1

Society: Coordination Problems & Economic Institutions

> Two neighbors may agree to drain a meadow, which they possess in common; because 'tis easy for them to know each others mind; and each must perceive, that the immediate consequence of his failing in his part, is the abandoning of the whole project.
>
> But 'tis very difficult and indeed impossible, that a thousand persons shou'd agree in any such action; it being difficult for them to concert so complicated a design, and still more difficult for them to execute it; while each seeks a pretext to free himself of the trouble and expense, and wou'd lay the whole burden on others.
>
> David Hume, *A Treatise of Human Nature*.[1]

DOING ECONOMICS
This chapter will enable you to:

- Use game theory to analyse how people interact in the economy, each affecting the conditions under which the others decide how to act.
- Understand why the outcomes of interactions are often worse than they could be when people fail to coordinate with each other and to take account of the effect of their own actions on others.
- Explain how problems like environmental damage and global poverty can be the result of failed coordination.
- Represent institutions as "the rules of the game" and see how changing these rules will change outcomes.
- See that economic institutions determine incentives for people's behavior and can affect how successfully we address coordination problems.
- Explain why when people have limited information and conflicts of interest they often fail to implement 'win-win' outcomes.

1.1 Introduction: Poor economics

At the turn of the present century, the process of economic development had bypassed almost all of the two hundred or so families that made up the village of Palanpur in the Indian state of Uttar Pradesh. But for the occasional watch, bicycle or irrigation pump, Palanpur appeared to be a timeless backwater, untouched by India's cutting edge software industry and booming agricultural regions. Less than a third of the adults were literate, and most had endured the loss of a child to malnutrition or to illnesses that had long been forgotten in other parts of the world.

Figure 1.1: **Palanpur farmers threshing and winnowing grain (separating grain from chaff)** . Photo courtesy of Nicholas Stern.

EXAMPLE In this video tinyurl.com/y5fg3whx Thomas Piketty and James Heckman explain why data is fundamental to their work (from the CORE project. www.core-econ.org)

A visitor to the village approached a farmer and his three daughters weeding a small plot of land. The conversation turned to the fact that Palanpur farmers plant their winter crops several weeks after the date that would maximize the amount of grain they could get at harvest time. The farmers knew that planting earlier would produce larger harvests, but no one, the farmer explained, wants to be the first farmer to plant, as the seeds on any lone plot would be quickly eaten by birds.

Curious, the visitor asked if a large group of farmers, perhaps members of the same extended family, had ever agreed to plant their seeds earlier, perhaps on the same day to minimize the individual losses. The farmer looked up from his hoe and made eye contact with the visitor for the first time "If we knew how to do that," he said, addressing the visitor as "bhai" or brother,"we would not be poor."[3]

HISTORY In his address accepting the Nobel Prize for economics in 1979, University of Chicago economist T.W. Schultz said: "Most of the people in the world are poor, so if we knew the economics of being poor, we would know much of the economics that really matters." He was right then and he is right now. What is called the Nobel Prize in economics, officially is the Nobel Memorial Prize in Economic Sciences.[2]

1.2 Societal coordination: The classical institutional challenge

For the Palanpur farmers, the decision when to plant is a **coordination problem**. A coordination problem is a situation in which people could all be better off, or at least some be better of and none be worse off, if they all jointly decided how to act – that is, if they coordinated their actions – than if they act individually.

COORDINATION PROBLEM A coordination problem is a situation in which people could all be better off (or at least some be better of and none be worse off) if they jointly decide how to act – that is, if they coordinate their actions – than if they act independently.

The planting choice is a *coordination problem* because:

- the farmer does better or worse depending on what other farmers do,
- all the farmers would do better if they could *coordinate* their actions by jointly agreeing to all do what would be mutually beneficial namely, planting early, but
- it is a *problem* because the farmers may not be able to coordinate, and as a result
- if they do *not* coordinate and plant late, then all of the farmers will do worse than they all could have done (that is, had they all planted early).

To stress the fact that coordination problems often affect an entire population (even though we explain them using two person examples) we sometimes use the expression *societal* coordination problems. Notice that one farmer cannot dictate the actions of the other farmers, nor can they come to a common agreement about what to do ("if we knew how to do that, we would not be poor") – the inability to come together and coordinate is at the heart of coordination problems.

Figure 1.2: **Poor economics.** Esther Duflo and Abhijit Banerjee founded the Massachusetts Institute of Technology's Poverty Action Lab to bring the best minds in economics to bear on eradicating global poverty. Their 2011 book is titled *Poor economics*.[4] In 2019 the two MIT professors were awarded the Nobel Prize in economics along with Micheal Kremer for their research on the causes of poverty and methods to raise the living standards of poor people.

EXAMPLE In this video, tinyurl.com/yxpf72hm Esther Duflo explains what happened when it was mandated that randomly selected villages elect a woman to head their local council (from the CORE project. www.core-econ.org)

David Hume (the 18th century British philosopher and economist quoted at the start of this chapter) used an example – two landowners considering draining a meadow – to pose what he considered the most important problem facing society, namely, devising institutions that would reconcile the pursuit of individual objectives (avoiding the "trouble and expense" in his example of the meadow) with getting desired societal outcomes (improving the value of the meadow by draining it). His simple two-person example was meant to illustrate the need (in a society of "a thousand persons") for a government to address the broader societal coordination problems of his day.

Though the term was invented only two centuries after Hume, he was using

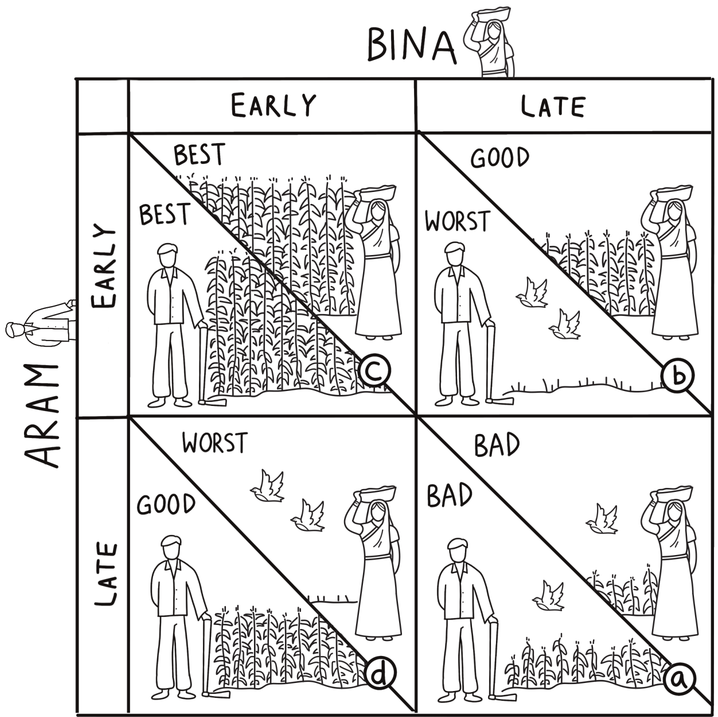


Figure 1.3: **Planting in Palanpur.** This figure shows "what-if" outcomes for planting in Palanpur. Each column represents a possible combination of Aram planting early or late and Bina planting early or late with the corresponding outcomes being worst, bad, good, or best in terms of how much grain they grow. Source: Anmei Zhi.

what we now call game theory to make his case. Let's apply his reasoning to the farmers of Palanpur. Like Hume we will consider just two farmers as a way of representing the institutional challenge faced by the entire village.

Figure 1.3 shows the outcomes for two players, Aram and Bina, choosing when to plant their grain. The figure illustrates the values of the farmers' crops, whether they consume the crop themselves or sell it for money to spend on other things.

Each farmer can either plant early or plant late, and while (also as in Hume's example) two people could probably come to some agreement about what to do, remember that we are using this two-person example to illustrate the entire village of about 200 families of farmers. So we assume that they cannot coordinate on some agreed upon actions for the two jointly. There are four possible outcomes:

- If both players plant early, they each achieve their *best* possible harvest, because they grow the most grain through sharing the risk of having their seeds eaten by birds (outcome c in Figure 1.3).
- If Aram plants early while Bina plants late, Aram has his seeds eaten by birds and gets no harvest (the *worst* outcome for him), whereas the late planter gets a *good* (but not the best) harvest. While none of her seeds are eaten by the birds, planting late is not the best for growing the most grain

(outcomes b and d in Figure 1.3). The same is true if Bina planted late when Aram planted early.

