriented firms, non-profit organizations, and government agencies. These are all responsible for the mediation of the production and distribution of goods and services, all with their own distinctive logic for providing and distributing goods and services. Any economic theory that aims to understand the reasons for the existence and boundaries of these different ways to organize economic activity must go beyond the First and Second Theorem of Welfare Economics.

Therefore, the challenge is to identify the missing concept that explains institutional diversity. In order to do so, it makes sense to look at the logic of the First Theorem of Welfare Economics from a different perspective. This allows one to reach a deeper understanding of the reasons why markets can be efficient, but also points towards possible explanations for the limitations of markets.

On a very basic level, scarcity implies that individual acts and consequences are interdependent. My decision to drink this glass of wine implies that no one else can drink it. My decision to wear a blue sweater implies (a) that no one else can wear this sweater at the same time and (b) that everyone passing along my way has to see me wearing it. In a world without scarcity, acts would be independent from each other and, therefore, individual goals would not compete with each other. Therefore, what scarcity does is to make individual acts interdependent. As a result, my decisions have repercussions on some other peoples' well-being, and the question is whether I take these consequences into consideration when I make a decision. Efficiency, from this perspective, requires exactly this: that each and every person takes the effects of his or her decisions on others into consideration and behaves accordingly. The technical term is that the person *internalizes* the effects of his or her behavior on others.

However, if I am selfish or ignorant, or both, then I do not care about the effects of my behavior on others. This is the point where markets step in: if I own a car and I consider driving it myself, I am in principle also aware of the fact that I could alternatively sell it on the market. What I am doing in this situation is comparing the monetary value of using the car myself with the market price. If the market price is higher, I want to sell my car; otherwise, I prefer to use it myself.

What does this almost trivial observation have to do with other people? Remember what one has found out about equilibrium prices thus far. The market price in a competitive market is equal to the willingness to pay of the consumer, who is just indifferent between buying and not buying. Thus, prices reflect the willingness to pay of other market participants: my ability to sell the good makes me implicitly internalize the effects that my choices have on others, with the consequence that I only use the good if my willingness to pay exceeds the willingness to pay of other potential users. This is the deeper meaning behind Adam Smith's famous remark on self-interest mediated by the market: "It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages." The self-love of the baker leverages into one's well-being, because one pays him to do so. Prices, in this respect, have two very powerful functions in an economy: they motivate the selfish to care about the effects of their actions upon others, and they also help the benevolent, because prices considerably reduce complexity. The question remains: why does this mechanism of internalization not always prevail?

6.2.1 An Example

Assume a firm produces some good (bread) by means of capital and labor. The capital (oven) is debt-financed and labor (the baker's time) is employed. This economic activity has three effects. First, the bread makes those people eating it better off (it is crispy, tasty, nourishing bread). Second, it ties capital to the specific use, which has opportunity costs in the sense that it cannot be used elsewhere. Third, the baker spends some time baking bread, which also has opportunity costs either in the form of forgone alternative earnings or in the form of forgone leisure time. With competitive markets for capital, labor, and goods, there will be market prices for both inputs and the output. The owner of the bakery has to decide how much bread to bake, how much capital to invest in, and how much labor to hire. The price for bread indicates the social value of an additional loaf of bread, which implies that one correctly internalizes the additional effect on well-being that one creates with the bread. The price for capital (the interest rate) signals the opportunity costs of the next-best use of capital, which implies that the owner correctly internalizes the "damage" that one creates by detracting capital from alternative uses. Additionally, the price for labor (wage) signals the opportunity costs of labor, i.e., the loss in welfare that results because the baker cannot do anything else during the time of

bread baking. This example illustrates not only that decisions are interdependent but also that markets make sure that they are made in an efficiency-enhancing way.

So far, so good, but one still is not at the point where it becomes apparent how markets are *not* efficient. In order to reach this point, I will modify the above example. In the first modification, the production of the product now has sewage as a necessary byproduct, which is dumped into a nearby lake. This solely reduces the profit of a local fisherman. Can one still count on markets doing their magic and leading the economy towards efficiency? The answer is that it depends, and this is where the legal side of the problem enters the picture. There are three possible scenarios:

1. The firm has the legal right to dump sewage.
2. The fisherman has the legal right to prohibit the dumping of sewage.
3. The existence and allocation of rights is unclear.

The first and second scenarios are qualitatively identical to the example before the modification: property rights are completely assigned, which is a prerequisite for bilateral negotiations between the firm and the fisherman. Assume that the reduction of sewage by 10,000 liters reduces the profit of the firm by CHF 1000 and increases the profit of the fisherman by CHF 1500. In this case, there are gains from trade between the fisherman and the firm, and the fisherman can buy "sewage-abatement rights" from the firm, in case the firm owns the rights (scenario 1). Any price for a 10,000-liter reduction between CHF 1000 and CHF 1500 increases the profit of both, the firm and the fisherman, and it is *a priori* not clear why negotiations should not be successful. However, the same holds true if the fisherman is the initial owner of the rights (scenario 2). In this case, the firm can buy "sewage rights" from the fisherman. Assume that an initial increase of sewage from 0 to 10,000 liters increases the profit of the firm by CHF 1500 and decreases the profits of the fisherman by CHF 1000. Again, there are gains from trade between the fisherman and the firm. In the next chapter, we will show that from, the point of view of efficiency, there is no assignment of rights that is more preferable than the other. However, both scenarios lead to different distributions of economic rents, because the owner of the right gets paid. This is no different from the case of, for example, apples: ownership rights, of course, have a value, but they are irrelevant with respect to the efficiency of the resulting allocation.

It is only case three where markets cannot do their magic. If there is no "owner" of the lake, the fisherman and the firm can haggle until eternity without ever reaching a legally binding agreement. Therefore, what one can learn from the example is that markets can only be established if property rights are well defined. These findings motivate the following definition:

▶ **Definition 6.1 Externality** An institution is inefficient if not all interdependencies caused by the individuals are internalized. These non-internalized interdependencies are called *externalities* or *external effects*.

This definition is sufficiently general to include non-market as well as market institutions. In a market context, the institution is, for example, a system of competitive markets and the internalization takes place by means of market prices. A situation where externalities exist in a system of markets is sometimes also called a *market failure*. If the institution is a firm, the internalization could take place by means of internal transfer prices between divisions or by means of wage contracts for employees. It is important to stress that the concept of external effects refers to the institutional framework in which transactions take place; in general, externalities are not properties of goods and services *per se*. However, sometimes specific goods have properties which are likely to produce external effects. We will come back to this later.

The above example has shown that incompletely specified property rights are likely to lead to externalities, because markets cannot emerge. This is an example of what is called *incomplete markets*, and the key question is whether markets are necessarily incomplete, because it is impossible to assign property rights, or if the problem can be fixed by "closing the gaps" and assigning previously unassigned property rights.

The narrative of the example has purposefully been developed around an environmental problem, because many people think that there is something inherent in environmental goods that prevents markets from being efficient. This is a profound misunderstanding, as the example shows. The fact that the interdependency between the fisherman and the baker is caused by sewage is inconsequential for the ability of markets to steer incentives efficiently; what is relevant is the existence and enforcement of property rights and contracts. The same type of problem, as in case 3, would occur if the property rights for bread were not assigned or unclear. If everyone were to enter the bakery and take as much bread as he or she could, the allocation of bread would likely be inefficient since the owner of the shop would lose any incentive to continue production. Thus, why is it that, especially environmental goods are prone to market inefficiencies? There are several reasons, but none of them is causally linked to the "environmental quality" of a good or service. One reason is that, for a long time in human history, a lot of environmental goods have not been scarce. Fresh air and water became only scarce in a lot of areas over the last century. However, without scarcity, it is not necessary to think about efficient uses and there is, therefore, no need to assign property rights to these goods. Thus, part of the problem is a lagging behind of the assignment of property rights when scarcity finally kicks in. For the better part of human history, humans simply did not have the technology to completely deplete fishing grounds, so there was no need to regulate access, and the same goes for other natural resources. However, these problems are relatively easy to solve because, in principle, one can assign rights.

In addition to incompletely assigned property rights, there is another reason why markets may fail. Assume, in the above example, that property rights are completely assigned, so that either the firm or the fisherman have the user rights for the lake. Therefore, in principle, it should be possible to set up a contract that specifies the quantity of sewage the firm is allowed to emit into the lake. The problem may then be that the contracting parties are not able to verify if the other party sticks to the

terms of the contract. There may be emissions by the firm that cause a reduction in the population of fish that is not easily detectable or even impossible to detect. In a situation like this, setting up a contract that specifies emissions may be insufficient to reaching efficiency because the contract cannot be enforced if neither party can verify a potential breach of it (in front of a court, for example).

However, there are other reasons why markets may fail. In order to get to this point, consider a further variation of the example. In this case, the firm no longer produces sewage as a byproduct that impedes with a single fisherman, but pollutes the air with negative consequences for all the residents in the nearby city. Now, one can look into what property rights and markets can do in this example.

1. The firm has the right to emit.
2. The residents have the right to prohibit emissions.

In the first case all the residents have to find an agreement with the firm. However, given that there are a lot of them, reaching such an agreement is likely to be very costly (think of the opportunity costs of time the residents and firm representatives have in reaching an agreement). Thus, it is very likely that negotiations will break down. The same is also true, if the residents hold the rights. Here is a numerical example: assume, as before, that the reduction of pollution by 10,000 liters reduces the profit of the firm by CHF 1000 and has a monetary value for each of the 10,000 residents of CHF 2. So there are huge gains from trade (CHF 20,000 − CHF 1000 = CHF 19,000), but each resident is only willing to negotiate up to the point where his or her opportunity costs of time are smaller than CHF 2, which is if they take no longer than, say, five minutes.

These opportunity costs are an example for a type of costs that turned out to be the key for understanding the economic role of institutions:

► **Definition 6.2 Transaction Costs** Transaction costs are the costs of economic activity that are caused by the institutional framework.

Transaction costs are, therefore, the costs of organizing economic activities, of measuring and policing property rights, of lobbying, or of monitoring performance, to mention a few potential sources.

One can check if the above-mentioned type of opportunity costs prevents successful negotiations. Assume that mutual negotiations take longer than five minutes for each resident. In this case, the potential gains from trade are more than consumed by the transaction costs of negotiations (the transaction costs of five minutes of negotiations are CHF 2 times 10,000 residents = CHF 20,000), so it is very unlikely that negotiations will be successful and, even if they are, they create a negative net value.

The fact that markets will likely lead to inefficient outcomes is not an argument against markets *per se*. The question is if alternative institutions exist that economize on transaction costs. In the above case, the residents could, for example, delegate the

authority to negotiate with the firm to a single representative. Even if some residents have to accept a compromise in the negotiations because of their very specific preferences, this compromise may be better than the externalities that would result from decentralized negotiations. To be more specific, assume that the opportunity costs of reaching an agreement to delegate authority to a representative are CHF 1 per resident, and that the subsequent negotiations between the representative incurs additional opportunity costs of CHF 1000 but reach an efficient agreement. In order to calculate the "net" gains from trade one has to subtract the sum of transaction costs from the gains from trade, i.e., CHF 20,000 − CHF 1000 − CHF 10,000 = CHF 9000. In this case, delegating authority consumes part of the gains from trade, but dominates the decentralized market outcome in terms of welfare, because it reduces the transaction costs. Note that the resulting arrangement can no longer be described as a decentralized market mechanism, but more closely resembles what might be called "representative democracy."

Digression 6.1 (Class Action)
A class action is an element of the U.S. legal system that allows a group to sue another party. It is a way to overcome the collective-action problem that exists, if many people are harmed by the actions of one party. The problem, in cases like these, is often that the small recoveries that can be expected by any individual do not provide an incentive to sue individually, despite the fact that the aggregate recoveries may be very high. Such a situation creates an incentive for parties to take disproportionately high risks, because the likelihood that they will be brought to court in case of harm is inefficiently small without class action. This problem leads to externalities.

Class action is a means to internalize these externalities. This argument has been explicitly used by the United States Court of Appeals. In Mace v. Van Ru Credit Corporation (1997), the court argued that "[t]he policy at the very core of the class action mechanism is to overcome the problem that small recoveries do not provide the incentive for any individual to bring a solo action prosecuting his or her rights. A class action solves this problem by aggregating the relatively paltry potential recoveries into something worth someone's (usually an attorney's) labor."

This point is also stated in the preamble to the Class Action Fairness Act of 2005: "Class-action lawsuits are an important and valuable part of the legal system when they permit the fair and efficient resolution of legitimate claims of numerous parties by allowing the claims to be aggregated into a single action against a defendant that has allegedly caused harm."

Swiss law, on the contrary, does not allow for class action. When the government proposed a new Federal Code of Civil Procedure in 2006, replacing the cantonal codes of civil procedure, it rejected the introduction of class actions. In the message to Parliament on the Swiss Code of Civil

(continued)

Digression 6.1 (continued)
Procedure (Federal Journal 2006, p. 7221) it has been argued that "[it] is alien to European legal thought to allow somebody to exercise rights on the behalf of a large number of people if these do not participate as parties in the action."

6.2.2 Analysis of Externalities on Markets

Externalities in markets can be analyzed using the supply and demand diagram that was introduced in Chap. 4. We have seen that the demand function can be reinterpreted as a function that measures the customers' marginal willingness to pay, and the supply function as a function that measures the producers' marginal willingness to sell. If interdependencies between individuals remain uninternalized, there is a gap between the individual and the social valuations of economic transactions, implying that individual demand and supply do not adequately reflect the social value of the transaction. Take the emissions problem from above as an example and assume that the residents do not figure out ways to organize collective action. In this case, markets are incomplete, because a market for emissions does not come into existence. If one wants to analyze this problem using standard supply and demand diagrams one, therefore, has to focus on the existing market for bread, which is given in Fig. 6.1.

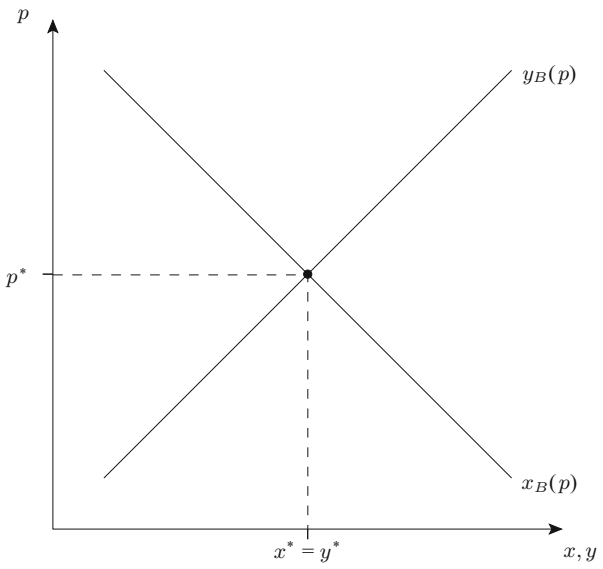

Fig. 6.1 Supply and demand in the bread market

The figure shows the demand function $x_B(p)$ which represents the marginal willingness to pay of the consumers, and the supply function $y_B(p)$ which represents the marginal willingness to sell of the producer. Note, that $y_B(p)$ only represents the *private* costs of production, i.e., it does not reflect the additional costs which emerge due to the sewage that accompanies the production of bread. In this case, the equilibrium quantity is x^* and equilibrium price is p^*. Next, we focus on the case where the baker pays for pollution, i.e., the case where the residents are the owner of the rights. One knows, from the above reasoning, that there must be a difference between the baker's marginal willingness to sell if he does not have to pay for pollution, and the marginal willingness to sell in case he has to pay. If pollution is proportional to the quantity of bread, the supply curve with internalized interdependencies must be *above* the supply curve with uninternalized interdependencies. The former thus represents the *social* costs of production: Making the baker pay for pollution increases his opportunity costs of production, which should influence his marginal willingness to sell any given quantity of bread. Production becomes more expensive, so his marginal willingness to sell should be higher than with uninternalized interdependencies. This situation is given in Fig. 6.2.

Assume that the upward-shifted supply curve has been derived for the hypothetical case of complete markets, where the baker has to pay for pollution. The

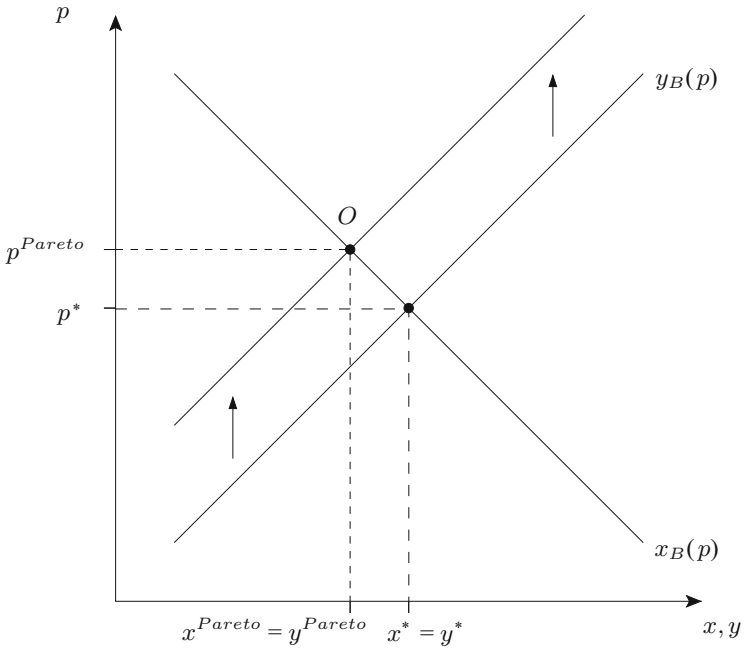

Fig. 6.2 Bread market, if the baker has to pay for pollution or if the fisher has to pay for the omission of pollution

intersection between the demand curve $x_B(p)$ and the "truncated" supply curve $\tilde{y}_B(p)$ (point O in Fig. 6.2) then represents the Pareto-efficient solution. Apparently, uninternalized interdependencies lead to inefficiently high levels of production ($x^* > x^{Pareto}$) at disproportionately cheap prices ($p^* < p^{Pareto}$): too much for too little. A situation like this is also called a *negative externality in production*. But we can do even more than this. By means of Fig. 6.2 we can now quantify the loss in welfare due to the uninternalized interdependencies: for $x \leq x^{Pareto}$ we find that the marginal willingness to pay (represented by $x_B(p)$) is at least as large as the marginal willingness to sell in case of internalized interdependencies (represented by $\tilde{y}_B(p)$), so that gains from trade exists. This no longer holds true for $x > x^{Pareto}$. Here, we find that $\tilde{y}_B(p) > x_B(p)$, i.e., the marginal willingness to sell exceeds the marginal willingness to pay, which in turn destroys welfare. Thus, the grey triangle above the demand function and between x^{Pareto} and x^* represents the loss in welfare due to the uninternalized interdependencies.

Digression 6.2 (Externalities, "Polluter-Pays Principle", and the "Principle of Minimum Harm")
In environmental law, the "polluter-pays principle" makes the party that produces pollution responsible for paying for the damage. It has support from the Organization for Economic Co-operation and Development (OECD) and the European Union, and it seems to make a lot of sense intuitively: in the above example, the baker is responsible for the pollution of the lake, so why not making him pay for cleaning up his mess?

Before one rushes to this conclusion, however, it makes sense to hold on for a second. It is correct to say that the baker causes the pollution, but this does not mean that he also causes the externality. This claim seems odd at first and it is one of the many counterintuitive insights from Ronald Coase to stress that externalities, necessarily, involve more than a single party. The externality exists only because both, the baker and the fisherman, are located on the same lake. If one of them would move away, the externality would cease to exist. In other words, externalities must be treated as a *reciprocal* problem. The polluter-pays principle ignores the fact that externalities are jointly caused by all involved parties: to avoid harm to a pollutee necessarily inflicts harm on the polluter.

If one is still not convinced, because it *is* the baker who *pollutes* the lake, think about a situation where a dynasty of bakers has been living at the lake for generations. Then, from one day to the other, a fisher decides to settle and set up his business. A few days later, he starts complaining about the pollution. Is it still so obvious that the baker causes the *externality*?

The polluter-pays principle is one way to assign rights, because it implies that one party, and not the other has to pay and, with adequately set payments, the externality gets internalized. One has, however, also seen that the same

(continued)

Digression 6.2 (continued)
type of solution can be reached if the baker has the right to pollute and the fisherman pays for reductions in pollution. Such a "pollutee-pays principle" may be at odds with one's intuitions of fairness but, from an efficiency point of view, one has no reason to assume that it is better or worse than the polluter-pays principle. If one sticks with efficiency as a normative principle, it makes sense to replace the principle with a "cheapest cost avoider principle". The idea behind this principle is that it cannot be assumed, in general, that both assignments of rights are equally efficient. With differences in transaction costs, however, it makes sense to assign the rights in a transaction-costs minimizing way.

The above discussion was exclusively concerned with the normative criterion of efficiency, which is an example of an anthropocentric ethic. The reason why the normative problem of externalities vanishes, if the fisherman moves away, is because there is no human being left to be harmed. Environmental ethics like "deep ecology" make the point that such an ethic is too narrow, because the lake, as an ecosystem, still gets harmed and the only way to solve this problem is to reduce pollution. If one includes considerations like this, the polluter-pays principle requires a different interpretation, because it is the only one that respects the integrity of nature. From this perspective, it can be seen as a special case of the more general *principle of minimal harm* or *ahimsa* that is a fundamental moral position of Jainism, Hinduism, and Buddhism. A very popular proponent of the principle of ahimsa was Mahatma Gandhi, and it also shaped Albert Schweitzer's principle of "reverence for life".

Case Study: Fossil Fuels The above analysis was despite its importance rather abstract. So it makes sense to practice a little bit by focussing on an example. The use of fossil fuels creates two main effects: It creates value within the present generation (mobility, power to produce stuff, stuff itself, etc.), and it contributes to the climate crisis which harms our and future generations. We may therefore ask the question if we can expect markets for fossil fuels to be efficient, and in case there is evidence that they are not, what can be done about it.

To tackle this problem we go back to the *willingness to buy* and *willingness to sell* interpretations of the demand and supply functions that we have introduced in Chapter 5. Formally, the marginal willingness to buy function is the inverse of the demand function $x(p)$, and we call it $P(x)$. By the same token, the marginal willingness to sell function is the inverse of the supply function $y(p)$, which is called $Q(y)$.

The potential economic value that can be created for the present generation is summarized by the market demand function for fossil fuels $x(p) = 1100 - p$, $p \geq 0$. We further assume that the market supply function for fossil fuels is equal to $y(p) = -100 + p$, $p \geq 100$. According to scientific consensus, there is an expected

damage that is caused by each unit of fossil fuel used in the present on the well-being of future generations. We denote it by q and assume for illustrative purposes that it is equal to $q = 400$.

The first thing that we have to discuss is whether the intergenerational interdependency measured by q is an externality or not. The answer to this question is straightforward: Future generations cannot participate in today's markets. Therefore, their interests cannot be reflected in today's market prices. The only exception to this rule would be if present generations were perfectly altruistic with respect to the future and take their effects on future well-being into consideration. This may partly be the case, but it seems safe to say that in a large number of cases, this is not what motivates supply and demand decisions. Hence, q is an externality. However, the problem is tricky from a philosophical point of view because future generations do not exist yet, and their existence depends at least partly on today's decisions. We neglect this problem that has spurred discussion in philosophy and simply assume that the mere existence of future generations suffices to argue that they can be harmed by today's decisions. So, a market without further interventions is inefficient.

Next, let us calculate the equilibrium on the market for fossil fuels. As we have seen in Chap. 4, an equilibrium is a price p^* such that $x(p^*) = y(p^*)$. If we insert the given demand and supply functions, we get $1100 - p^* = -100 + p^*$, and the resulting equilibrium price is $p^* = 600$. We can then insert the price into either the supply or the demand function to get $x(600) = y(600) = 500$. We call it *night-watchman equilibrium* because it results without further interventions by the state.

What would happen to either supply of demand if one would internalize the harm imposed on future generations? In order to answer this question we have to determine the Pareto-efficient equilibrium. There are two ways to do so, and we will look at both of them to show that they are in fact equally effective to internalize externalities.

- The first one is to let consumers pay for the externality $q = 400$ per unit of fossil fuel. In this case, the new demand function (from the point of view of the producers) can be determined as follows: First, we have to determine the inverse function, $x = 1100 - p(x) \Leftrightarrow P(x) = 1100 - x$, subtract the additional costs, $p^{EX}(x) = 700 - x$, and get back to the initial function: $x^{EX}(p) = 700 - p$. This demand function internalizes the externality imposed on future generations. We can use it to calculate the new equilibrium $p^{EX} : x^{EX}(p^{EX}) = y(p^{EX}) \Leftrightarrow 700 - p^{EX} = -100 + p^{EX}$. If we solve for p^{EX}, we end up with $p^{EX} = 400$ and $x^{EX}(400) = y^{EX}(400) = 300$.
- The second one is to let producers pay for the externality $q = 400$ per unit of fossil fuel. In this case, the new supply function (from the point of view of consumers) can be determined as follows: First, we need the inverse function, $y = -100 + p(y) \Leftrightarrow p(y) = y + 100$, add the additional costs, $p^{EX}(y) = y + 500$, and get back to the initial function: $y^{EX}(p) = -500 + p$. This supply function internalizes the externality imposed on future generations. We can use it

to calculate the new equilibrium in the same way as before, and we end up with $p^{EX} = 800$ and $x^{EX} = y^{EX} = 300$.

If we compare the night-watchman equilibrium with the equilibria that internalize externalities, we see that the use of fossil fuels is inefficiently large, $dx = 500-300$, and this is exactly at the heart of the problem with most of the activities that contribute to the climate crisis. So the natural next question is, what kinds of interventions into the night-watchman market would allow it to move closer to the Pareto-efficient solution? We will discuss four potential solutions.

- The first solution is not an option here. One could argue that at the heart of the problem is a missing market. But of course we cannot create such a market because future generations cannot participate by definition.
- So, the second solution could be to impose a tax on fossil fuels. This is in principle possible and is also done in practice. A disadvantage of taxes in a situation with a lot of heterogeneity between firms and users is, however, that it does not take these differences into consideration. Therefore, in practice a tax cannot be expected to be fully efficient. In general, taxes have another disadvantage. Taxes that are levied to finance public projects like schools, roads, etc., are a source of inefficiency in themselves because they usually distort individual decisions. This is where taxes that are used to internalize externalities come into play. They do not only reduce an externality, but they can also be used to reduce other, distortionary taxes if total tax revenues are constant. This is called a *double dividend*.
- The third solution is the creation of an artificial market, for example, for the right to emit the pollutants that are responsible for climate change (like CO_2 or methane). In this case, a government agency makes these permits mandatory and sets a total supply for these permits. Firms then have to buy permits if they use resources that issue the respective pollutant. This solution is in principle possible, and a lot of economists see it as the best available alternative because it allows firms to adjust according to their specific circumstances. This flexibility avoids some of the inefficiencies that come with taxes.
- A fourth solution is price regulation or quantity control. In this case, the government intervenes with the existing markets by either setting maximum or minimum prices or restricts the quantities that are allowed to be traded on markets that strongly correlate with the emissions. This instrument is rather bold and usually comes with efficiency costs because it is practice usually not tailored to the individual characteristics of the pollutant. In addition, there is no double dividend that counterbalances these losses in efficiency.

From a purely theoretical point of view it is impossible to rank the different solutions according to their efficiency. The best solution in practice depends on a number of things like the information available to the regulating authority, the strengths of the institutions to credibly implement the measures, the sensitivity of the environmental problem, etc. This ends the case study.

As previously stated, there are always two ways (that may differ in transaction costs) of internalizing interdependencies, depending on which side of the market holds the property rights. Therefore, the alternative in the example above would be to analyze the effects on the bread market, if the residents pay the baker. Is the effect on the bread market identical to the example above or can one expect something different? If the previous analysis is correct, then the assignment of property rights should not influence the efficiency of the solution (without transaction costs), so both scenarios had better yield the same effects on the bread market. In order to check this, assume that the baker gets paid for the reduction in emissions and emissions are again proportional to bread production. For simplicity, assume that one loaf of bread produces one unit of emissions. Let the price of bread be p^b and the price for each unit of omitted emissions be p^e. In this case, an additional loaf of bread has two effects on the baker: it increases his revenues by the market price for bread, p^b, and he reduces his revenues because of the additional emissions by p^e. The total effect on the baker's revenues is therefore $p^b - p^e$, whereas it had been p^b with uninternalized interdependencies. The effect of this change is that the supply curve moves *upwards* as in Fig. 6.2: the only way to convince the baker to sell as much bread as with uninternalized interdependencies is to pay him *more*. This finding verifies the conjecture that, in the absence of transaction costs, it is irrelevant which side of the market pays for the interdependency: it is only relevant that one side does. This insight plays an important role in the literature on law and economics that tries to understand the behavioral consequences of different legal rules.