- If both plant late, they harvest a smaller crop while also sharing the risk of their seeds being eaten, a *bad* outcome, but not the worst (outcome a in Figure 1.3).

COORDINATION FAILURE A coordination failure occurs when the non-cooperative interaction of two or more people leads to a result that is not Pareto efficient.

The people of Palanpur are stuck in the bad outcome even though they would all be better off if they all planted early (they would both move from a "bad" outcome to the "best" outcome in the figure). They are experiencing a **co-ordination failure**, namely a coordination problem that is not addressed by appropriate institutions. A modern day David Hume would point out that a government could simply impose a sufficient tax on those planting late to ensure that most farmers would plant early.

Adam Smith, a generation after Hume, would stress the value of the exchange of privately owned goods on competitive markets as a way of coordinating the actions of large numbers of people, who would be guided (even without knowing it) by what he termed "an invisible hand." Hume, Smith and the other founders of European political philosophy and political economy posed what we call the classical institutional challenge.

HISTORY Adam Smith wrote: "[E]very individual [. . .], indeed, neither intends to promote the public interest, nor knows how much he is promoting it [. . .] he intends only his own security; . . . he intends only his own gain, and he is in this . . . led by an invisible hand to promote an end which was no part of his intention . . . By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it."[5]

These philosophers and economists wanted to know how *to design* institutions *so that people could be left free to make their own decisions, and at the same time avoid outcomes that were inferior for everyone.* More precisely, how do we design institutions which encourage coordination by free choice while avoiding poor outcomes such as planting late in Palanpur? The 18th and 19th century political economists and philosophers who founded the field of economics were attempting to provide solutions to coordination problems.

Checkpoint 1.1: Planting in Palanpur: A coordination problem

Imagine that you are Bina in the figure above, and that you did not know whether Aram would plant early or late. What would you do? Suppose, contrary to what we have assumed, you and Aram were neighbors and you could talk with him. What would you say?

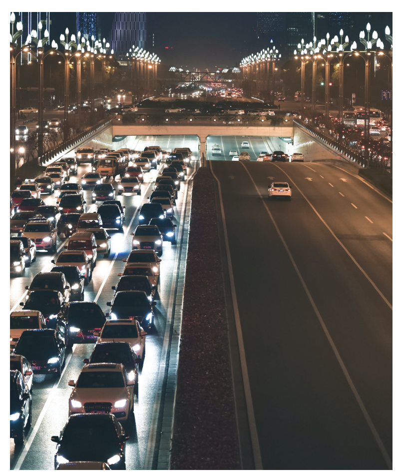


Figure 1.4: **Traffic headed out of a major city.** Image Credit: Photo by Preillumination SeTh (@7seth).

1.3 The institutional challenge today

The classical institutional challenge remains with us, although some of the forms that it takes today were unknown to the great 18th and 19th century thinkers.

Consider the following coordination problems:

- How do we sustain the global environment? To avoid damaging climate change we need to coordinate our reduction of emissions. Many people and firms would prefer that someone *else* reduce *their* carbon footprint. How can we address climate change in a way that is both fair and imposes the least possible costs?

- How do we make the best use of our ability to create and use knowledge? If we all agree to share the knowledge we have with others we may all be better off: when I transfer my knowledge to you I do not lose the ability to continue using it. But each of us may profit by restricting others' use of our knowledge by means of patents, copyrights and other intellectual property rights.
- How do we move around a city without overcrowding streets and causing delays? My decision whether to drive, walk, or take public transport affects not only my own travel time, but also the degree of traffic congestion and delays experienced by everyone else. Everyone might be better off if the use of private vehicles was substantially reduced, but few will reduce their driving unless some way is found to implement a general reduction by everyone.

These are all coordination problems because an outcome that is better for all is possible if people find a way to jointly agree to a course of action. But for reasons we will explain in detail, people routinely fail to coordinate and suffer bad consequences as a result, including the following:

- *over*use of some resources illustrated by pollution, over-grazing, traffic congestion, and climate change; and
- *under*use of other resources such as the productive capacities and creativity of people and the knowledge that we have created, illustrated by unemployment and the enduring poverty of the people of Palanpur and villages like it around the world.

Checkpoint 1.2: Coordination problems you have known

Think of a social interaction in which you have been involved that was a *coordination problem*, and, using the description of why planting in Palanpur is a coordination problem (the bulleted points above), explain why it was a problem and how coordination might have (or did) address the problem.

1.4 *Anatomy of a coordination problem: The tragedy of the commons*

The over use of environmental resources provides a good illustration of why coordination problems arise.

In 1968, Garrett Hardin, an ecologist, famously described what he called the **tragedy of the commons**, an example of a coordination failure.[6] He told a story about a group of herders who share a pasture. The pasture was common land – hence a "commons" – shared by many herders. But why was his story a tragedy?

TRAGEDY OF THE COMMONS The tragedy of the commons is a term used to describe a coordination problem in which self-interested individuals acting independently deplete a common resource, lowering the payoffs of all.

Each herder could put as many animals in the pasture as they wished, and overgrazing will lead to erosion and the ruin of the pasture. Hardin reasoned that if the land is common to all and no one herder owns it, each herder has no interest in limiting how many animals they put in the common pasture. A ruined pasture is of no value to any of the herders. But each herder's self-interest leads them to neglect the effect their actions have on others. The outcome is a tragedy.

With the term *tragedy of the commons*, Hardin gave social science one of the most evocative metaphors since Adam Smith's "invisible hand." Indeed Hardin called his tragedy a "rebuttal to the invisible hand." The two metaphors are powerful because they capture two essential yet contrasting social insights. When guided by an invisible hand, social interactions reconcile individual choice and socially desirable outcomes. By contrast, the actors in the tragedy of the commons pursue their private objectives to tragic consequences for themselves and others.

The natural setting for Hardin's tragedy was chosen for its imagery. The underlying problem applies to many situations where people typically cannot or do not take account of the effects of their actions on the well-being of others. You can think of a city's streets as a commons, and people deciding to drive rather than walk, bike, or use public transport as similar to the herders putting cattle on the common. A modern day "tragedy of the roadways" is a traffic jam.

What are the common elements in Hume's drain the meadow problem, the farmers in Palanpur planting late, Hardin's herders overgrazing their pasture and our modern city dwellers clogging the streets with their vehicles?

EXTERNAL EFFECT An external effect occurs when a person's action confers a benefit or imposes a cost on others and this cost or benefit is not taken into account by the individual taking the action. External effects are also called externalities.

In each of these three cases, the reason why uncoordinated activities of people pursuing their own ends produce outcomes that are worse for all is that each participant's actions affect the well-being of others, but these effects are not taken into account by the individual actors when they decide how to act. These impacts of our actions on others that we do not take account of in deciding what to do are termed **external effects**.

Here are the external effects (italicized in the list below) that actors in our four examples do not take into account when deciding what to do:

- The person who lives in a city who drives to work, *adds congestion* to the streets, and therefore increases the travel time of others.
- Hume's farmer who does not drain the swamp and *imposes the cost* of

doing so on the other farmer.

- The Palanpur farmer who plants late, *imposes a cost* on the other farmer who will have his seeds devoured by birds if he plants early.
- Likewise the farmer who plants early *confers a benefit* on the other farmer who can benefit by planting at the right time (early) without severe losses of seed to the birds.
- The herder who places additional cattle on the common pasture *reduces the grass available* to the other herders stock.

Addressing coordination problems by internalizing external effects

Simply abolishing these and other external effects that are the root of coordination problems is not an option. There is no way to organize society so that nothing that we do would affect others, each person on his or her self sufficient island.

Apart from not being much fun, life would be impossible in a society of total social isolates (just think about how the next generation would be born and raised!). So, to address the classical institutional challenge to prevent or at least minimize coordination failures we need to find ways of inducing each participant *to take adequate account of the effects of their actions on others.*

This is called *internalizing* an *external effect*. We use the term *external effect* because the effect is *outside* of the individual's process of decision-making when taking the action. To internalize the external effect, you ensure that the person who acts bears the costs of their negative effects on others and reaps the rewards of their positive effects on others. In this way the otherwise "external" costs and benefits become part of the individual's decision-making process, leading them to "take adequate account of the effects of her actions on others."