If there are negative externalities in production, it should not be too surprising that there can also be positive externalities in production, negative externalities in consumption, and positive externalities in consumption:

- **Negative externality in production:** The behavior of an individual causes non-internalized interdependencies, the internalization of which would increase the opportunity costs of production. An example is the above-mentioned problem of uninternalized environmental interdependencies.
- **Positive externality in production:** The behavior of a firm causes uninternalized interdependencies, the internalization of which would reduce the opportunity costs of production. An example is the pollination of fruit trees by bees. The presence of a beekeeper in proximity of a fruit farmer increases the crop of the farmer, because more blossoms get pollinated. If there is no market for "pollination services," the resulting equilibrium is inefficient with too few bees, honey, and fruits. Such a situation can, for example, be analyzed in the market for honey, where the individual's marginal willingness to pay falls short of the social value of honey, because the quantity of honey is positively correlated with the pollination services provided, for which the beekeeper is not paid. (This is an assumption to illustrate how this type of interdependency can be analyzed using supply and demand diagrams if, in fact, the interdependency causes an externality. In practice, farmers and beekeepers are likely to figure out ways to pay for the services). Alternatively, one can focus on the fruit market, in

which the individual's marginal willingness to sell is higher than the efficient one, because the same quantity of fruit is more costly to produce, if there are not enough bees around.

Digression 6.3 (Pollination Services)
The first reaction of a lot of people when they first hear about pollination services is to discard them as a slightly idiosyncratic curiosity, without much economic relevance. The truth is that pollination services are the backbone of agriculture and are also a very important economic factor.

Pollination makes a very significant contribution to the agricultural production of fruits, vegetables, fiber crops, and nuts. Estimates show that pollination services contribute between US $6 and US $14 billion to the US economy per year (Southwick & Southwick, 1992; Morse & Calderone, 2000). The United Nations Environment Programme (UNEP, 2016) estimated that pollination services are worth between US $235 billion and $577 billion globally.

Given the economic importance of pollination services, it should not come as a surprise that commercial pollination services have emerged, mostly provided by honeybees through a long-standing and well-organized market. Californian Almonds are a good example to study the functioning of this market. Almonds are one of the most profitable agricultural products. Recently, honeybee pests and other problems have reduced available bee supplies. At the same time, the high profit margins led to an expansion of almond acreage. Standard supply and demand analysis predicts that this trend—shortage in supply and increase in demand for pollination services—leads to an increase in the price. Figure 6.3 shows that this has, in fact, been the case: the average price per colony almost tripled between 1995 and 2006.

Pollination services are an example of what is called an *ecosystem function*, which is defined as "the capacity of the ecosystem to provide goods and services that satisfy human needs, directly or indirectly" (De Groot, 1992). These services are not only provided by bees, but by a wide variety of insects, birds, and mammals (like bats). A study for the UK found that insect-pollinated crops have become increasingly important in UK crop agriculture and, as of 2007, accounted for 20% of UK cropland value. Bees account for only about 34% of pollination services, down from 70% in 1984 (Breeze et al., 2011). Unlike with bees, it is very difficult to create markets for pollination services provided by other species, which leads to externalities. One of the consequences is that the conservation status of pollinating bird and mammal species is deteriorating.

- **Positive externality in consumption:** The behavior of an individual causes non-internalized interdependencies with other individuals, which increases the value of their consumption. An example is the decision to buy a product that

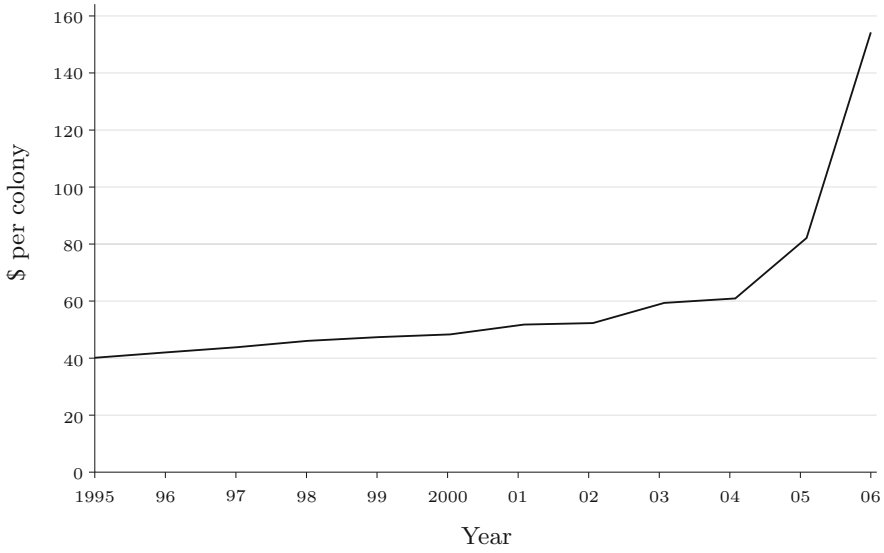

Fig. 6.3 Price level for pollination services (source: Sumner & Boriss, 2006, p. 9)

is interconnected in a network, like a specific type of software or cellphone. The more users coordinate on a given standard, the more valuable the standard becomes for others. For example, the more people use the same text editor, the easier it becomes to exchange text documents. This means that the social value of consumption exceeds the individual value. In other words, the Pareto-efficient demand function lies above the market demand with uninternalized interdependencies. Other examples are individual education decisions that raise individuals' qualifications, but also have an effect on the average literacy of a community, or maintenance of housing that not only increases the value of the individual property but also the attractiveness of the neighborhood.

- **Negative externality in consumption:** The behavior of an individual causes uninternalized interdependencies with other individuals, which reduces the value of their consumption. An example is noise from gardening that annoys the neighbors. In this case, the Pareto-efficient demand function for gardening activities lies below the market demand with uninternalized interdependencies. Another example is vaccination. The World Health Organization (WHO) estimates that vaccination averts 2–3 million deaths per year (in all age groups) and that up to 1.5 million children die each year due to diseases that could have been prevented by vaccination. The individual decision to vaccinate against a pathogen creates a positive interdependency, because it makes the spreading of pathogens more difficult, reducing the risk of other people getting infected. By the same token, the individual decision not to vaccinate creates a negative interdependency. The transaction costs of internalizing these interdependencies on markets are prohibitive, leading to negative externalities in consumption.

Case Study: Vaccination The Corona pandemic is an ideal example that allows to illustrate the concept of interdependencies, transaction costs, and externalities. Among the different aspects, we take vaccinations as an example. The following model can, however, be easily reinterpreted to allow insights into another important phenomenon of a pandemic: that people behave in a way that is inefficiently risky from a societal point of view. We come back to this reinterpretation at the end.

Assume the following simplified model that provides the flavor of the problem. In a society with $n > 1$ individuals, each one can evaluate utilities and costs of their health status as well as their costs of getting vaccinated, and they are all identical in these respects.

- $u_h > 0$: utility (in monetary terms) of being healthy,
- $u_s \geq 0$: utility (in monetary terms) of getting sick with a virus ($u_h > u_s$),
- $c \geq 0$: costs (in monetary terms) of vaccination (direct plus health),
- $1 \geq p \geq 0$: probability of catching the virus without vaccination (probability in case of vaccination is assumed to be 0),
- $r \geq 0$: effective reproduction number (average number of persons that get infected by a person infected with the virus).

We neglect problems that stem from the heterogeneity of individuals (in reality they differ with respect to all of the above variables) and the long-run endogeneity of probabilities (r and p influence but also depend on individual behavior). You can make the model more complicated later; the basic message would stay the same.

We assume that the only choices that the individuals can make is to either vaccinate themselves or not. We will come back to this assumption when we discuss the results. With this information, we can calculate the expected utilities of the individuals for the two options they face, vaccination or no vaccination:

- expected utility (in monetary terms) of an individual without vaccination: $(1 - p)u_h + pu_s$,
- expected utility (in monetary terms) of an individual with vaccination: $u_h - c$.

If people make their decisions in terms of opportunity costs, they get vaccinated if and only if $u_h - c > (1 - p)u_h + pu_s$, and you can simplify this inequality to get $p(u_h - u_s) > c$.

Is this decision rule rational from a societal perspective? In order to be able to answer this question, we have to determine the expected societal utility of an individual without vaccination, which is $(1 - p(1 + r))u_h + p(1 + r)u_s$, and compare it with the expected societal utility of an individual with vaccination, which is $u_h - c$. The difference is the interdependency caused by the individual decision. It results from the additional expected infections that are a consequence of a lack of vaccination.

If you compare both terms, you come to the conclusion that an individual should get a vaccination if and only if $u_h - c > (1 - p(1 + r))u_h + p(1 + r)u_s$, which again

can be simplified to get $p(1+r)(u_h - u_s) > c$. We can now compare the individual with the societal decision rule:

- Observation 1: If it is optimal for the individual to vaccinate, it is also optimal for society: $p(u_h - u_s) > c \rightarrow p(1+r)(u_h - u_s) > c$.
- Observation 2: If it is not optimal for society to vaccinate, it is also not optimal for the individual: $p(1+r)(u_h - u_s) < c \rightarrow p(u_h - u_s) < c$.
- Observation 3: It is possible that it is not optimal for the individual to vaccinate despite the fact that it is optimal for society: $p(u_h - u_s) < c \wedge p(1+r)(u_h - u_s) > c$.

Observation 3 shows that there are parameter values that lead to a negative externality. If one would follow the optimal societal decision rule, every person would profit by $EX = p(1+r)(u_h - u_s) - p(u_h - u_s) = rp(u_h - u_s)$ if an individual gets vaccinated. The negative externality of not getting vaccinated is equal to the non-internalized effect of individual behavior on society which is equal to the weighted (by the expected utility difference between being healthy and sick) effective reproduction number. This externality occurs in cases where the loss of utility from an infection compared to the costs of vaccination is neither very high nor very low.

Why does the underlying interdependency (individual decisions have an impact on the health of other people) turn out to be a negative externality in this case? We did not model transaction costs explicitly, but it is not difficult to see why decentralized solutions cannot be expected to be effective if n is sufficiently large. We have assumed that individuals have only two choices, to get vaccinated or not. It is not surprising that this restriction leads to an inefficiency if people feel no direct, moral obligation to care for others, because with selfish individuals, an internalization of the interdependency requires inter-individual payments that make them internalize the effects of their behavior on others and to nudge them to act efficiently. Hence, up to this point the externality is a result of the modeling strategy not to allow these payments. How realistic and convincing is this strategy? If n would be very small (for example, 2), not very, because it would be relatively straightforward to negotiate between only two individuals. However, if n gets large, the network of required decentralized payments will likely be ineffective because the transaction costs of negotiating them would be too high.

What are the alternatives to internalize the externality? We discuss two policies that require a centralized agency (the state) whose ability to act in a unified and coordinated way reduced transaction costs.

- The agency could subsidize each vaccination by an amount of EX. The advantage is that such a subsidy is sufficient to induce optimal behavior while at the same time making the decision to participate voluntary. The disadvantage is that the agency needs sufficient financial resources to finance the subsidies, which in practice means either higher taxes and/or more public debt.

There is a second disadvantage that is not apparent in this simplified model but that should be discussed to illustrate the limits of simplifying assumptions. If all individuals are identical, one needs exactly one type of subsidy. With heterogenous individuals, the subsidies that are necessary to induce the right type of behavior have to vary between different individuals. If individualized subsidies are not feasible in practice (for example, because the agency has incomplete information), an effective policy of subsidization has to make sure that a sufficient fraction of the population gets vaccinated. Hence, the subsidy is higher than necessary for those individuals who would get vaccinated anyway in order to induce the right kind of behavior in those who are more reluctant. This is called a *windfall gain*, and it makes this policy even more expensive.

- If the agency has sufficient coercive power, it could make vaccinations mandatory. The main advantage of this policy is that it implements the efficient behavior relatively easily without expensive subsidies (however, enforcement costs can be high if people resist). The disadvantage is also easy to identify: people do not like to be forced by the state to do something.

  This reduction in individual freedom does not show up as a direct cost in most economic models, which is why mainstream economics is sometimes criticized. However, it is important in practice not only because people do not like it but also for more systematic reasons. In a society that is built on liberal democratic principles, there is a qualitative difference between incentives (which leaves the ultimate decision with the individual) and compulsion, even if the final result is the same. This is not the place to get more deeply into these legal and political territories. But it should serve as a reminder that economic analysis has to be put into a broader context.

We conclude with the reinterpretation of the formal model that we have mentioned in the beginning. In a time without effective vaccination, the only two major means to prevent the spread of a virus is (1) controlling the spread of the virus by means of masks or social distancing and (1) reducing interactions with other people. (2) can be easily analyzed by a simple reinterpretation of the model, (1) requires a slightly more complicated model.

- Precautionary measures that reduce the spread of a virus are uncomfortable in general. Take masks as an example. In this case, the two options are to either wear one or not, and c measure the disutility from wearing a mask. The one major difference between vaccination and mask wearing is that the latter does not give the same protection as the former, which is why one has to modify probabilities. So, the expected utility of not wearing a mask would be $(1-p)u_h + pu_s$, and the expected utility of wearing a mask would be $(1-p^M)u_h + p^M u_s - c$. $p^M > p$ is the individual probability of staying healthy if one wears a mask. This extension makes the analysis more complicated, but as long as $p^M > p$, the qualitative results stay the same (you can do this as an exercise).
- Social contacts are an important part of what is means to have a good life. Hence, reducing social contacts is costly. In this case, the two options are to reduce

contacts (to zero for simplicity) or not, and c measure the disutility from isolation (in terms of opportunity costs). If contacts are reduced to zero, this measure is as effective as a perfect vaccine, so everything else stays the same.

We can get an important methodological insight from the reinterpretations of the model: The mathematical model is a structure that focusses on certain abstract properties of a problem. The specific interpretation of this abstract structure depends on the specific situation that one wants to analyze. The same mathematical model can fit several different scenarios. This ends the case study.

The literature on externalities is very inconsistent in its terminology, mixing the physical properties of activities, which we call interdependencies, together with the institutional properties, which either lead to an internalization of these interdependencies or do not. It should therefore be stressed again that, in a market context, the term externality relates to missing or imperfect markets. An analysis that makes the *assumption* that an externality exists does not ask for the deeper reasons for the externality and, therefore, risks that one will draw the wrong policy conclusions, which could have been derived from a more thorough analysis. The baker-fisherman problem is a good example. If one starts the analysis with the premise that there is a negative externality between the two, one *assumes* that the two gentlemen cannot figure out ways to fix the problem. However, it would be in their best interest to find a solution, because there are gains from trade. Therefore, one must explore the deeper reasons for this failure and the institutional alternatives. This points one towards a detailed analysis of transaction costs.

Economists can sometimes be blinded by their own theories. It was, for example, a staple in the profession that lighthouse services are a good example for positive externalities in production, because ships cannot be excluded from the insurance provided by the lighthouses. The obvious policy implication from this analysis would be that the state has to step in to provide these services, because markets must fail. However, a more detailed empirical analysis revealed that there are numerous examples for the provision of lighthouse services without government interventions and the key to understanding this "curiosity" was the realization that port owners have an incentive to provide these services to make their ports more attractive. The situation is similar to today's free TV or free services on the Internet. Content providers give away content for free because users allow those firms to make money on other markets, like advertising or data collection.

6.2.3 The Bigger Picture

It is now time to use these examples to develop a more comprehensive view on institutions and transaction costs. The idea that something may be missing in standard theory, which helps explain institutions, goes back to a paper by Ronald Coase that he wrote as early as Coase (1937). Standard theory models firms simply as technological phenomena transforming inputs into outputs, and makes a behavioral assumption that they seek to maximize profits. This "black-box

approach" to the firm had the advantage of simplicity and it allowed for generating a lot of deep insights into the functioning of markets, some of which the last chapter covered. However, the standard approach turned out to be ill-suited to answering the question of why firms exist in the first place, given the apparent efficiency of markets. Ronald Coase's major insight was that transaction costs are at the heart of the problem of optimal institutional design. Unfortunately, transaction costs are a vexed concept, because they turned out to be very difficult to define in a precise and useful way.

Much effort has been devoted to understand the exact conditions under which the invisible hand can leverage self-interest into social welfare and the most useful insight, for this purpose, goes back to another paper by Ronald Coase (1960). If society is interested in promoting efficiency, then every institution that is compatible with this goal must share the same structure: it has to make sure that individuals fully internalize the effects of their behavior on others.

As suggested above, internalization of interdependencies can be achieved by a complete set of competitive markets. The completeness of the markets implies the absence of an important category of market-related transaction costs. The term "transaction costs" is closely related to institutions, since transaction costs can be used to assess the relative "imperfectness" of different institutions (see Definition 6.2). This understanding allows it to put the First Theorem of Welfare Economics into perspective. It was clear from the work of theorists of socialist planning like Oskar Lange (1936; 1937) that, under the conditions of the First Theorem of Welfare Economics, a central planning mechanism is efficient as well. In order to find the equilibrium price, "the market" needs information that, in the hands of a central planner, would be sufficient to implement the efficient allocation directly without the detour of market transactions. This implies that, under ideal circumstances, the institutional structure does not matter for the efficiency of the resulting allocation.

Coase (1960) generalized this idea by creating the awareness that it is neither the complete set of markets nor the idealized planner mechanism that is responsible for the result, but two other, implicit assumptions, namely the rationality of economic actors and the absence of transaction costs.

▶ **Result 6.1 Coase Theorem** In the absence of transaction costs, rational actors will find an agreement that is both Pareto-efficient and independent of the initial allocation of property rights.

The insight is of striking simplicity: if individuals are rational and no transaction costs exist, they should always end up in a situation where gains from trade are completely exhausted; it would simply not be rational to leave them unexploited. In an ideal world, without transaction costs, potential externalities would be fully internalized by rational individuals, whether through market prices, centralized planning in either the form of centrally determined transfer prices or direct quantity control, or other institutional arrangements.

What is the importance of this result? Contrary to what is sometimes argued in the literature, the Coase Theorem is not a result about the efficiency of markets or the advantages of decentralized negotiations. It is rather a methodological critique of models and theories that either im- or explicitly compare the efficiency of different institutions or organizations without making the underlying assumptions regarding transaction costs explicit (or even better explain the specific transaction costs). It is like comparing different architectural designs that have been derived without incorporating gravity. The houses may look beautiful, but it is not clear what will happen if one actually builds them. Transaction costs, in this sense, are like gravity.

The First Theorem of Welfare Economics is a case in point. At the time the underlying theory had been developed, economists have not been aware of the key importance of transaction costs in explaining the comparative efficiency of different institutions. So, the underlying market mechanism is modeled as if transactions were zero. The implication is that from a transaction-cost perspective, the result cannot be used to argue that markets are superior to other institutions as long as the relevant transactions costs are not understood and compared to the transaction costs of alternative institutions. Without transaction costs, rationality alone makes any institution efficient. (This finding does not imply that the market model is useless. On the contrary, it can be extremely valuable in making predictions about all kinds of effects on market prices, the allocation of goods, etc. It can, however, not be used to infer anything useful regarding the comparative efficiency of market- compared to other institutions.)

The implication of this finding is, of course, not that institutions do not matter in reality, but that one has to identify the institution-specific transaction costs, if one wants to understand the relative efficiency of, for example, markets, firms, and government agencies. The transaction-costs-free economy plays the role of the frictionless pendulum in physics: it is not a good description of reality, but a benchmark that allows one to understand the role of friction (or, for that matter, transaction costs) better.

A number of important research areas emerged from this benchmark over the last decades, all of them unified by their attempt to understand transaction costs and their implications for efficiency and the organization of economic activity. The following are some of the most important types of transaction costs:

1. Transaction costs due to the formation of contracts: as shown in the above example, contracts are not just "there" but have to be negotiated, which requires investment of scarce time and effort. Thus, contracts will only be written (and market transactions will only be performed), if the gains from trade exceed the (opportunity) costs of negotiations. Even buying a smoothie requires that one enters the shop, checks the price and pays.

   A very dramatic example of market failure, due to the impossibility to cope with interdependencies by means of contracts, is the interdependencies between generations. Most of one's present decisions are likely to have long-term consequences far beyond one's own planning horizon. However, they will likely affect the well-being of future generations. The most prominent examples

are anthropogenic climate change and nuclear power. In both cases, there surely are intergenerational interdependencies and they cannot be internalized by the use of markets, because one side of the contracting table has not yet been born when the relevant decisions have to be made. Markets must create externalities almost by definition. On the other hand, if markets must fail, what other means does one have to include the interests of future generations? Given that unborn people cannot be part of *any* decision procedure, be it market-based or political, there is only one alternative left: the literal internalization of interdependencies by means of moral concerns of contemporaries. If the present generations are willing to think and act according to the legitimate claims of future generations, then and only then is it possible to internalize the otherwise existing externalities. Even if political decisions to, for example, raise the price of fossil fuels constrain individual behavior, the decision to implement these regulations is not a result of some kind of bargaining between all the affected parties. It is a commitment mechanism by contemporaries that makes it easier for them to follow their moral standards.

Digression 6.4 (Is There Someone to be Harmed? The Non-Identity Problem of Intergenerational Justice)
There is an aspect of the problem of intergenerational justice that makes it different from standard allocation problems between contemporaries. There is a debate in practical philosophy about the normative status of unborn people that focuses on the question, of whether unborn people have the same rights as contemporaries and whether and in what sense contemporaries can harm unborn human beings (Parfit, 1984). One of the key obstacles is the so-called *non-identity problem*, which argues that apparently trivial changes in one's plans are likely to change the identity of the future people (for example, because the egg is fertilized by a different sperm).

Thus, changes in the political environment are likely to have some influence on the identity of future generations but, if this is the case, it cannot be argued that anybody is worse off in the future because one is comparing different people. A pragmatic view would accept this problem as it is and declare the specific identity of a future human being to be morally irrelevant. The only fact that counts, one could argue, is that future generations will come into existence and that they can profit or can be harmed by present generations' choices. Plausible as this approach may sound, it implies a major deviation from standard welfarism, which builds on the idea that the welfare of actual people is normatively relevant.

1. Transaction costs due to the enforcement of contracts: even in a night-watchman state, property rights and contractual arrangements have to be backed by the police and courts. The capital and labor costs of maintaining these agencies must

be considered part of the transaction costs of markets. From an efficiency point of view, the police are only indirectly productive, because its presence creates the necessary environment in which people feel save to invest and trade but, if police were obsolete (for example, because individuals behave cooperatively out of an intrinsic motivation), capital and labor would be freed for other directly productive purposes.

2. Transaction costs due to the incompleteness of contracts: An extensively studied problem is the role that information plays in contract design and in the performance of institutions. There are several strands of literature that I will briefly discuss in turn.
   (a) Asymmetric information: Asymmetric information refers to a situation where one of the contracting parties is aware of information that is relevant for the contract and that of which the other contracting party is not aware. This situation is, of course, the rule rather than the exception, because the parties, in almost any buyer-seller relationship, are unaware of the other party's marginal willingness to pay or sell. Here is an example that highlights the specific problems that may be caused by asymmetric information. Assume a market for used cars, where the sellers are better informed about the quality of the cars than the buyers are. The representative buyer's marginal willingness to pay depends on her assessment of the *average* quality of the car, which implies that the price is not attractive for high-quality sellers. These sellers will withdraw from the market. If the buyers anticipate this incentive, they will further reduce their expectations about average quality and, therefore, their marginal willingness to pay. In the end, the market can completely unravel, leaving only cars of poor quality for sale. George Akerlof (1970), one of the pioneers of information economics, called this type of market a *market for lemons* (a lemon is an American slang term for a car that is found to be defective only after its purchase).

      It turns out that this informational incompleteness is especially relevant on insurance markets and explains why unregulated insurance markets are likely to be inefficient. Specific forms of regulation, like mandatory insurance and obligation to contract (plus some form of price regulation that is necessary to prevent insurance companies from levering out the obligation to contract by charging high prices), reduce these inefficiencies. This kind of regulation works on insurance markets, but not generally on other markets because the standard for efficiency is easy to set. If individuals want to avoid risk (they are *risk averse*), an efficient solution is one where everybody gets full insurance. Such a standard is relatively easy for a government to regulate.
   (b) Non-verifiable contracts: Some contractual arrangements may refer to properties of the good or service that are observable for both contracting parties, but are not verifiable, for example, in front of a court. An example would be a labor contract, where both parties know that the employee is cheating, but the employer is unable to prove it.
   (c) Imperfect foresight: Many contracts expand into the future, which makes the anticipation of future consequences of the contractual arrangement crucial.

> However, in a number of cases, the future cannot be foreseen with sufficient precision to allow for efficient contracts. An example is a different labor contract where a person is hired to conduct research for a company. By definition, the terms of the contract cannot be specified contingent on the outcome of the research project, because it would contradict the nature of research and development. Something completely and qualitatively new may come out of a research project, which makes contracts necessarily incomplete.

From the perspective of transaction-cost economics, climate change is maybe the worst problem someone could have invented to challenge humanity, because it combines a lot of elements that human beings are ill-prepared to solve. First of all, the very nature of intergenerational interdependencies makes it impossible for everyone who is influenced by a decision to participate in a market or any other form of negotiation. Therefore, the only way to incorporate the interests of future generations into today's decision-making is by means of the morality of the present generations. Second, even if one is sufficiently morally motivated to care for future generations, one has imperfect foresight about the future consequences of one's behavior. Third, humanity evolved as a species that had to solve small-group problems for the better part of its history. One's "hard wired" moral instincts are restricted to one's kin and tribe. Problems on a global scale require going beyond one's moral intuitions and caring for the lot of all human beings, not only one's relatives and fellow tribe members. However, reason is a lazy and easily exhausted companion. The executive summary of the so-called Stern Review (2007) makes this point in all clarity: "The scientific evidence is now overwhelming: climate change presents very serious global risks, and it demands an urgent global response. [...] Climate change presents a unique challenge for economics: it is the greatest and widest-ranging market failure ever seen. The economic analysis must therefore be global, deal with long time horizons, have the economics of risk and uncertainty at center stage, and examine the possibility of major, non-marginal change. [...] The effects of our actions now on future changes in the climate have long lead times. What we do now can have only a limited effect on the climate over the next 40 or 50 years. On the other hand what we do in the next 10 or 20 years can have a profound effect on the climate in the second half of this century and in the next. No one can predict the consequences of climate change with complete certainty; but we now know enough to understand the risks. [...] For this to work well, policy must promote sound market signals, overcome market failures and have equity and risk mitigation at its core."

The following subchapter will cover the examples of traffic congestion and environmental problems in order to illustrate how the concept of transaction costs can be used to understand the organization of economic activity better and to design solutions for externality problems.

6.2.3.1 Externalities in Traffic

> A society sufficiently sophisticated to produce the internal combustion engine has not had the sophistication to develop cheap and efficient public transport?
> Yes, boss . . . it's true. There's hardly any buses, the trains are hopelessly underfunded, and hence the entire population is stuck in traffic. (Ben Elton, 1991, Gridlock)

The most common feeling of car drivers who are locked in a traffic jam is anger, but these psychological costs are only the tip of the iceberg regarding economic costs caused by crowded streets and overburdened infrastructure. The main causes of traffic jams are accidents, poor infrastructure, peak-hour traffic, and variable traffic speeds on congested roads. The Centre for Economics and Business Research and INRIX (a company providing Internet services pertaining to road traffic) has estimated the impact of such delays on the British, French, German, and American economies. Here are some of the main findings (US data):

- The costs of congestion summed up to $124 billion in 2013. This cost is (ceteris paribus) expected to increase 50% to $186 billion by 2030. The cumulative cost over the 17-year period is projected to be $2.8 trillion.
- The annual cost of traffic for each American household is $1700 today. This cost is expected to rise to $2300 in 2030, with huge regional variations (the cost is $6000 in the Los Angeles area). To put these numbers into perspective, the median household income was $51,939 in 2013.
- The monetary value of carbon emissions caused by congestion was $300 million in 2013. By 2030, this is expected to rise to $538 million, totaling $7.6 billion over the 17-year period.