If the "others" are our family, our neighbors, or our friends, our concern for their well-being or our desire to be well regarded by others might get us to take account of the effects of our actions on them. Reflecting this fact, an important response to the classical institutional challenge – one that long predates the classical economists – is that caring for the well-being of others need not be confined to friends and relatives but may extend to all of those with whom we interact. Ethical guides such as the "golden rule" are ways that people often internalize the effects of our actions on others, even when the others are total strangers to us.

HISTORY The "golden rule" is "to do unto others as you would have them do unto you" (Matthew, 7: 12). Or, treat others as you would like to be treated yourself. The same ethical principle is found in Islamic scriptures and in the teaching of other religions.

But, over the past five centuries, people have come to interact not with a just few dozen people, as humans have for most of our history and pre-history, but directly with hundreds and indirectly with millions of strangers. The classical economists in the 18th century were responding to the fact that the generosity or ethical motivations that one might feel towards one's family or neighbors would not be sufficient to induce people to take account of the effect of their actions on others, once these external effects spread across the entire network of global interactions.

From its 18th century origins up to today, an objective that economics has set for itself, therefore, has been to design and implement institutions that would induce people to act *as if* they cared about those who were affected by their actions even when that was not literally true.

Checkpoint 1.3: External effects

a. Provide an example of a *negative external effect* that occurs in a social interaction. Explain *why* it is negative and why it is external.
b. Provide an example of a *positive external effect* that occurs in a social interaction. Explain *why* it is a positive external effect.

1.5 Institutions: Games and the rules of the game

Institutions

Institutions are the laws, informal rules, and mutual expectations which regulate *social interactions* among people and between people and the biosphere.

Think about driving on the right or on the left as a coordination problem (not a very challenging one). People adopt the behaviors prescribed by institutions (e.g. drive on the right if you are in the U.S.) because of some combination of

- *laws* enforced by a government (you will be arrested and fined for driving on the left in Brazil, the U.S, France, and other countries where driving on the right is the law.)
- *social pressures* – sometimes termed informal rules because they are not enforced by governments (your friends and neighbors will disapprove and think less of you if you drive on the left), and
- *mutual expectations* that you have about what others will do and have about what you will do (you expect others to drive on the right because they expect you to drive on the right, so you will avoid accidents by doing the same.)

INSTITUTIONS Institutions are the laws, informal rules, and mutual expectations which regulate social interactions among people and between people and the biosphere.

We refer to institutions as the rules of the game. To see what this means we now introduce an important conceptual approach for understanding society. **Game theory** uses mathematical models and verbal arguments to analyse how the outcomes of the interaction for the participants will depend on the rules of the game and the objectives of the players. It has been used exten-

GAME THEORY Game theory is the branch of applied mathematics that studies strategic interactions.

sively in economics and the other social sciences, biology, and computer science.

Game theory focuses on **strategic interactions** where participants are interdependent and are aware of this interdependence: one player's outcome depends on their own and other players' actions and all players know this. We can contrast strategic with *non-strategic* situations in which the effect of your actions on the outcomes you will experience is independent of what others do. An example: your enjoyment of the program you are streaming at home alone is substantially independent of what others may be doing.

STRATEGIC INTERACTION An interaction is strategic when participants' outcomes are interdependent – their profit, standard of living, or some other measure of their well-being depends on the actions that both they and others choose, and this interdependence is known to the actors. A short-hand expression for the term strategic is: mutual dependence, recognized.

But many of our economic and social interactions are strategic:

- those considering driving to work know that their travel time will depend on how others decided to get to work that morning;
- the Palanpur farmer knows that how his crop will fare if he plants early will depend on how many others planted early.

EXAMPLE In 2020 under the pressure of the popular protests, the government of Chile established a set of rules governing how the constitution of Chile would be amended. [7]

Checkpoint 1.4: Institutions

a. Give an example of a strategic and a non-strategic interaction.
b. Which of the three items on the list of reasons why people coordinate on the side of the road on which to drive – laws, social pressures, and mutual expectations – explain why the farmers in Palanpur plant late?

Figure 1.5: **John von Neumann (1903-1957)** was a Hungarian-American mathematician, computer scientist, and physicist who is regarded as the father of game theory,[8] which he hoped would allow us to better understand the anti-Semitism and fascist political upheavals that he had witnessed in the early 20th century and provide the basis for understanding how groups interact. Photo Credit: Wikimedia Commons.

Games

When we model strategic interactions using game theory we call the actors *players*. Players can be people, owners of firms, social movements, governments and a variety of other entities. In biology, where game theory has been extensively used, even sub-individual entities are "players" such as viruses "trying to" spread in a pandemic or genes "trying to" get as many copies of themselves made as possible. Players may choose from a list of possible strategies (called a strategy set). For example, a strategy set might include "Purchase a bicycle for $350." But the rules of the game reflect institutions: if private property is an institution that is present and enforced, then the strategy set would not include "Pick up any available bicycle," without specifying the possible penalties for stealing.

SET A set (in mathematics) is a collection of objects defined either by enumerating the objects, or by a rule for deciding whether any particular object is in the set or not. For example, the set of positive, even integers less than or equal to 10 is, $\{8,2,6,10,4\}$.

The Palanpur farmers' strategies are 'Plant Early' or 'Plant Late.' The strategies could also include a strategy based on what others did in the past (called a contingent strategy) such as: "Plant early as long as at least 5 others planted early last season." The description of a *game* requires us to identify the following:

EXAMPLE People can *change* these rules, so institutions can themselves be outcomes of games that govern how the rules of the game can be changed. FIFA governs how football (soccer) can be played by what are called The Laws of the Game. These institutions also change: the corner kick was introduced in 1872 when the U.K. Football Association changed the rules.

- *Players*: a list of every player in the game whether they be individuals (like the farmers in Palanpur), an organization such as Amazon or Alibaba, or some other entity that can be represented as a single actor choosing between alternative courses of action.

- *Strategy sets*: a list for each player of every course of action available to them at each point where they must make a choice (including actions that depend on the actions taken by other players, or on chance events). The strategies selected by each of the players – the outcome of the game – is called the strategy profile.

- *Order of play:* a game can be *simultaneous* such that players make their choices without knowing the choices of others, as in the game of rock-paper-scissors. Or a game can be *sequential* such that players move in sequence, one after the other, as in chess, so that each player knows and responds to the choices of the previous players.

- *Information*: A game also specifies
  - who "knows" what,
  - when do they "know" it,
  - if what they "know" is known to others as well,
  - if what they "know" can be used in a court of law to enforce a contract, and
  - if what they "know" is true (this is why we use the quotation marks)?

- *Payoffs:* Are numbers assigned to each possible outcome of the game (each strategy profile) for each player; a player chooses a strategy with the intention of bringing about the strategy profile with the highest number.

EXAMPLE When we model the coordination problem of the Palanpur farmers as a game we assume they plant simultaneously. But when we model the interaction between a bank and a borrower we assume that the banks first makes an offer (the loan size, interest rate and schedule of repayment) and the prospective borrower responds.

It is often useful to consider payoffs as something that the players actually get. For example, considering the farmers in Palanpur again, an outcome of the game is a strategy profile indicating who plants early and who plants late, and the payoffs could be the amount of grain each farmer harvests. We say that the payoff associated with a particular outcome of a game is how much the player *values* that outcome. But that means nothing more than that a player will choose a strategy resulting in an outcome with a higher payoff number if possible.

An important distinction concerning strategy sets is whether or not one of the strategies open to the players is to jointly agree on a strategy profile – that is to deliberately coordinate their actions. This is possible in what is called a **cooperative game.**

COOPERATIVE GAME A strategic interaction for which the players' choice of a strategy is subject to a binding (enforceable) agreement.

We use the set of *players*, their *strategy sets*, their *payoffs*, the *order of play*, and the *information* the players have to describe the institutions governing some economic interaction, whether it is between an employer and an employee, or a central bank like the U.S. Federal reserve and a commercial bank.

But even this detailed description of the interaction does not give us enough information to predict how the game will be played.

EXAMPLE Watch this video tinyurl.com/y2cosf8v to see how a sophisticated game theorist brings an episode of the show Golden Balls to a surprising close (from the CORE project www.core-econ.org)

The outcome of a game – how it will be played, resulting in a particular strategy profile – is called a solution. To determine the solution as a way of predicting the outcome of a game we need what is called a **solution concept**. A solution concept for a cooperative game would include some rule for deciding on what the coordination would be, for example allowing one player selected at random to dictate the outcome, or a particular system of voting.