Congestion costs of traffic can legitimately count as an externality, because the main causes of these costs are (a) opportunity costs of time, (b) costs of carbon and other emissions, and (c) price effects of higher transportation costs. In order to understand this conjecture, it makes sense to look at a car driver's decision problem. When deciding if, when or where to use streets, she takes individual costs and benefits into consideration. However, the lion's share of costs and benefits spills over onto other traffic participants and the general public. Emissions cause either regional or global effects, which are not included in the individual's decision problem, and other drivers' wasted time is also neglected. The reason is that decentralized negotiations about when and where to use the streets would lead to prohibitive transaction costs.

What else can one do to make traffic more efficient? What are the institutional alternatives? Solving congestion is not easy. Building more roads, or widening existing ones, can encourage people to drive even more. Charging road users for travelling at busy periods can help to solve the efficiency problem, but it may cause other problems. To highlight them, one can focus on the *London Congestion Charge*. The standard charge in 2016 was £11.50 on most motor vehicles operating within the Congestion Charge Zone (Central London) between 07:00 a.m. and 08:00 p.m., Monday through Friday. In theory, the charge should be set such that the individual

driver pays a price that is equal to the costs caused by his decision to use a specific network of streets during a given time period. Hence, if the charge is calculated correctly, one can infer that the externality caused by a single driver is approximately £11.50. If the price of going to central London goes up, demand should go down and one gets the desired increase in efficiency, because congestion is reduced. What makes this instrument problematic is that it has distributional consequences, because the fee is especially burdensome for the relatively poor, who are disproportionately deterred from coming to the city center by car.

Digression 6.5 (The Role of Public Space in Democracy)
Congestion charges or road prices not only have distributive consequences, which one might find objectionable, but also have more profound effects on how one thinks about the societal role of public space. In a democracy, public spaces have an important role in the expression of political opinions, as locations for spontaneous gatherings and, more generally, places where a representative profile of people comes together and has the right to do so. A public space is a site where democracy becomes possible. Henri Lefebvre (1974) made this point quite poignantly: "(Social) space is a (social) product [...] the space thus produced also serves as a tool of thought and of action [...] in addition to being a means of production it is also a means of control, and hence of domination, of power." Charging high prices for the access to public space, which makes it more difficult for specific groups to access them is, therefore, politically questionable. A narrow economic view, which focuses on efficiency gains, easily loses sight of the bigger context in which the instruments are embedded.

A good example for the relationship between democracy and public space is the *Landsgemeinde* (cantonal assembly). This is a Swiss institution where eligible citizens of the canton meet on a certain day in a public space and debate and decide on laws and public expenditures. Another example is the *Speakers' Corner*, an area for unrestricted public speaking, debate and discussion, which became a symbol for the importance of unrestricted access to public space in a democracy. An interesting, yet unresolved, question is whether virtual public space on the internet can take over the role of physical public space, thereby overcoming physical and legal boundaries.

6.2.3.2 Environmental Externalities

> Climate change is a result of the greatest market failure the world has seen. (Nicholas Stern, 2007)

> The metaphor is so obvious. Easter Island isolated in the Pacific Ocean—once the island got into trouble, there was no way they could get free. There was no other people from whom they could get help. In the same way that we on Planet Earth, if we ruin our own [world], we won't be able to get help. (Jared Diamond, 2005)

Oil spills that waste beautiful beaches and wilderness areas are only the tip of the iceberg of environmental externalities. The following are some examples of environmental externalities in production that lead to social costs that are not internalized by market prices. Unregulated air pollution from burning fossil fuels becomes a problem, if no market for pollutants exists. Anthropogenic climate change, as a consequence of greenhouse gas emissions, involves future generations. Negative effects of industrial animal farming include, for example, the overuse of antibiotics that results in bacterial resistance and the contamination of the environment with animal waste. Another problem is the cost of storing nuclear waste from nuclear plants for very long periods of time.

There is a broad consensus among scientists that the rate of species loss is greater now than at any time in human history. In 2007, the German Federal Environment Minister acknowledged that up to 30% of all species would be extinct by 2010. The Living Planet Report (World Wildlife Fund, 2014) comes to the conclusion that "the number of mammals, birds, reptiles, amphibians and fish across the globe is, on average, about half the size it was 40 years ago." If one follows the scientific consensus and assumes that part of the loss in biodiversity is a consequence of the economic system, the question is whether this loss is a result of externalities. Is it possible that mass extinction of species can be Pareto-efficient? This is a tough question, because it requires information about the role of biodiversity in supporting human life on this planet and it relies on assumption about the way humans value biodiversity *per se*. If one starts with the conservative assumption that biodiversity has only instrumental value in supporting human life and if one admits that intergenerational externalities exist, because current generations do not adequately take the interests of future ones into consideration, then one can make a case for the existence of an externality. This is if one assumes that a more diverse biosphere is more likely to support human life than an impoverished one. This latter conjecture, however, is built on deep uncertainty of the complex role of the biosphere in supporting human life. The deeper problem is that the concept of Pareto efficiency, as seen before, is blind with respect to the distribution of gains from trade, and, more generally, economic welfare. A policy where the present generation has a big "party" and uses up most of the natural resources, leaving a devastated planet where future generations scrap along at the subsistence level, is Pareto-efficient as long as there is no alternative policy to make future generations better off without harming the present ones.

The concept of Pareto efficiency has a lot of shortcomings when it comes to long-term problems, which is why is has been supplemented, and even replaced, by the concept of *sustainability* in the normative social and natural sciences and in politics. The most popular definition of the concept of sustainable development goes back to the so-called Brundtland Commission of the United Nations (1987): "sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs." This concept implicitly acknowledges the right of future generations to live a decent life and is, therefore, stronger than the Pareto criterion. However, it still suffers from

the need to understand the complex role of ecosystems and it is anthropocentric in nature. I will come back to this latter point at the end of this subchapter.

Returning to a less complex externality, the example of an oil spill illustrates the basic problems and solutions. Assume that a company operates a fleet of oil tankers, which move large quantities of crude oil from its point of extraction to the refineries. The environmental risk of this business model is that oil spills, due to accidents, affect the (marine) environment and may also affect the fishing industry. One can divide the discussion into two parts. Part one assumes that it is possible to attach a meaningful monetary value to the damage caused by oil spills and to ask for institutional arrangements that lead to efficient outcomes. Part two scrutinizes this assumption and takes a closer look at the normative issues that are involved when attaching price tags to oil spills.

The risk of an accident can be influenced by the shipping companies' investment in safety technology. A profit-oriented company faces a tradeoff between the costs and benefits of such investments and the question is whether it adequately reflects the social costs and benefits when it makes its decisions. In an unregulated market, with only property rights and contract law, this is very unlikely, because many people are potentially influenced by an oil spill, so decentralized negotiations cannot solve the problem efficiently. Therefore, safety standards are presumably inefficiently low in an unregulated market. How can one internalize these externalities? I will discuss three different instruments:

- A very direct and crude way of enforcing safety standards is by setting and enforcing mandatory standards. This instrument is effective, if enforcement is guaranteed, but not necessarily efficient. It becomes the more efficient, the more homogeneous the global fleet is because, in this case, the costs and benefits of a reduction in the risk of accidents are the same for all tankers. Unfortunately this is not the case and, the more heterogeneous the ships are, the less efficient a homogenous regulation will be. One could argue that this is not a problem, as long as regulation can be fine-tuned to the specific characteristics of the tanker, but regulations that are more complicated are more difficult to enact and enforce. Therefore, it is very likely that, in practice, standards would lead to some efficiency losses.
- It is also possible to tax activities that are positively correlated with risks and to offer subsidies for activities that are negatively correlated with risks. Taxes and subsidies change the perceived prices, either making risky activities more costly or making risk-avoiding activities cheaper. The effect is that one creates incentives to influence investments into safety in a socially desirable way. The major advantage of this solution is that, unlike with standardization, this instrument works selectively for different types of tankers and it is, in principle, able to avoid inefficiencies that result from the one-size-fits-all approach of standards. However, a tax-subsidy system has to be administered, which causes transaction costs of its own.
- Last, but not least, one can react with the introduction of liability law. Liability law makes shipping companies pay in case of damage. Liability law increases

the costs of the firms in case of an accident and is, therefore, a theoretically promising instrument for internalizing externalities. When it becomes more costly to have an accident, the company will be more prudent and invest in higher safety standards. However, this legal instrument can conflict with other legal instruments, which have legitimizations of their own. For example, most countries have an insolvency law that restricts the risks of firms and individuals. If such a law is in place, the worst that can happen to a firm is for it to become insolvent, which effectively restricts its monetary risks. Since oil spills are usually big events, liability law can, therefore, be a toothless tiger, if the owners of the company are protected by insolvency regulations.

The above discussion has shown that there are several tools for coping with environmental externalities in the economist's toolbox and it depends on the case at hand which tool (or combination of tools) will work best.

The second aspect of the problem, which one should at least briefly consider, is the question of whether it is possible to attach a price tag to environmental damages. It is relatively uncontroversial that it is possible to get reasonable estimates of damages to the local fishing or tourism industries, because the goods and services they provide have market prices and past experience gives a good proxy for the loss in revenues and profits that result from environmental damages. The question becomes more involved, if one tries to estimate the non-economic costs to human beings that result from the depletion of resources such as air, water, and soil, the destruction of ecosystems and the extinction of wildlife. What is the value of a species of beetle to humankind, which is threatened to become extinct?

However, a radical position would even go beyond the evaluation of non-economic (in a narrow sense) damages and scrutinize the implied anthropocentrism implicit in the normative values underlying Pareto efficiency (or more generally welfarism) and also in the idea of sustainability in the sense of the Brundtland report. According to, for example, the *deep ecology movement*, heavily influenced by the Norwegian philosopher Arne Næs, animals, wildlife, and biosystems have intrinsic value, whereas the mainstream approach is to see them exclusively as means for human ends. The latter approach would deny wildlife a right to existence, if it does not serve any needs of human beings. The deep ecology movement would reject the characterization of non-human life as a means to an end. The core principle is the belief that the living environment, as a whole, should be respected and regarded as having certain inalienable legal rights to live and flourish, independent of its utilitarian instrumental benefits for human use. This has far-reaching consequences for normative economics, which are based on welfaristic ideas about ends and means, good and bad, right and wrong. From the perspective of deep ecology, classifying a meat market as being efficient is completely off the mark, because animals are ends and not means to human needs. A comparison to slave markets is illuminating: trading slaves on markets can be classified as Pareto-efficient, as long as one denies slaves human and civil rights and does not see them as ends, but rather as means for the needs of the class of "non-slaves." Hence, it is a meaningful problem to discuss the efficiency properties of slave markets in such a society. As

soon as one extends basic human and civil rights to all human beings and declares them unalienable, however, there is no meaningful way to discuss the efficiency of such a market, because the traded "resources" are no longer means, but rather ends in themselves. One gets the same fundamental transformation if one grants rights to non-human species.

It would be far beyond the scope of an introductory textbook to dig deeper into the thorny issues of environmental ethics and the consequences for one's perception of economic systems. What the above discussion should have made clear, however, is that our perception of markets relies on normative principles that are—despite their widespread acceptance—far from obvious and innocuous.

There are other, less obvious, ways to cope with externalities and also other, less obvious, sources of externalities in markets. The next two examples focus on business ethics and, especially, the concept of *corporate social responsibility* (CSR), status concerns, and relative-performance measurement as illustrations.

6.2.3.3 Morality and Corporate Social Responsibility

> Globalization makes it clear that social responsibility is required not only of governments, but of companies and individuals. (Attributed to Anna Lindh, 2002)

In the realistic case that the institutional structure of a state is imperfect, in the sense that it does not always provide incentives for (Pareto-)efficient behavior, the question is how the people within society do or should deal with these inefficiencies. An example of this is when property rights are imperfectly enforced because of high transaction costs. The better part of everyday transactions is, for example, formally but not materially protected by property rights, because it would be too costly to enforce them. If a customer buys a bottle of orange juice at a kiosk and the retail clerk refuses to give back the change, the opportunity costs of calling the police, verifying the tort (which is difficult, if the retail clerk refuses to confess), etc. are likely prohibitive. Alternatively, on that subject, it is equally unlikely that the retail clerk can do much to prevent the customer from saying thank you and walking away with the bottle of juice without paying for it. Property rights cannot explain the fact that the overwhelming number of these transactions take place smoothly and efficiently.

There must be other mechanisms at work, and I will briefly discuss two of them. First, the interaction may not be singular but rather repeated and, if there is always a probability that the customer and the retail clerk will meet again in the future, it would be rather shortsighted to sacrifice future trades for the (relatively small) present gain. Repeated interactions can, therefore, be used to build up a reputation as a reliable trading partner, which can stabilize transactions, even in situations where formal property rights cannot be protected by the state. Second, the trading partner may have an intrinsic motivation to play fair. There is broad, scientific consensus by now that individuals are, for good evolutionary reasons, not always selfish, but have the ability and also (sometimes) the desire to act morally. The marginal willingness to keep one's promises, to pay one's bills, etc., however, depends very much on

the perception of the situational context. If people have the feeling that—by and large—society gives everyone his or her fair share, their willingness to cooperate, to act fairly and to voluntarily follow certain moral standards of behavior is much larger than in a situation that is considered unfair from the beginning. Social norms and the intrinsic desire to act morally are then substitutes for formal property-rights enforcement. The more porous the system of property-rights enforcement is, the more important moral behavior becomes.

How relevant is the above observation? Is moral behavior, as the example suggests, only necessary for small-scale transactions, like buying soft drinks at kiosks, or is there more to the story? Here is an example. As one has seen, imperfect and asymmetric information is potentially a major cause of transaction costs. Therefore, in all cases where the better-informed party can exploit the other party, moral behavior can reduce transaction costs and facilitate trade. This view has been nicely expressed by Kenneth Arrow (1971): "In the absence of trust [...] opportunities for mutually beneficial cooperation would have to be forgone [...] norms of social behavior, including ethical and moral codes (may be) [...] reactions of society to compensate for market failures." What could be scrutinized in this quote is the implied supremacy of markets. It is too narrow of a view to see morality only as a repair shop for market failures. However, the general point is irrespectively valid: if specialization, exchange, and trust go hand in hand, it is much easier for a society to flourish.

As one will see below, the existence of public goods, like infrastructure, basic research or defense, is a reason why the state can improve efficiency by playing a role beyond the enforcement of property rights. In order to be able to do so, the state needs to have access to finances, which are primarily collected as taxes. The process of globalization has, however, created opportunities for (multinational) firms and (mainly) wealthy individuals to minimize their tax burden by ever more complicated financial constructions. It may be a good deal for small countries to attract big companies by low tax rates, but the result is a global tax structure and provision of public goods that is inefficient. The point is that the international system of sovereign national states and international tax treaties creates loopholes and leads to discretionary power for firms and wealthy individuals and, despite the OECD initiatives, it is unrealistic to close those loopholes by means of enforceable treaties. As a consequence, one can either accept the resulting inefficiencies or appeal to the moral responsibilities of these firms or people. This is what the former Swedish politician Anna Lindh had in mind in the quote from the beginning of this section and international tax evasion strategies, of course, do not exhaust the number of challenges imposed by globalization.

In the field of Business Ethics, corporate social responsibility (CSR) emerged as a separate field of research, exactly because current trends in international markets led to a redistribution of power from the institutions of the traditional state into the hands of corporations. One of the key questions, in this literature, is whether this increase in power goes hand in hand with the moral responsibilities of the managers and the corporation as an institutional actor.

6.2.3.4 Status

> From whence, then, arises that emulation which runs through all the different ranks of men, and what are the advantages which we propose by that great purpose of human life which we call bettering our condition? To be observed, to be attended to, to be taken notice of with sympathy, complacency, and approbation, are all the advantages which we can propose to derive from it. It is the vanity, not the ease, or the pleasure, which interests us. (Adam Smith, The Theory of Moral Sentiments, Chapter II.)

> Comparison brings about frustration and merely encourages envy, which is called competition. (Jiddu Krishnamurti, 2005)

Social comparisons and the urge to outperform others seem to be deep motivational factors for human beings. Humans are, sometimes, not only a cooperative but also a competitive species and there are good evolutionary reasons for why relative performance is an important factor in mate selection. The human metabolism also requires that some absolute standards are met to stay healthy (for example, daily caloric intake) but, apart from that, relative (status) concerns are an important factor for the explanation of human behavior across cultures and times.

However, this powerful drive has a dark side to it and almost all spiritual traditions from Christianity to Buddhism warn people that the concern for relative status is the road to unhappiness and suffering and that the way towards a fulfilling live is to free oneself from social comparisons. Mark Twain has the, perhaps, shortest account of this fact: "Comparison is the death of joy."

May this be as it is; is there anything that one can say, as an economist, about the functioning of markets, if demand is driven by status concerns? The first observation that one can make is that scarcity works differently for status than for other goods. Say one is eating apples for nourishment. An additional apple makes one better off, irrespective of what the other people are doing. Thus, if everybody in society eats twice as many apples, everybody is better off. This is not true for status goods. If cars are acting as a status symbol (or for that matter, smartphones or Prada shoes) and one buys a bigger car, while no one else in one's neighborhood does the same, then one gains in status and prestige. However, if everyone buys a bigger car, the effects neutralize and one ends up in the same status position as before, when everyone had the smaller car. It is like running a race: if everyone trains harder and runs faster, the odds of winning remain the same but, if one is the only one who "goes the extra mile," then one can tip the balance in one's direction.

What this example shows is that technological progress or an increase in material well-being can alleviate scarcity for ordinary goods (people live healthier, longer lives, are better nourished, etc.), but not for social status. This is why status is called a *positional good*. It is the relative ranking in the pecking order that determines one's position and, if everyone works twice as hard to improve, then no one will be better off in the end. It might even be the case that a point is reached when everybody is worse off, when people start paying tribute for working longer hours. However,

do status concerns create externalities in market economies? To understand this, one can use the extended supply and demand analysis introduced in Sect. 6.2.2. Assume that x measures supply and demand for a status good (mechanical watches), which means that part of the reason for buying this good is to impress the neighbors. Assume further that this interdependency between one and one's neighbor cannot be internalized directly (think about it: "how much would you pay me to not buy that Rolex?"). In that case, the individual value of the status good is, in general, higher than the social value.

One may wonder if *positional externalities* are a mere theoretical curiosity, or if anything more significant is going on. One way to approach this problem is to look for empirical evidence that is anomalous, given the predictions of standard theory (without status concerns), but that can consistently be explained if one accounts for status. In fact, such evidence exists and it became famous as the "happiness paradox." However, it is still highly contested whether the empirical findings are valid.

The happiness paradox refers to patterns in empirical research on happiness. Two findings are key for it. The first relates to the relationship between average subjective happiness and average income. The findings are summarized in Fig. 6.4.

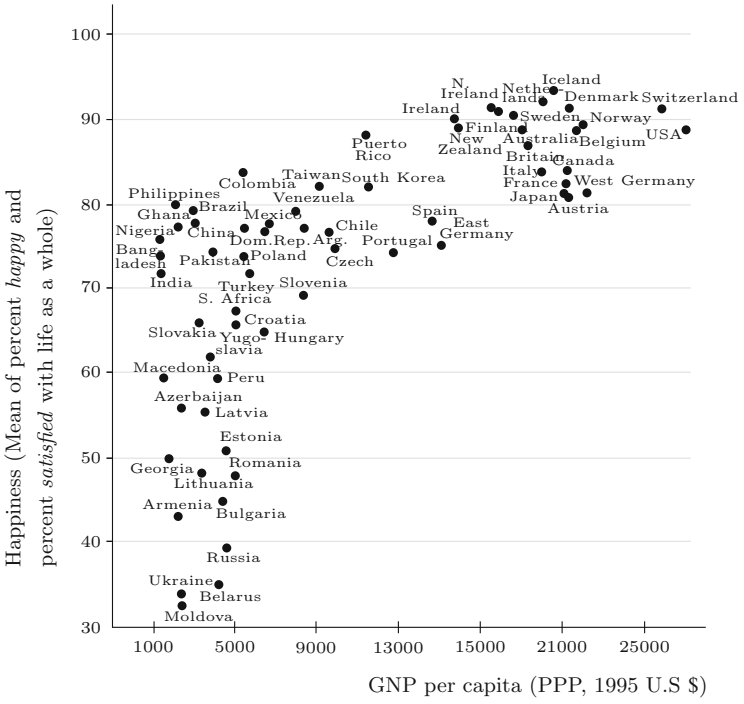

Fig. 6.4 Average subjective happiness and average income (source: Inglehart & Klingemann, 2000, p. 168)

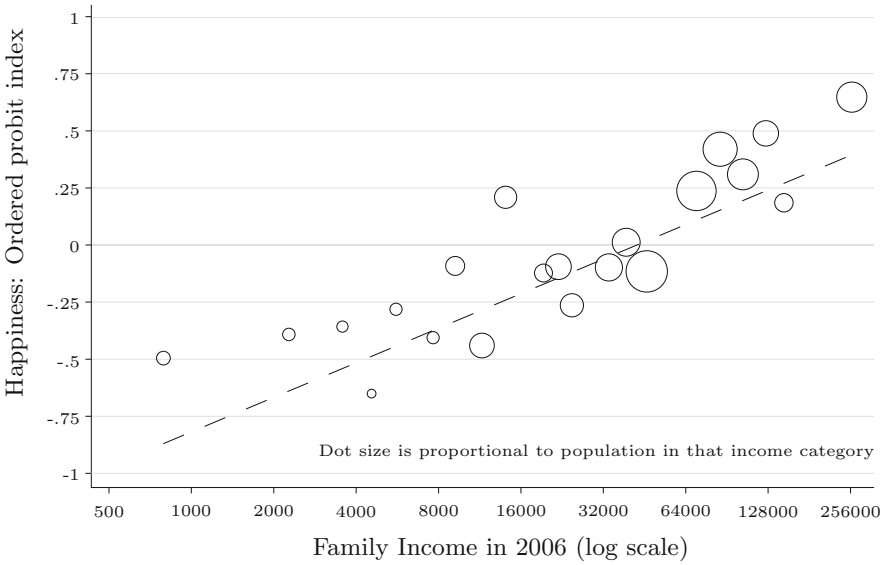

Fig. 6.5 Individual happiness as a function of individual income (source: Stevenson & Wolfers, 2008; General Social Survey, 2006)

The figure depicts *average* happiness levels and *average* income levels in different countries. It shows that there is a positive association between average happiness and average income up to an annual income of about $12,000. However, there is no positive association between average happiness and average income for higher income levels. This "flatness" is sometimes also referred to as the "hedonic treadmill" where one runs faster and faster without moving forward. This finding is difficult to square with the idea that individuals are mutually unconcerned or selfish, because this assumption would imply that increases in material well-being (and average income should be a proxy for this) increase subjective well-being (i.e., happiness). This is apparently not the case. Here is a nice summary of this aspect of the paradox: "People in the West have got no happier in the last 50 years. They have become much richer, they work much less, they have longer holidays, they travel more, they live longer, and they are healthier. But they are no happier." (Layard, 2005)

The second finding refers to *individual* happiness levels as a function of *individual* income within countries, see Fig. 6.5.

The figure reveals that richer individuals are happier than poorer individuals, in a given society, and this is the puzzle: how is it possible that individual income is a good proxy for individual happiness while, at the same time, these effects net out for the whole society as soon as the average income exceeds a certain minimum?

Status preferences provide a missing link for resolving this puzzle: in poor countries, where individuals have to fight for subsistence, the relative importance of non-status-related compared to status-related consumption and production is high.

However, the richer a society gets, the more important status concerns become. Thus, in these societies, there is no longer a positive association between *average* income and *average* happiness, because status effects "net out." If one climbs up on the status ladder, someone else necessarily has to climb it down. However, increases in *individual* income make a difference, because one climbs higher on the status ladder, and the negative happiness effects on others that result from this improvement are irrelevant for one's individual happiness.

As mentioned before, the findings and the interpretation of the happiness paradox are contested. This results from the fact that the exact meaning of the paradox is sometimes misinterpreted. Critics usually start from the premise that the paradox claims that the initially positive correlation between income and subjective happiness disappears above a certain income (often estimated at around $75,000) *at the individual level*. The existence of such an individual kink has been re-examined by Killingsworth in 2021 on the basis of a very large data set and new methods, and he finds no evidence for such a kink. But this finding is without relevance for the paradox, as we have seen.

To attribute the existence of the paradox to a status effect based on relative positioning is in line with a lot of evidence from other fields, like evolutionary biology, where *relative* fitness is key for survival and mating and therefore evolutionary success of ones' genes. Additionally, it comes as no surprise that all the major spiritual traditions humans have created attach large warning signs to individual comparisons. However, even if one takes the interpretation at face value, the policy implications are complicated. Should one infer from the hedonic treadmill that the state has an active role in the internalization of status externalities that is similar to its role in the internalization of, for example, environmental externalities (taxation of status goods, etc.), or should one leave it to the individual to overcome the attachment to status? These are deep questions and they are even more pressing because, as long as social norms declare that social status is a function of material well-being, one straps oneself to the wheel of consumerism and materialism, which is, at least partly, responsible for the environmental externalities mentioned above.

6.3 Four Boundary Cases

> [T]hey devote a very small fraction of time to the consideration of any public object, most of it to the prosecution of their own objects. Meanwhile each fancies that no harm will come to his neglect, that it is the business of somebody else to look after this or that for him; and so, by the same notion being entertained by all separately, the common cause imperceptibly decays. (Thucydides, 2013, The Peloponnesian War, Book 1, Section 141)

Coming back to the variations of the bakery example from the last subchapter, the distinctive difference between the two types of environmental interdependencies (sewage and air pollution) was the physical "reach" of the interdependency-causing activity. In the sewage-case, there was only one person, the fisherman, who was affected by the interdependency with the bakery whereas, in the air-pollution

case, the bakery influenced all the residents. These differences in the number of people, who are influenced by economic activities, are an important element in the classification of goods and services and in developing an understanding of the functioning of markets.

The implicit assumption behind the model of competitive markets discussed in Chap. 4 was that the interdependency is *bilateral*. A typical example for a bilateral interdependency is an apple. Either one or the other person can eat an apple (one cannot eat the same apple twice), so Ann's decision to sell an apple to Bill has no direct physical consequences for third parties. The same was true in the sewage example. However, the bilateralism of the interdependency was a result of the fact that only one fisherman made his living from the lake. If two fishermen had have cast their nets into the lake, the interdependency would have been trilateral, because the emissions by the factory would have reduced the catches of both fishermen. In the air-pollution example, the reach of the interdependency was even larger, covering all residents of the area. This observation motivates the following definition.

▶ **Definition 6.3 Reach** The reach of an economic activity is the set of people directly influenced by the activity.

It is important at this point to say a little bit more about individual motives for consumption. Let us therefore come back to the apple from above. Most people see an apple as food. In this case, it causes a binary interdependency. However, it cannot be excluded that people like apples for aesthetic reasons and simply like to look at them. In this case, the interdependency is no longer necessarily binary, more than two people can profit from the apple. The reason why this example may sound rather awkward is because it is. But it makes an important point: economists usually do not care about motives to act because these motives are hard to measure. However, the motive to act may have an impact on the reach of the act, as one can see from the example, with important economic consequences . So, the reach of an act depends on the specific mental context of the act, not on the physical activity itself.

The two meaningful boundary cases are the minimum and the maximum reach. The use of a good with minimum reach has an effect on only one person, and the use of a good with maximum reach has an effect on all individuals in an economy. A good with minimum reach is called *rival in consumption*, and a good with maximum reach is called *non-rival in consumption*. Combined with scarcity, goods with minimum reach create a bilateral interdependency. One has already seen that an apple is an example for a rival good and it is either person A or B who gets nourished by the apple. An example for a global non-rival good is a fossil fuel combustion increasing CO_2 levels which, in turn, contribute to anthropogenic climate change, which has an impact on all individuals on the planet. Finally, an example for a national, non-rival good is the protection against foreign aggressors due to national defense.

A good part of the goods and services fall in-between these extremes. The reach of national defense, for example, is the boundaries of the nation-state. A live sports event or a music concert has a reach that is limited to the visitors of the stadium

or concert hall. Additionally, a piece of music uploaded on YouTube has everyone with internet access within its potential reach. Even though reach can vary widely in range, it is customary to start with a discussion of the two extreme forms of rival and non-rival goods and this book will stick to this custom here, keeping in mind that the understanding that one can develop from these cases must be somewhat modified, when applied to intermediate cases.