SOLUTION CONCEPT A solution concept is a rule for predicting the outcome of a game, that is, how a game will be played.

But by positing some way that people could jointly implement some outcome, cooperative game theory *assumes away* the problem of coordination. And the problem of how coordination is to be achieved is at the heart of the classical institutional challenge whether it takes the form of climate change or traffic jams.

So we need to see how players might coordinate in what is initially a non-cooperative setting – one in which coordination is not assumed at the outset – lets take a concrete example: people interacting in a way that results in the over-exploitation of an environmental resource. We will use this example to illustrate a basic solution concept for non-cooperative games: the Nash equilibrium.

Checkpoint 1.5: Games
a. What is a game?
b. How do you describe the *outcome* of a game?

1.6 Over-exploiting nature: Illustrating the basics of game theory

Figure 1.6: **Elinor Ostrom (1933-2012).** Elinor Ostrom was an American political scientist who won the Nobel Prize in economics for her contributions to understanding coordination problems, such as that encountered by Alfredo and Bob in the Fishermen's Dilemma, and on the institutions that promote cooperation in groups. Especially important are her empirical studies showing how people address the coordination problems that they encounter in seeking to maintain their livelihoods, such as grazing herds of cattle, fishing, or managing shared forests. Photo Credit: Holger Motzkau. Wikimedia Commons.

People who fish for a living interact with each other regularly. Each of them is aware that how much they benefit from fishing depends not only on their own actions, but on the actions of others. This is because the more others fish, the more difficult it will be for each to catch fish. The fishermen therefore impose negative external effects on each other. And this, along with the difficulty they face in agreeing on a common course of action, is why they face a coordination problem. Given that they cannot jointly decide on how much to fish, each faces a basic question: how much fishing to do given the strategies adopted by others who are fishing the same waters?

The game set up

Specifically, we consider two fictional fishermen, Alfredo and Bob, who share access to a lake, and catch fish, which they eat. There are no other people affected by their actions.

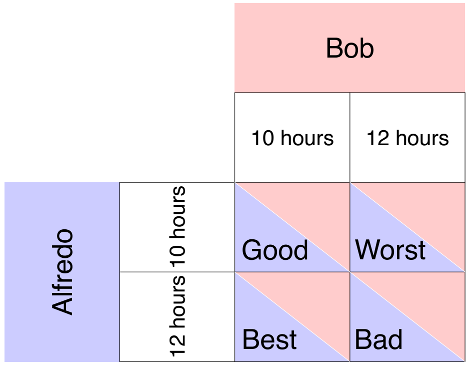


Figure 1.7: **Alfredo's payoffs to fishing more or less depend on how much Bob fishes.** Alfredo's payoffs are described using the words like we used for the coordination problem: Planting in Palanpur. Alfredo ranks his outcomes from best to worst: Best > Good > Bad > Worst. Alfredo's strategies and outcomes are highlighted in Blue. Bob's strategies and outcomes are highlighted in pink (but we have not put the words to describe Bob's outcomes in the figure).

Here we illustrate the basic concepts of game theory in a game we call the Fishermen's Dilemma. We chose the name because it is an example of what is probably the most famous game, the Prisoners' Dilemma.

The Fishermen's Dilemma game is non-cooperative, which for two people fishing in the same lake may seem unrealistic because as neighbors they might be able to come to some kind of agreement about what each will do. We do not consider this option in the two-person case because the model illustrates a large number of people interacting. When many people interact arriving at and enforcing such a cooperative agreement would present serious challenges.

Here is the game.

- *Players:* Alfredo and Bob, two fishermen.
- *Strategy sets*: Each may fish for either 10 or 12 hours.
- *Order of play*: They simultaneously select a strategy, resulting in the game's strategy profile
- *Payoffs*: The players catch and eat an amount of fish given by the strategy profile they have implemented.

This ends the game.

Payoffs

The payoff of each player is composed of two parts:

- The amount of fish they are able to catch and consume, which they value and would like to increase; and
- The amount of time they spend fishing, which they find tiring and would like to decrease.

We can describe the fishermen's interaction in the form of a *payoff matrix*.

We first present a version of the payoff matrix with words to represent Alfredo's payoffs (but not yet Bob's) in 1.7. Read the table this way: If Bob fishes 12 hours (the right hand column) and Alfredo fishes 10 hour (top row) this is the *worst* outcome for Alfredo. A payoff matrix presents hypothetical 'if-then' information; it presents all of the possible sets of payoffs, whether or not each is likely ever to occur.

M-CHECK A *matrix* is a rectangular array of quantities or other quantitative information).

The complete payoff matrix for the Fishermen's Dilemma is represented in Figure 1.9 with numbers indicating the two fishermen's evaluation of how good the outcome indicated is. So for example the payoff to each if they both fish ten hours (3) is fifty percent greater than if they both fish twelve hours (2).

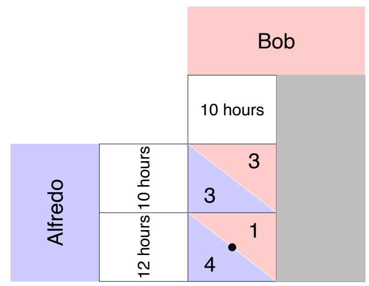
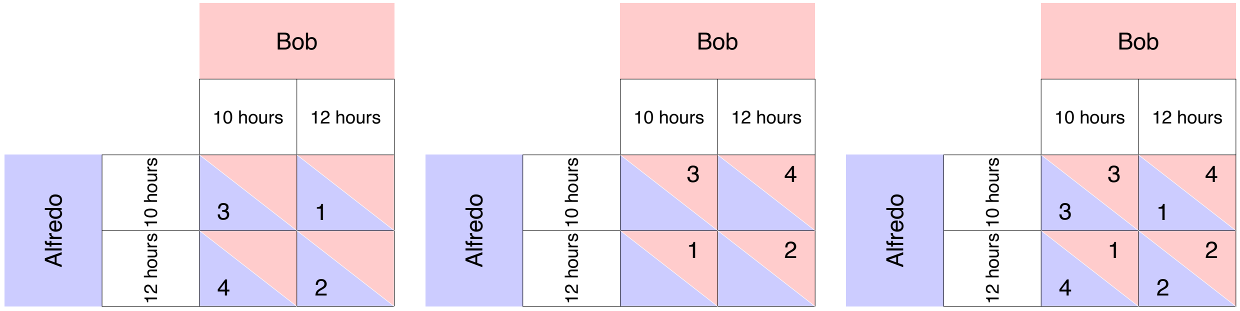


Figure 1.8: **Imagine Bob playing Fish 10 hours to assess which of Alfredo's strategies is a best response.**

The convention we will use throughout this book is to list the row player's payoffs first and in the bottom left corner of the cell and the column player's payoffs second in the top right corner. So, in the Fishermen's Dilemma game, we list Alfredo's payoffs first and Bob's payoffs second. We shade each players payoffs to make them easier to differentiate: blue for the row player (Alfredo) and red for the column player (Bob).

Many of the games in this book involve two players and each player has two possible strategies. We often call a game like this a "2 $\times$ 2" game (a "two-by-two" game). We now have all the elements we need for the complete description of the Fishermen's Dilemma and its strategy profiles and associated payoffs.

- *Alfredo fishes 12 hours, Bob fishes 12 hours*: When both fishermen fish 12 hours, they each catch fewer fish per hour of work, while they also have a higher cost of effort because they've spent a lot of time fishing. Each fisherman ends up with 2.
- *Alfredo fishes 10 hours, Bob fishes 10 hours*: When both fishermen spend less time fishing they catch a decent amount of fish and they haven't fished so long that the other fisherman catches fewer fish. They also benefit from a lower cost of time spent fishing. Each gets a net benefit of 3.
- *Alfredo fishes 10 hours, Bob fishes 12 hours*: Because Bob fishes 12 hours, Al catches many fewer fish and because Bob still fishes for another two hours, he catches a lot of fish while Al doesn't fish. Consequently, with the cost of time and catching fewer fish, Al ends up with net benefits of 1 and Bob ends up with net benefits of 4.
- *Alfredo fishes 12 hours, Bob fishes 10 hours*: This is symmetrical to the previous description, so now Al gets net benefits of 4 and Bob gets net benefits of 1.



(a) A's payoffs only (b) B's payoffs only (c) Payoffs for both players

Figure 1.9: **Payoffs of players in the Fishermen's Dilemma.** Alfredo's payoffs are in the bottom-left corner of each cell and are shaded blue. We include Alfredo's payoffs in the right-hand and left-hand panels. Bob's payoffs are in the top-right corner of each cell and are shaded red. We include Bob's payoffs in the center panel and the right-hand panel.

Checkpoint 1.6: Payoff matrixes

a. Fill in the blank pink triangles showing Bob's payoffs in Figure 1.7 using the payoffs shown in Figure 1.9.
b. You can now see that the cartoon in Figure 1.3 is a payoff matrix. What are the main differences in the payoffs of the Planting in Palanpur game and the Fishermen's Dilemma in Figure 1.9?

1.7 Predicting economic outcomes: The Nash equilibrium

As you already know, to predict a game outcome – the strategy profile that will result– we need more than the description of the game alone. We need to add what is termed a solution concept – a statement about *how* players will behave in the game – that can be the basis of a *prediction* of the game's outcome. Predicting the outcome of a game – based on the rules of the game and the solution concept – is especially important if we are evaluating policies to improve the functioning of the economy by changing the rules of a game so as to change the outcome of a game.

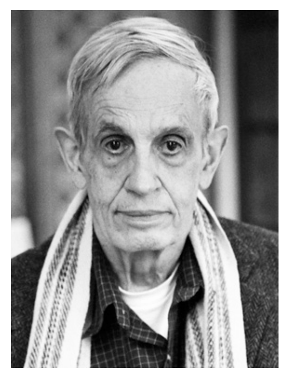


Figure 1.10: **John F. Nash (1928-2015)** was an American mathematician who contributed the to the theory of bargaining and to the concept of equilibrium that bears his name. He won the Nobel Prize in economics in 1994. His life was documented in the book and movie *A Beautiful Mind*.[9] Source: Peter Badge, Wikimedia Commons.

Equilibrium and prediction

The key idea on which a solution concept is based is **equilibrium**. An equilibrium is a state in which there is nothing in the situation that will cause the state to change. A *predicted outcome* will be an equilibrium, that is, an outcome that is stationary (not changing). To understand why, imagine this were not the case. You make a prediction, but then the outcome changes. Your prediction would *no longer be true* because the outcome had changed.

EQUILIBRIUM An equilibrium is situation that is stationary (unchanging) in the absence of a change external to the model.

Applying this reasoning to games, if we were to predict the outcome of a game to be a strategy profile under which one or more players would have reason to change their strategy, then the prediction would be falsified as soon as they carried out the change. So the status of stationarity – change-less-ness – is a property of a prediction; and this is why equilibrium is fundamental to making predictions about game outcomes.

Think of a concrete example. Suppose you want to predict where a marble will be if all that you know is that it is going to be somewhere in a round bottomed salad bowl sitting on a table. If I predicted that the marble would be somewhere halfway up the side of the bowl you would doubt my prediction. The reason is that any marble in that position would move downward in the bowl, that is, its position would not be stationary so, if it ever were (for some reason) where I predicted it would be, it would not be there any longer. It is not that the prediction would necessarily be wrong. It could be true for a millisecond after I placed the marble in the bowl just above my predicted spot, for example.

The only predicted position in the salad bowl that would not immediately falsify itself in this sense is the bottom. So a reasonable prediction of the location of the marble would be "the bottom of the bowl."

There are some situations in which a prediction based on an equilibrium would be likely to be incorrect. Change the marble-in-bowl example by filling the bowl with very thick honey. Then if you were asked to predict where the marble would be found, you would want to know how long it had been in the bowl, did have time to reach the bottom? If the marble had been placed in the bowl just a second ago, then you might be better off predicting that it would be where it had been placed, rather than the bottom of the bowl.

The marble-in-bowl-of-honey is often a better illustration of how economic processes work than the initial example. Markets are often out of equilibrium. Predicting things in motion is a much more challenging task than predicting them when they are stationary. We provide an example in a model of residential the segregation (Section 1.15) where we follow the process of change step by step. But for the most part we study equilibria and how to change them so as to improve outcomes.

In the marble-in-bowl illustration (without the honey) what is the solution concept that let us arrive the "bottom of the bowl" prediction? It is gravity, which is our understanding about a reasonable way for the marble to "behave." In modeling an economic interaction, the game structure is analogous to the salad bowl. What is the analogy to gravity? The answer is the players' *best response.*

Best-response strategies

BEST RESPONSE A strategy is a player's best response to the strategies adopted by others if no other strategy available would result in higher payoffs.

By far the most widely-used solution concept, the Nash equilibrium, is based on the idea that players choose **best-response** strategies; they do the best they can given the strategies adopted by everyone else

To understand better what a best response is, think about Alfredo in the Fisherman's Dilemma and imagine each of the possible situations that might occur and what would be best for him in each of these hypothetical situations.

First, what strategy should Alfredo adopt in order to gain the highest payoff if Bob were hypothetically to play Fish 10 hours as shown in Figure 1.8. We do not ask why Bob would do this. We are mapping all of the possible situations that Alfredo might encounter.

- Against Bob playing Fish 10 hours, Alfredo can get a payoff of 3 for fishing 10 hours or a payoff of 4 for Fishing 12 hours.
- $4 > 3$ therefore fish 12 hours is Alfredo's *best response* to Bob playing fish 10 hours.
- Place a solid dot in the cell (Alfredo plays Fish 12 Hours, if Bob plays Fish 10 hours) to indicate that it is Alfredo's best response. We will use this "circle and dot" method to find the Nash equilibrium.

Let's repeat the analysis and imagine Bob playing Fish 12 hours, as shown in Figure 1.11.

- Against Bob playing Fish 12 hours, Alfredo can get a payoff of 1 for playing Fish 10 hours or a payoff of 2 for playing Fish 12 hours.
- $2 > 1$ therefore Fish 12 hours is Alfredo's *best response* to Bob playing fish 10 hours.
- place a solid dot in the cell (Alfredo plays Fish 12 Hours, Bob plays Fish 12 hours) to indicate that it is Alfredo's best response.

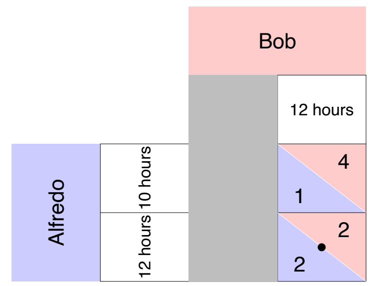


Figure 1.11: **Imagine Bob playing Fish 12 hours to assess which of Alfredo's strategies is a best response.**

Checkpoint 1.7: A best response for Bob
Repeat the process we went through for Alfredo, but do it for Bob instead. Notice that when you do so, you will blank out a *row* for Alfredo to imagine him playing the strategy in the other row, whereas you blanked out a *column* for Bob. What are Bob's best responses? Show his best responses using a hollow circle.

M-CHECK: STRONG AND WEAK BEST RESPONSE. A best response may be either *strong* or *weak*. A strong (also called strict) best response yields higher payoffs than any other: it is strictly "better" than any other strategy. There can be no strategy that is better than a weak best response but a weak best response need not be better than any other; it may be "as good as" (the payoffs to the strategy and some alternative strategy being equal.)

Nash equilibrium and the outcome of a game

Some games do not have a Nash equilibrium and you will see shortly that some have more than one.

EXAMPLE The Rock, Paper, Scissors game (also called ro-sham-bo and by many other names in other languages) originated in China about two thousand years ago. It does not have a Nash equilibrium.

NASH EQUILIBRIUM A Nash equilibrium is a profile of strategies – one strategy for each player – each of which is a best response to the strategies of the other players.