What one has also seen from the air-pollution example is that different types of transaction costs exist that have an impact on the functioning of markets, as well as on other institutions. In order to be able to use markets to allocate goods and services, one relies on the ability of the owner of the good to exclude others from its use. Without excludability, people would freely use the goods and services provided by others with the consequence that market transactions would never take place. Excludability of goods and services is, therefore, a necessary condition for the establishment of markets and the (opportunity) costs of exclusion are a major source of transaction costs in the market mechanism. This motivates the following definition:

▶ **Definition 6.4 Exclusion Costs** The transaction costs that are necessary to exclude third parties from the appropriation and use of goods and services owned by a person are called exclusion costs.

The reach of an economic activity and the exclusion costs span a two-dimensional map where goods can be pinned down according to their specific characteristics. Figure 6.6 illustrates this point.

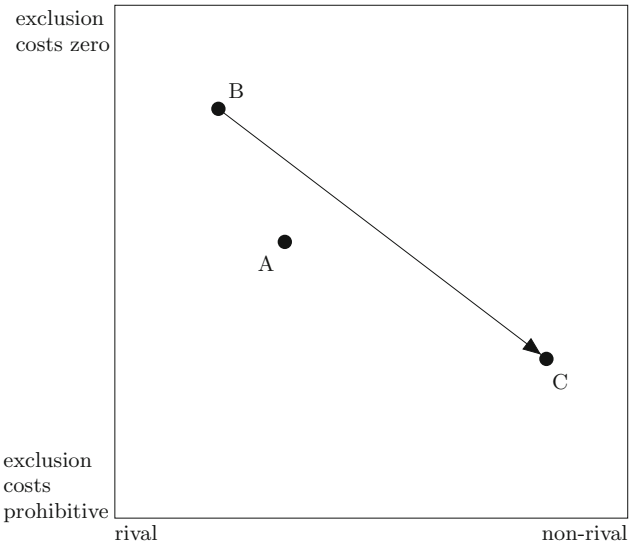

Fig. 6.6 Types of goods according to reach and exclusion costs

Table 6.1 A taxonomy of goods

	Rivalrous	Non-rivalrous
Excludable	Private goods	Club goods
Non-excludable	Common goods	Public goods

The four corners of this "map" define the boundary cases of minimum and maximum reach and zero and prohibitive exclusion costs. In reality, all goods are located somewhere in between. A point like *A* could, for example, be a car. Its reach is to carry up to five persons and exclusion costs are given by the price for locks and the alarm system.

Exclusion costs can vary over time. Take music as an example. In the good old days of the phonograph record, excluding third parties from the illegal consumption of music was relatively easy: in the absence of technologies for copying, exclusion required investments to prevent the theft of the physical record. The piece of music, as such, was non-rival in consumption, but the specific physical "carrier", the record, made it *de facto* rival (a point like *B* in the figure). With the invention of music cassettes, copying music became easier, which had an impact on the way property rights had to be protected. However, the big change came, of course, with the digitalization and distribution of music via the internet. This technological change essentially transformed music from a rival to a non-rival good and had a huge impact on the ability of the owner to exclude people from the illegal use of music (a shift from point *B* to point *C* in the figure). It took the music industry years to cope with this problem and to develop new business models. Technological inventions like copy and data-storage devices can, therefore, cause externalities for other products, like music or software.

Again, custom has it that one focuses on the two most extreme manifestations of exclusion costs. If exclusion causes zero transaction costs, then the good or service is called *(perfectly) excludable*. If exclusion causes prohibitive transaction costs, then the good or service is called *(perfectly) non-excludable*. Perfect excludability is, obviously, a simplifying assumption. To quote James Madison in the Federalist Papers No. 51 (p. 377), "If men were angels, no government would be necessary," because mankind would never steal, which is the only way perfect exclusion is possible without any costs. Otherwise, shop owners protect their shops by locks, security systems, and guards, all of which contribute to transaction costs. The same is true for the general public that protects its flats, houses, and cars against theft. However, some goods come relatively close to the ideal of perfect excludability, for example, the above-mentioned apple. An essential good that is non-excludable is oxygen in the air. Just try to enforce any property rights on a specific molecule.

The extreme cases of rivalry and excludability give rise to a two-by-two matrix of goods that is useful for a first discussion of the different types of challenges that have to be overcome, if one wants to organize economic activities. Table 6.1 gives an overview. The four boundary cases are called private goods, common goods, club goods, and public goods, and I will discuss them in turn.

Private Goods One does not have to devote much time and attention to private goods, because they are the ones whose efficient production and distribution can be organized relatively easily, at least in principle. They are also the type of good that is implicitly assumed in the theory of competitive markets, which is analyzed in Chap. 4. Their minimum reach makes the interdependence bilateral under conditions of scarcity, and market prices induce efficient incentives to produce and exchange these goods. If it is, furthermore, costless to exclude others from the use of these goods, without consent from the owners, there is nothing standing in the way of establishing markets.

Common Goods Things are getting much more involved when it comes to common goods, in the literature sometimes also referred to as Common-pool goods. These goods share the minimum-reach property, but it is not possible to allocate them using market mechanisms, because the owner of these goods cannot prevent others from their use, which is a prerequisite for the functioning of market transactions. The ability to exclude others from the use of resources, goods, and services depends very much on the state's ability to function properly. Even the night-watchman state needs laws and law enforcement to support the development of markets and, with weak state institutions (insufficient funding of the police, corrupt officials, etc.), excludability is far from guaranteed, which prevents markets from functioning efficiently. Irrespective of the quality of institutions, though, there are some goods and resources, whose inherent qualities make exclusion very costly. Examples are migratory species, like fish and birds, or oxygen. In comparison to cattle, where the assignment of property rights to specific animals is possible and effectively enforceable in principle, it is very hard to assign and enforce property rights to individual fish. This need not be an impediment to effective exclusion, as long as close substitutes to property rights for fish exist, and a close substitute could be property rights over the part of the sea where a shoal of fish lives. For example, the United Nations Convention on the Law of the Sea assigns exclusive economic zones (EEZ) to states. These zones grant special rights regarding the exploration and use of marine resources to nation states. They stretch from the baseline out to 200 nautical miles from its coast (as long as there are no overlaps between different countries). Territories beyond this stretch are international waters without exclusively assigned user rights.

Exclusion is, therefore, possible for all fish that migrate only within the boundaries of a given EEZ. However, for fish that migrate beyond or across the EEZs or in international waters, property rights over the sea are no effective substitutes for property rights over the fish themselves. The result is often overexploitation, because sustainable shoal management is not in the interest of the states or the fishery fleets: they have to bear the (opportunity) costs of sustainable management, but part of the revenues spill over to other nations or fleets. In order to understand this problem better, it makes sense to dig a little deeper into the economics of renewable resource management.

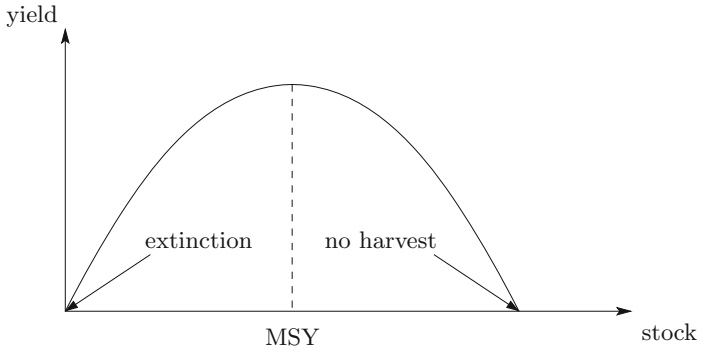

Fig. 6.7 Maximum sustainable yield

For all renewable resources, there is a causal relationship between the size of the stock and the yield. If the stock size is zero, obviously, the yield is zero, as well. Increasing stock size makes positive yields possible and the yield increases with stock size up to a certain point, where larger stocks require smaller yields again, up to the point of maximum stock size, which can only be sustained if the yield is zero. Figure 6.7 shows this relationship.

The *maximum sustainable yield* (MSY) is the largest yield (or catch for the fishery example) that can be taken from a species' stock over an indefinite period. It is given by point *MSY* in Fig. 6.7. Given this biological law, it is in the interest of a long-term business to adjust the yield around *MSY*. Underexploitation would leave money on the table and overexploitation would trade long-term for short-term profits. However, if the stock is not excludable, the incentives to act according to the long-run interests are diminished, because no user of the stock can be sure that the stock will still be there tomorrow. There is a tension between the logic of individual and the logic of collective action. I will come back to this point in Chap. 9.

Digression 6.6 (Cod)
One of the most "famous" examples for the overexploitation of marine resources is *gadus morhua*, or cod. Cod has been a very important commodity for about 600 years and dried cod (also called stockfish or clipfish) was an essential food for mariners. During the Middle Ages and the Age of Discovery, it was one of the most important commodities that made seafaring possible, because dried cod was one of the world's first non-perishable foods. It also became a popular food in Europe and, for about 250 years, 60% of all the fish eaten in Europe was cod. As early as 1620, cod fishing was at the center of international conflict, because various nations attempted to monopolize rich fishing grounds. Even the King of Spain married off his son

(continued)

Digression 6.6 (continued)
to the royal house of Portugal, because of fishing rights. By the late 1700s, codfish made New England an international commercial power.

For a very long time, it was beyond imagination that human activity could negatively impact the species, because it was famous for its reproduction rates. In the words of Alexandre Dumas (1873), "It has been calculated that if no accident prevented the hatching of the eggs and each egg reached maturity, it would take only three years to fill the sea so that you could walk across the Atlantic dryshod on the backs of cod." Human imagination proved to be too limited. Since the late 1950ies, technological advances, which have made fishing more effective, have heralded the start of a period of overfishing, which led to a first partial collapse of Atlantic northwest cod fishery in the 1970ies and a complete collapse in the 1990ies. In the summer of 1992, the Northern Cod biomass fell to 1% of its earlier level, see Fig. 6.8.

Cod is only a very prominent example of the problem of overfishing: the Peruvian coastal anchovy fisheries crashed in the 1970s after overfishing, the sole fisheries in the Irish Sea and the west English Channel have become hopelessly overfished and many deep-sea fish are at risk, as well as a number of species of tuna. A 2008 UN report asserts that the world's fishing fleet could be halved with no change in catch. Even more fundamental is the impact on the whole marine biosystem. Scientific evidence regarding the impact of humans on marine life is nicely summarized in a recent paper by McCauley et al. (2015): "Three lessons emerge when comparing the marine and terrestrial defaunation experiences: (i) today's low rates of marine extinction may be the prelude to a major extinction pulse, similar to that observed on land during the industrial revolution, as the footprint of human ocean use widens; (ii) effectively slowing ocean defaunation requires both protected areas and careful management of the intervening ocean matrix; and (iii) the terrestrial experience and current trends in ocean use suggest that habitat destruction is likely to become an increasingly dominant threat to ocean wildlife over the next 150 years. [...] Human dependency on marine wildlife and the linked fate of marine and terrestrial fauna necessitate that we act quickly to slow the advance of marine defaunation."

Unresolved commons problems even led to the collapse of whole societies throughout human history. As far as we know today, examples are the Greenland Norse, Easter Island, the Polynesians of Pitcairn Island, the Anasazi of southwestern North America and the Maya of Central America (Diamond 2005). There are, of course, always several factors that contribute to the collapse of a society, but overexploitation of natural resources plays a very prominent role.

Markets are only sufficient, not necessary, to reach efficiency. Humans have developed other effective means to cope with commons problems and property

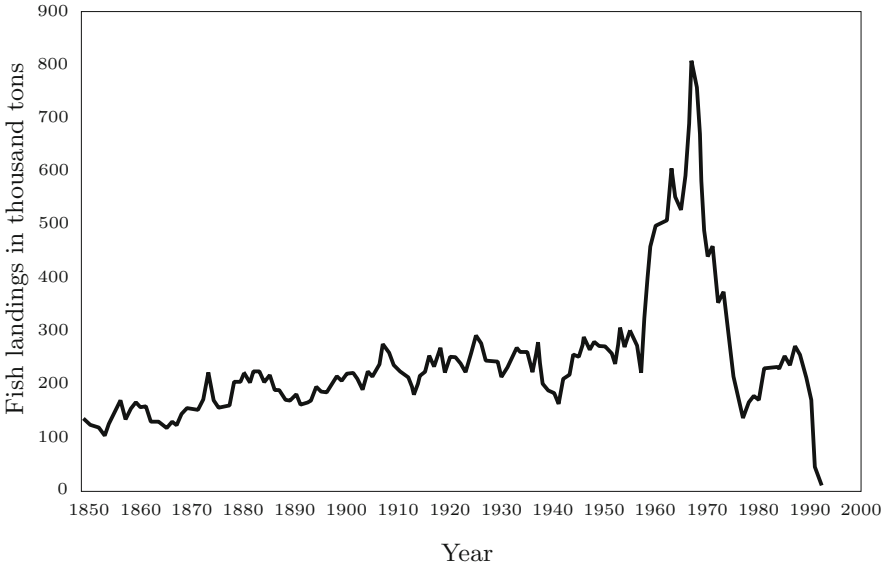

Fig. 6.8 Collapse of the North Atlantic cod fishery (source: Millennium Ecosystem Assessment, 2005, p. 12)

rights plus trade, interestingly, is not one of the most common forms of resource management, as is stressed by Elinor Ostrom who systematically analyzed solutions to the commons problem in different societies. She came to the conclusion that the absence of private property and markets need not be an impediment to the efficient and sustainable use of common goods. On the contrary, evidence suggests that well-maintained systems of resource and ecosystem management can, in fact, yield better results than markets can. These findings are of considerable importance for the way one should think about institutions, because they point towards the blind spots of the standard model in economics, which puts (too?) much emphasis on markets. What makes Ostrom's findings difficult to integrate into this discussion about common goods is, however, a tension between the microeconomic definition of common goods and the one she uses. This book defines them by the "technological" property of non-excludability, which excludes certain institutions *by definition*. Ostrom (2005) starts from a different perspective, focusing on goods and resources for which common property exists *de facto*. The set of admissible institutions, therefore, remains unclear and some of her criteria for successful institutional principles show that they rely on excludability. With this caveat, one can briefly discuss the basic principles of successful management of common goods that have been identified.

- Pretty much in line with the standard model, precise delineations of the resources and effective exclusion of externals are important. Hence, even if exclusion is not practiced within the group, it is important to exclude outsiders.