Using the best responses of the players we can now predict the outcome of a game using as our solution concept the **Nash equilibrium.** A Nash equilibrium is a profile of strategies – one for each player – each of which is a best response to the strategies of the other players. A Nash equilibrium is also called a *mutual best response*. Because at a Nash equilibrium all players are playing their best response to all of the others, it follows that no player

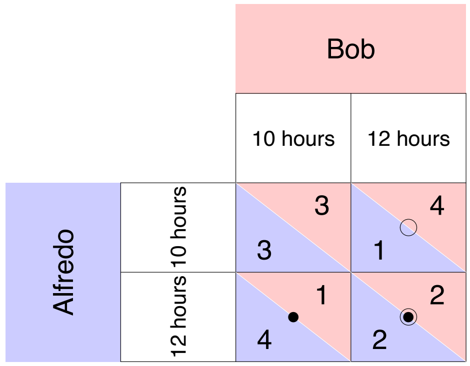


Figure 1.12: **Payoff matrix for the Fishermen's Dilemma.** The solid dots indicate Alfredo's best responses. The hollow circles indicate Bob's best responses. A Nash equilibrium is a cell that contains both. In this case there is just one Nash equilibrium: both fishing 12 hours.

has a reason to change his or her strategy as long as the other players do not change theirs. In Figure 1.12, Alfredo's best responses are shown by the solid black dot in the cell. Bob's best responses are shown by the hollow circle. Their best responses coincide at the Nash equilibrium (Fish 12 Hours, Fish 12 Hours) with payoffs (2, 2) shown in the cell where the solid dot is inside the hollow circle. You can use the "dot and circle" method to find one or more Nash equilibria (if they exist) for games that can be represented by a payoff matrix like Figure 1.12.

The outcome demonstrates how Nash equilibrium can initially seem counter-intuitive. Both would have had higher payoffs if they could have agreed to restrict their fishing to 10 hours (they could have had 3 each if they both fished 10 hours and $3 > 2$). But suppose both were restricting their fishing to 10 hours; then each would an incentive to fish for 12 hours (because $4 > 3$) and unless they had a binding agreement to continue fishing less, each would choose to fish more.

The Fishermen's Dilemma is therefore a coordination problem and it returns us to the classical institutional challenge. Without institutions to align the individual interest of the participants with their shared interest, they get an outcome that is worse for both of them than other possible outcomes. We will later show how a change in the institutions regulating how Alfredo and Bob interact – that is, changing the rules of the game – might address this coordination failure.

Checkpoint 1.8: Nash equilibrium

a. Explain why none of the other three outcomes (those that are not (Fish 12 Hours, Fish 12 Hours)) of the Fishermen's Dilemma satisfy the definition of Nash equilibrium.

b. At each of the other three outcomes, which player has an incentive to change strategy and in what way? Explain.

c. Explain why a game like Rock Paper Scissors would not be much fun if there was a Nash equilibrium.

Dominant strategies

In the Fisherman's Dilemma (and all Prisoners' Dilemmas) there is a single strategy that yields the highest payoffs to a player independently of which of the strategies the other player adopts. A strategy is a player's **dominant strategy** if it is the player's best response to all possible strategy profiles of the other player or players. That is, a strategy is a dominant if by playing it the player's payoff is *greater than or equal to* the payoff they would get by playing any other strategy for every one of the other player's profiles of strategies.

DOMINANT STRATEGY A strategy is dominant if it yields the highest payoff for a player, for any strategy chosen by the other players. Weak dominance refers to the case where there is another strategy yielding the same payoff.

Likewise we say that strategy A is *dominated* by another strategy B if the payoff to playing B is at least as great or greater than playing A for every strategy profile of the other players. If there is a strategy that dominates all of the other strategies that an player may choose, then it is a dominant strategy. If each player in a game has a dominant strategy, then the strategy profile in which all players adopt their dominant strategy is called a **dominant strategy equilibrium**.

We can apply the concept of *dominant strategy* equilibrium to the Fishermen's Dilemma. To do so, we need to understand whether each player has a dominant strategy.

- When Alfredo fishes 10 hours, his payoff is 3 if Bob fishes 10 hours and 1 if Bob fishes 12 hours.
- When Alfredo fishes 12 hours, his payoff payoff is 4 when Bob fishes 10 hours and 2 when Bob fishes 12 hours.
- So, when Bob fishes 10 hours, fishing 12 hours gets Alfredo a higher payoff $(4 > 3)$ and when Bob fishes 12 hours, fishing 12 hours gets Alfredo a higher payoff $(2 > 1)$
- Therefore, Alfredo gets a higher payoff from fishing 12 hours against each of Bob's strategies
- Fish 12 hours is therefore Alfredo's *dominant strategy*.

Fishing 12 hours is also Bob's dominant strategy. Because each player has a dominant strategy to fish 12 hours, the *dominant strategy equilibrium* is (Fish 12 hours, Fish 12 hours) with payoffs (2, 2). The dominant strategy equilibrium of a game is always a Nash equilibrium.

DOMINANT STRATEGY EQUILIBRIUM A dominant strategy equilibrium is a strategy profile in which all players play a dominant strategy.

The fact that the Fishermen's Dilemma has a dominant strategy equilibrium makes it a particularly simple problem for us, studying it. It also makes it simpler for Bob and Alfredo because what is best for each does not depend on what the other does. But this does not mean that they will be happy with the result.

Checkpoint 1.9: Dominance and Nash equilibrium

a. Repeat the analysis we did for Alfredo for Bob and confirm that 12 hours is a dominant strategy for him too.
b. We said that a dominant strategy equilibrium is always a Nash equilibrium. But do you think that a Nash equilibrium is always a Dominant Strategy equilibrium? Why or why not?

1.8 Evaluating outcomes: Pareto comparisons and Pareto efficiency

The Nash equilibrium can help us predict the result of an interaction. But it does not tell us anything about whether some outcome is good by any standard, or even better or worse than some other outcome. Economists, policy-makers and others would like to evaluate whether some outcomes are better or worse so that we can try to work out which rules of the game would make the better outcomes Nash equilibria, and therefore more likely to be what we observe.

The challenge in making these comparisons is that whether some outcome is better than another depends on what you value. There is no agreed upon standard of what makes one outcome better than another. Returning to our fishermen, here are some of the values that we could use to evaluate an outcome:

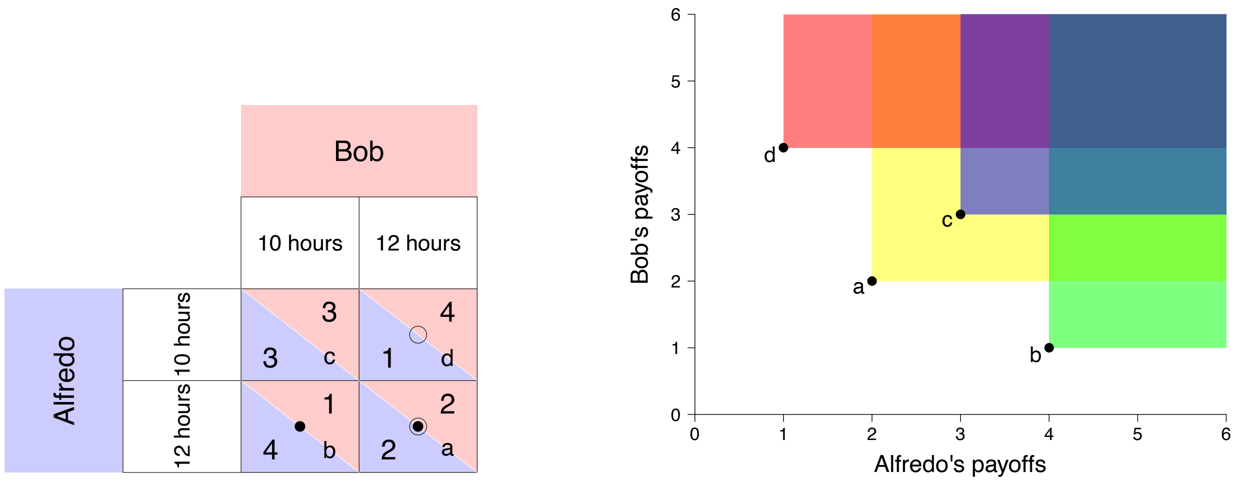
- *Fairness in the distribution of payoffs* among the players; is it fair that Alfredo receives 4 times what Bob gets when Alfredo does *not* limit his fishing hours and Bob does?
- *Are the rules of the game itself fair?* In the Fishermen's Dilemma the same rules applied to both players; but were the game a bit different, many would think it unfair if Alfredo could simply order Bob to fish 10 hours, or to hand over half of all the fish Bob caught.
- Setting aside fairness, is the outcome a *reasonable use of available resources* including the working time of the two fishermen and the sustainability of the lake itself and the living things that it supports.

There are many other standards that could be proposed.

A concept that is widely used to evaluate economic outcomes involving 2 or more people is called **Pareto efficiency**. The idea is simple: an objective of public policy and institutional design – the rules of the game – should be to avoid those outcomes – like traffic jams, planting late in Palanpur, and overfishing the lake – that are worse for everyone, compared to an alternative outcome that also would have been feasible.

PARETO EFFICIENCY An allocation with the property that there is no alternative technically feasible allocation in which at least one person would be better off, and nobody worse off.

PARETO-SUPERIOR Outcome A is Pareto-superior to outcome B (it Pareto-dominates outcome B) if, in outcome A, at least one player is better off than in outcome B without anyone being worse off.



(a) Fishermen's Dilemma with labeled points

(b) Analyzing points for Pareto efficiency

Figure 1.13: **Three Pareto-efficient outcomes of the Fishermen's Dilemma**. Panel a is the same as Figure 1.12 except that each of the four squares in the payoff matrix has been assigned a letter. In Panel b we show the Fishermen's Dilemma indicated by the payoffs of the two at the four possible outcomes given by the same letters that appear in each of the cells of the payoff matrix. We use shaded colors indicating 90 degree angles to the northeast of the feasible outcomes (each of the lettered points).

Pareto comparisons

Pareto-efficiency is based on **Pareto comparisons** of outcomes. Consider two outcomes, A and B, with resulting payoffs for two or more players. Outcome A is Pareto superior to outcome B if in outcome A at least one player is better off than in outcome B without anyone being worse off. A change in the outcome from a Pareto-inferior situation like B to a Pareto-superior outcome like A is called a *Pareto-improvement*. This is a Pareto comparison. An outcome is Pareto efficient if no other feasible outcome is Pareto superior to it.[10]

Figure 1.13 depicts the outcomes of the interaction between Alfredo and Bob. Figure 1.13 a is the Fishermen's Dilemma payoff matrix with each outcome given a label **a**, **b**, **c**, or **d**. These payoffs are indicated by points in Figure 1.13 b where you can read on the vertical and horizontal axes the payoffs to the two players that you see in the payoff matrix.

The Pareto comparison is easy to see in this type of plot. An outcome A is Pareto-superior to another outcome B if the point indicating the payoffs from A lies to the "north-east" of point indicating the payoffs from B. "North-east" in this figure is "better for both." So looking at the colored areas whose lower left corners are points **a, b, c,** and **d**, then a Pareto-efficient point is one that has no other point in its "colored shadow" extending upwards and to the right of the point. By this standard, three of the points – **b**, **c**, and **d** – are Pareto efficient, while **a** is not, because point **c** is in the yellow "color shadow" of point **a**.

We say that two outcomes can be Pareto compared, or Pareto ranked, if one of them is Pareto superior to the other. But as you can see from the figure Pareto comparisons (or rankings) are often not possible. Specifically, when

HISTORY: PARETO EFFICIENCY The idea is attributed to the Italian economist and sociologist Vilfredo Pareto (Pa RAY to) but it was first introduced by the British economist, Francis Edgeworth. However, an important concept that we use extensively in later chapters, called the Edgeworth Box, was actually invented by Pareto.[11]

two outcomes are both Pareto-efficient, they cannot be Pareto-compared or Pareto-ranked. We could rank **c** above **a** because both players were better off, but with **b**, **c** and **d** we cannot move from one outcome to another without worsening outcomes for at least one of the players.

Checkpoint 1.10: Pareto improvements in the Fishermen's Dilemma

Referring to Figure 1.13 do the following:

a. Is any point dominated by some other point? Say which, if any?
b. At which point is the total payoff of the two fishermen the greatest?
c. Would a change from any other point to that "total payoff maximum" point be a Pareto improvement?
d. Explain what you think is the meaning of the expression "there is no such thing as a free lunch" and say whether this saying is true in Figure 1.13 (at all of, some of, one of the, none of the points.)

1.9 The value and limitations of Pareto efficiency

Pareto efficiency gives us a way to identify "lose-lose" outcomes we should seek to avoid, namely those "that are worse for all than they could be." But except in special cases, Pareto efficiency does not provide a rule to select what we might call "the best" outcome.

To see why this is true, suppose we have a cake and we are dividing it among people, all of whom enjoy eating cake. An outcome in which one person gets the entire cake is surely Pareto-efficient because in any other allocation that lucky person would get less. Likewise an allocation in which everyone got the same sized slice of the cake is Pareto-efficient, for in any other allocation at least one person would have to get less.

Pareto efficiency is *not* about how something of value should be divided up. All it says is "make sure there's no cake left on the table!"

Most economic problems that we face are similar to the cake example in that there are a great many Pareto efficient outcomes. Think about the Fishermen's Dilemma game: all of the possible outcomes of the game *except one* are Pareto efficient. When there are many Pareto-efficient outcomes there is a *conflict of interest* among players over which Pareto-efficient outcome they would prefer. We cannot say that one is "more Pareto-efficient" than the other.

It is also perfectly sensible to prefer an outcome that is not Pareto efficient but is more fair over an alternative Pareto efficient outcome that is unfair. To continue the cake example, if there are two people between whom the cake will be divided many people would reject the (Pareto efficient) outcome in which one person gets the entire cake in favor of a Pareto inefficient alternative in which each gets a third of the cake (the remaining third perhaps being thrown away or destroyed in the conflict over its distribution). So we would prefer a Pareto-inefficient outcome over a particular efficient outcome (one person gets the whole cake). But the Pareto comparison would remind us that each person getting half of the cake is preferable to each getting a third with the rest being wasted.

Pareto efficiency is a useful device for screening out those outcomes (like throwing away some of the cake in the above example, or planting late in Palanpur, or over-fishing the lake) that should not be among the list of candidate feasible outcomes among which the choice of better or best should be made or grounds of fairness or other bases.

Checkpoint 1.11: Pareto efficiency

Consider these questions about Pareto efficiency.

a. True or False (and explain): "The fact that an outcome is Pareto-efficient does *not* imply that it is preferred by all the actors to all the other outcomes."
b. Can two Pareto efficient outcomes be Pareto compared? Why or why not? Explain.
c. Imagine you are an impartial observer evaluating the possible outcomes that might occur for Bob and Alfredo. Are there any reasons why you might judge the outcome **a** in the figure to be better than the Pareto efficient outcomes **b** and **d**, despite the fact that **a** is Pareto-inefficient?

1.10 Conflict and common interest in a Prisoners' Dilemma

PRISONERS' DILEMMA In a 2-by-2 game, a Prisoners' Dilemma is a social interaction in which there is a unique Nash equilibrium (that is also a dominant strategy equilibrium), but there is another outcome that gives a higher payoff to both players (and a higher total sum of payoffs than any other outcome), so that the Nash equilibrium is not Pareto-efficient.

You know that the game the fishermen are playing is a particular case of the **Prisoners' Dilemma**. We now point out some of the general characteristics of this particular kind of coordination problem.

A Prisoners' Dilemma is a two-person interaction in which there is a unique Nash equilibrium (that is also a dominant strategy equilibrium), but there is another outcome that gives a higher payoff to both players, so that the Nash equilibrium is not Pareto-efficient. This means that in the Prisoners' Dilemma both players get their second worst payoffs in the game by playing their strictly dominant best-response strategies.

In Figure 1.14 we show the familiar payoff matrix for the Fishermen's Dilemma, but instead of the numbers indicating the payoffs of the players now we label the payoffs w, x, y, and z. We label the fishing 10 hours strategy "Cooperate" because it is the mutually beneficial action the two fishermen could take if they could coordinate their actions. The strategy fish 12 hours is labeled "Defect" because choosing to fish 12 hours instead of 10 is deviating from a mutual cooperate outcome on which the two fishermen might be able to coordinate.

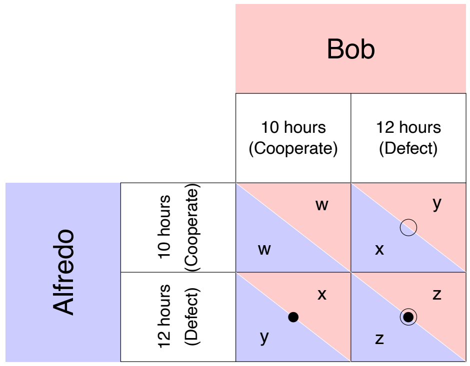


Figure 1.14: **A general Prisoners' Dilemma**. For the game to be a Prisoners' Dilemma, we require $y > w > z > x$ and $2w > y + x$ (this is like $4 > 3 > 2 > 1$ and $2 \times 3 > 4 + 1$ from the numerical example).