- One needs rules regarding the appropriation and provision of the common goods and these rules have to be adapted to local conditions. This property shows that institutional diversity is key, because there are close ties between the norms and cultures of groups and the environment.
- Rules for collective decision-making play an important role by giving voice to as many users of the common resource as possible and allowing the management system to adapt to changing environmental and social conditions.
- Monitors maintain compliance effective and are part of or are accountable to the users.
- One needs a scale of graduated sanctions for resource appropriators, who violate community rules.
- One needs mechanisms of conflict resolution that are cheap and are easily accessible for the conflicting parties.
- The self-determination of the community is recognized by higher-level authorities.

This list shows that institutions, which can effectively manage common goods, are diverse, but share common patterns. One-size-fits-all solutions that rely on property rights and markets should, therefore, be considered with caution, because they are only one means to cope with commons problems and may even be maladapted to local norms and traditions. However, given that the above principles have been identified in mostly stable and small communities, it remains an unresolved question whether they can be "scaled up" to cope with large or even global commons problems. Trust and sanctions are relatively easy to establish in small and stable communities and small-scale communities are also the environments in which human beings evolved and developed their intuitions about fairness and justice. A suggestion about how to deal with larger common-goods problems, which comes out of this line of research, is to organize them in the form of multiple layers of nested organizational units.

Club Goods If exclusion is possible and the good is non-rival, then it is a club good. The name sounds strange at first, but it will become clear as I discuss the implications of this combination of factors. Think of a live music concert or sports event. In order to be able to enjoy it, one has to enter a stadium or concert hall and this physical barrier can be used to exercise exclusion and to force one to buy an entrance ticket. Further examples for club goods are Pay TV, lectures, music and software or—to a certain extent—roads. I will briefly discuss them, to see if there are interesting patterns to be found.

Lectures at universities, for example, are pretty much like live music and sports events club goods, because one can, in principle, exclude people from attending and thereby enforce the payment of prices. These prices are sometimes also called entrance or user fees. Moreover, if the primary motive for attending lectures is a grade certificate, one can enforce the payment of user fees by withholding the certificate. Given that exclusion is possible in principle, it is mainly a political

decision of whether access should be regulated by the price mechanism and whether it shall be complemented by other mechanisms (like making a high-school degree a mandatory prerequisite for applications). A lot of public universities in Europe charge only moderate or no tuition fees, whereas private universities and also public universities in countries like the USA charge substantial amounts. MIT, for example, charged its students an annual fee of approximately $44,525, (academic year 2019/20), which is pretty much in line with other top US universities. The University of Cambridge discriminates tuition fees between UK and international students. Students from the UK pay £9250 for the academic year 2021, and international students pay £22,227 (for their economics programs).

Another aspect of lectures, music and sports is that the "live event" has a limited reach defined by the capacity of the lecture room, concert hall or stadium. Therefore, the maximum supply is defined by this capacity. To make sure that supply meets demand, one can rely on the price mechanism, adjusting user fees accordingly or one can use alternative rationing schemes. Universities, for example, screen students by means of entrance tests, and so on.

Why is there a difference in the way demand is rationed between, for example, music events and university programs? Profit-oriented universities face a tradeoff between short- and long-term profits. Assume that, at given tuition fees, demand exceeds supply, such that entrance tests must be used to ration. In the short run, the university could increase its revenues by increasing tuition fees, but this may have a negative impact on the selection of the student body, which may have a negative effect on the future reputation of the university, which—as the last step in a complex causal chain—has a negative effect on future entrance fees. This is not the case with other commercial events, like concerts and sports, because the talents and motivation of the audience has only a very limited effect on the quality of the event.

One has already seen that live events face certain capacity constraints, which limits the reach of a club good. These limitations can, in principle, be overcome by "going digital." Broadcasting sports events or live music and selling studio music via, for example, Spotify markedly extended the reach of these goods such that, at the maximum, everyone with access to the internet can get access to the product, which creates huge profit potentials for firms. However, every distribution channel has its own enforcement costs and the music industry had to learn this the hard way during the early days of the internet, when it was almost impossible to prevent illegal downloads. Thus, digital products somehow oscillate between the characteristics of a club good and a public good.

Last, but not least, roads are an example of a good for which regulating access via price mechanisms is becoming increasingly popular, partly because of changes in the available technologies of exclusion, and partly because of other trends. Access to most roads in the majority of countries is free, and traffic infrastructure is financed by taxes. One of the reasons is that road pricing and the investments in the setup and maintenance of the necessary exclusion technology is very costly in general. Furthermore, there is a lot of evidence that, as long as congestion is not an issue, a region's traffic infrastructure creates huge positive externalities, because it facilitates trade. For example, Paris experienced a boost in its economic

development after the abolishment of bridge tolls by Baron Haussmann in the nineteenth century. However, positive externalities caused by traffic infrastructure can easily be compensated by negative externalities, if traffic gets congested. The current trend to (re-)introduce tolls on highways, bridges, and other major roads is, to a certain extent, a reaction to the increasing economic costs of a congested traffic infrastructure, combined with more efficient technologies for the enforcement and collection of tolls that bring down the transaction costs of enforcing fees.

If one looks at club goods from a slightly different angle, one observes an interesting property because, as long as no capacity constraints are binding (there are still empty seats in the lecture room), an additional user of the good causes no additional costs. This property has two interesting implications.

First, from an efficiency perspective, it makes sense to increase the number of users to the largest extent possible, because each additional user increases the gains from trade (no additional costs, but additional consumer surplus). It follows that actually excluding people from using the club good can never be efficient. Exclusion is a mechanism that can be used to establish a market and, therefore, has to be distinguished from the act of actually excluding people. The threat of exclusion makes the enforcement of prices (like tolls or tuition fees) possible, but it depends on the actual prices whether potential users will be excluded or not.

Second, the fact that firms can serve additional customers at approximately zero additional costs creates a tendency towards the monopolization of markets for club goods. Take software as an example. From the point of view of a software developer, the lion's share of the costs she has to invest is caused by the development of the product. As soon as the product is on the market, each additional user causes approximately zero additional costs. Hence, the more users there are the better, for the software developer. The fact that the minimum price that is necessary to break even falls as the number of users increases creates an inherent tendency for market concentration: firms with larger market shares can outcompete their smaller competitors, because they can charge lower prices without running a deficit. This is the reason why club goods are sometimes also called *natural monopolies*. I will come back to this point when I discuss production costs in Chap. 12 and monopoly pricing in Chap. 14.

Public Goods The last type of good is non-rival in consumption and exclusion is impossible. If exclusion is impossible, markets cannot be used to incentivize the production and allocation of these goods, so one has to look for alternative ways to organize economic activity.

Examples for public goods are fireworks, basic research, national defense, avoidance of climate change and legal systems, and the following paragraphs will discuss all five examples in turn.

Fireworks are an example for a (local or regional) public good, because no one in a city can be effectively excluded from the spectacle and it is also non-rival. Arguably, the other examples are more important than fireworks.

Basic research is non-rival, because the fact that I understand a mathematical theorem does not make it impossible for another person to understand it at the same time. All knowledge, in this sense, is non-rival. The difference between basic and applied research is, therefore, not the degree to which goods are rival, but the ability to exclude. Applied research usually builds on basic research and "brings it to the market." A good example is quantum physics. Without quantum mechanics, there would be no transistor and hence no personal computer and no laser. Therefore, the development of quantum mechanics made the development of a large number of products possible, without which today's world would be impossible. Products or components thereof, like transistors or computers, can be effectively protected by patent law. However, the protection of property rights for the mathematical formulation of quantum physics, like the Schrödinger equation or the uncertainty principle, is not as easy to do. Even if a formal property right exists, one cannot sell the Schrödinger equation directly and it is, in general, very hard to establish a causal link between abstract physical principles and marketable goods, such that potential property-rights infringements would be hard to detect.

The public-goods character of basic research requires alternatives to the market mechanism and one can find essentially two different ways to organize the production process. One is public financing. Major resources for basic research at universities and research institutions are provided by the state and financed by taxes, and career incentives for scientists have the form of a contest, where the relatively most successful qualify for professorships and research money. The alternative is to interpret education and research as complementary bundles where basic research is, at least partly, financed by tuition fees and students profit indirectly from the direct access to a research-intensive environment, because new ideas disseminate earlier, which gives them a competitive edge in developing new, marketable products. A good example is the synergistic relationship between Stanford University and Silicon Valley startups and companies.

A staple example for public goods is national defense. It is relatively obvious that, within certain geographic limits, a military of a given size provides a non-rival service to its citizens. By its very nature, the reach of national defense is the people living within the territory of a nation state (protecting people living abroad is much more difficult). The non-excludability of national defense becomes apparent, if one distinguishes between an actual military conflict and the insurance against attacks provided by the military. In case of an act of aggression, exclusion of specific citizens is, in principle, possible. One could escort them to the border and hand them over to the enemy. However, it is virtually impossible to exclude people within the territory from the insurance provided by the existence of the military, which results from the fact that one is not attacked at first place.

Last, but not least, the avoidance of climate change has important properties of a truly global public good: CO_2 emissions have global effects on the climate, so measures to slow down climate change cause non-rival effects. Similarly, no one can be excluded from the effects of climate change (or the effects from slowing it down). The global nature of climate change is what makes the problem so difficult to solve. The expected costs and benefits of climate change are unevenly distributed

between countries and regions and international negotiations take place within the holey network of international law. International agreements are difficult to reach and they are even more difficult to enforce. If one would ask a group of social scientists and psychologists to design a problem that is hard to solve for human beings, I am pretty sure that it would look very much like climate change.

The last example for a public good that I will discuss is the legal system of a country, because it allows one to focus one's attention on the fact that excludability need not be a physical characteristic of a good. The legal system of a country is clearly non-rival. If A uses contract law to set up contracts, it does not impede B from using the contract law herself. Things get more complex when it comes to excludability. Technically, it is no problem to exclude people from contract law because the courts could decide not to apply it to contracts signed by specific people. However, contract law is embedded within the rest of the legal system, which makes such restrictions illegal. It can (and, in practice, usually does) specify that all laws apply equally to all citizens of a country. Such a norm creates a legal non-excludability and the system depends on levels of analysis to determine whether such constraints are taken as a given or if they are subject to scrutiny.

References

Akerlof, G. (1970). The market for lemons: Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, *84*(3), 488–500.

Arrow, K. (1971). *Essays in the theory of risk-bearing*. Amsterdam: North-Holland.

Breeze, T. D., Bailey, A. P., Balcombe K. G., & Potts S. G. (2011). Pollination services in the UK: how important are honeybees? *Agriculture, Ecosystems and Environment*, *142*(3–4), 137–143.

Carson, R. (1962). *The silent spring*. Mariner Books.

Coase, R. (1937). The nature of the firm. *Economica*, *4*(16), 386–405.

Coase, R. (1960). The problem of social cost. *Journal of Law and Economics*, *3*, 1–44.

De Groot, R. S. (1992). *Functions of nature: Evaluation of nature in environmental planning, management and decision making*. Groningen: Wolters-Noordhoff.

Diamond, J. (2005). *Collapse: How societies choose to fail or succeed*. Penguin Books.

Dumas, A. (1873). *Grand Dictionnaire de Cuisine*. Hachette.

Elton, B. (1991). *Gridlock*. Sphere Books.

Inglehart, R., & Klingemann, H. D. (2000). Genes, culture, democracy, and happiness. In E. Diener, & M. Eunkook (Eds.), *Culture and subjective well-being*. MIT-Press.

Killingsworth, M. A. (2021). Experienced well-being rises with income, even above $75,000 per year. *Proceedings of the National Academy of Sciences, 118*(4), e2016976118.

Krishnamurti, J. (2005). *Life ahead: On learning and the search for meaning*. New World Library.

Lange, O. (1936). On the economic theory of socialism: Part one. *The Review of Economic Studies*, *4*(1), 53–71.

Lange, O. (1937). On the theory of economic socialism: Part two. *The Review of Economic Studies*, *4*(2), 123–142.

Layard, R. (2005). *Happiness: Lessons from a new science*. Penguin Books.

Lefebvre, H. (1974). *The production of space*. John Wiley.

Lindh, A. (2002). Speech by Anna Lindh in the Helsinki Conference.

McCauley, D. J., Pinsky, M. L., Palumbi, S. R., Estes, J. A., Joyce, F. H., & Warner R. R. (2015). Marine defaunation: Animal loss in the global ocean. *Science*, *347*(6219), 1255641.

Millennium Ecosystem Assessment, (2005). *Ecosystems and human well-being: Synthesis*. Island.

Morse, R. A., & Calderone, N. W. (2000). The value of honey bees as pollinators of U.S. crops in 2000. *Bee Culture*, *128*, 1–15.
Ostrom, E. (2005). *Understanding institutional diversity* Princeton University Press.
Parfit, D. (1984). *Reasons and persons*. Oxford University Press.
Southwick, E. E., & Southwick, L. (1992). Estimating the economic value of honey-bees (Hymenoptera, Apidae) as agricultural pollinators in the United States. *Journal of Economic Entomology*, *85*, 621–633.
Stern, N. (2007). *Stern review on the economics of climate change*. Cambridge University Press.
Stevenson, B., & Wolfers, J. (2008). Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox. Discussion paper, National Bureau of Economic Research.
Sumner, D. A., & Boriss, H. (2006). Bee-conomics and the leap in pollination fees. *Agricultural and Resource Economics Update*, *9*(3), 9–11.
Thucydides, (2013). *The history of the peloponnesian war*. CreateSpace Independent.
World Wildlife Fund. (2014). Living Planet Report 2014. http://www.footprintnetwork.org/images/article_uploads/LPR2014summary_low_res.pdf.

Further Reading

Arrow, K. J. (1969). The organization of economic activity: Issues pertinent to the choice of market versus non-market allocations. In *Analysis and evaluation of public expenditures: The PPP system*. Washington, D.C: Joint Economic Committee of Congress.
Coase, R. (1998). The new institutional economics. *American Economic Review*, *88*, 72–74.
Furuborn, E. G., & Richter, R. (2005). *Institutions and economic theory: The contribution of the new institutional economics*, 2nd edn. University of Michigan Press.
Laffont, J. J. (2008). Externalities. *The New Palgrave: Dictionary of economics*.
Marney, G. A. (1971). The 'Coase Theorem:' A reexamination. *Quarterly Journal of Economics*, *85*(4), 718–723.
Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.
Swiss Code of Civil Procedure. (2006). https://www.admin.ch/opc/en/classified-compilation/20061121/201601010000/272.pdf. Accessed 08/04/2017
Williamson, O. E. (1985). *The economic institutions of capitalism: Firms, markets, relational contracting*. Free Press.