The interaction is a Prisoners' Dilemma if two conditions hold:

- $y > w$ and $z > x$ means that fishing Defect is a strict dominant strategy
- $w > z$ means that mutual cooperation is Pareto superior to mutual defection.

For Alfredo, 12 Hours is a best response to Bob playing 10 Hours because $y > w$; 12 Hours is also a best response to 12 Hours because $z > x$ (both best responses are shown in Figure 1.14 by the solid dot). Similarly, for Bob, 12 Hours is a best response to Alfredo playing 10 Hours because $y > w$; 12 Hours is also a best response to 12 Hours because $z > x$ (both best responses are shown by the hollow circle). The dot-inside-circle in Figure 1.14 confirms that if the game is played non-cooperatively the Nash equilibrium is (12 Hours, 12 Hours) with payoffs (z, z) when by coordinating their choices they could have each received (w, w) (where $w > z$).

M-CHECK A third condition is sometimes added, namely $x + y < 2w$ which means that the sum of payoffs when both players cooperate is greater than the sum of payoffs when one cooperates and the other defects. This condition makes (Cooperate, Cooperate) preferable to any outcome in which one defects and the other cooperates.

Checkpoint 1.12: When is cooperation *not* the best they could do?

Show that if the condition $x + y < 2w$ is violated (i.e. if $x + y > 2w$) then the two players could do better by

a. one defecting on the other and then sharing their total payoffs equally or
b. if the game is played many times, alternating who cooperates and who effects.

Economic rent: The incentive to coordinate

Both players have a good reason to try to change the rules of the game so that they can agree on both cooperating. How much more they would get if they were to mutually cooperate than if they mutually defected – in this case $w - z$ – is called an **economic rent**, meaning the difference between the payoff that they would get if they cooperated and their next best alternative. Their next best alternative to cooperating, in this case, is mutual defection, also known as their **fallback position**.

Economic rents and the fallback position play a central role in microeconomic theory, so it is a good idea to master them. The meaning of the term fallback position is intuitive: it is what you *fall back* to if some particular outcome is not possible, in this case if the mutual cooperation should not work out. A player's **fallback position or fallback option** is the *payoff* they receive in their *next best alternative*.

FALLBACK POSITION A player's fallback position (or reservation option) is the payoff they receive in their next best alternative.

The term "economic rent" may at first seem surprising, because the word "rent" also means a payment for the temporary use of something to a landlord or a car rental agency. The term economic rent means something entirely different. A participant's *economic rent* is the payoff they receive in excess of what they would get in their fallback position.

ECONOMIC RENT A participant's economic rent is the payoff they receive in excess of what they would get in their fallback position.

We shall use the idea of a fallback often, from social interactions like the Prisoners' Dilemma, to worker-employer relationships where a worker's next best alternative may be unemployment and the rent she receives as an employee is the difference between her wage and the government transfer she would receive were she to lose her job. The next best alternative for a person applying to a bank for a loan is trying to get money from friends or family along with future obligations. As these examples indicate, the fallback position will differ depending on the details of the situation and what our modelling assumptions are designed to illuminate.

Impediments to coordination: Limited information and conflicts of interest

EXAMPLE The term economic rent is what is known in the study of language as a "false friend," a term that you think you know the meaning of but means something entirely different in the new language you are now learning. "Sensible" in English means "reasonable" but in Italian it means "sensitive."

If $w - z$ is substantial – meaning substantial rents associated with cooperation for each player – then it might seem a simple matter for the players to agree to cooperate. But people often fail to reach or enforce such an agreement, for two main reasons:

- *Limited information:* The participants may lack the information needed to monitor and enforce an agreement. How can a participant know or verify what other participants do?
- *Conflict over distribution of the economic rents from cooperation:* Disagreement about who gets what – for example who gets to fish more – may make it impossible for the two to agree.

Concerning the information problem, the fishermen, for example, may have no way of enforcing an agreement, or even knowing if the agreement has been violated. While each may know how many hours the other has fished on day with clear and sunny weather, on a foggy day it may be impossible to know. Even if one fisherman knows how much the other fished, that knowledge may be insufficient to enforce an agreement through a third party such as a court of law.

ASYMMETRIC INFORMATION Information is asymmetric if something that is relevant to the parties in an economic interaction is known by one actor and is not known by another.

This is the problem of **asymmetric information** or **non-verifiable informa-**

tion. Information is asymmetric if people know different things, or if what one person knows (for example how many hours he fished), the other person does not know. Information is not verifiable if people cannot use it to enforce an agreement or a contract. Asymmetric and non-verifiable information will play a central role in our analysis of how the labor market, the credit market and other markets work.For example most courts will not accept “hearsay” (meaning “second-hand”) information, so if one of the fishermen had heard from someone else that the other had fished 12 hours, this would be non-verifiable information.

NON-VERIFIABLE INFORMATION Information that cannot be used in legal proceedings to enforce a contract or other agreement.

Concerning the second problem for coordination, conflicts over the distribution of the economic rents from cooperation, in the Fishermen’s Dilemma, the agreement to restrict fishing to 10 hours a day divides the benefits of restricting fishing in a particular way, namely equally. But the fishermen need not agree on 10 hours each. Alfredo might insist that he will fish 12 hours and Bob only 10 hours. Or Bob might insist on the opposite.

Or Bob might insist that both fish 10 hours, but that Alfredo give him most of Alfredo’s catch, leaving Alfredo with just enough of his catch to be no worse off than had they both fished 12 hours, namely with a payoff of z. Which of these we will observe depends on rules of the game we have not yet introduced, including differences in the bargaining power exercised by the two. Unless they can find a mutually acceptable solution to the distribution problem they may end up having no agreement at all, and then simply fish at 12 hours each, at their fallback position.

The fishermen’s distribution conflict highlights a challenge that arises in any voluntary economic interaction. Consider their possible agreement to limit their fishing time:

- The agreement is voluntarily entered into. This means that neither player can force the other to accept terms worse than their fallback position.
- The agreement therefore must allow each participant to achieve a payoff greater than (or at least not worse than) had the individual not agreed to cooperate. In other words, there must be some economic rents made possible by a voluntary cooperative outcome.
- This being the case, the participants have to find a way that the total rents will be divided. If they are to agree to cooperate by restricting the total time they spend fishing, they must also agree on how these economic rents will be distributed.
- Conflict over the distribution of the economic rents (who gets what amount of economic rent) may prevent the fishermen from coming to an agreement.

FACT CHECK In the next chapter we will see that across many cultures of the world, people would rather get nothing than get what they consider to be an unfair share of the economic rents, and as a result cooperation breaks down and nobody gets any rent at all.

We sometimes think of cooperation and conflict as opposites, as for example when members of a team cooperate in their efforts to win some conflict with another team. But the Prisoners' Dilemma is a scenario of *conflict and cooperation among the very same people.* They have common interests in getting some share of the economic rents by cooperating; but they have conflicting interests in how the total will be divided into the rents received by each.

A catalogue of games: And their challenges to coordination

Some interactions present greater impediments to coordination than others; the Prisoners' Dilemma is in some respects the most challenging of all.

We can classify coordination problems and the challenges they present by the relation between Nash equilibria and Pareto-efficient outcomes of the games that represent them.

- In the Prisoners' Dilemma, you know, there is a unique Nash equilibrium that is Pareto-inefficient. Because this outcome is also a dominant strategy equilibrium, coordination on mutual cooperation will require some change in the rules of the game (making it a cooperative game) or a change in the player's payoffs, for example, if they dislike harming the other player by defecting on them.
- In interactions like Planting in Palanpur, which are often called *Assurance Games*, there are two Nash equilibria, (both Plant Early and both Plant Late) one of which (Plant Early) is Pareto-superior to the other (Plant Late). In these games if one of the players plays the strategy making up the Pareto superior equilibrum (Plant Early) then the best response of the other will be to do the same. Finding institutions that will implement the preferred plant early outcome in a game like this will be a lot less challenging than in a Prisoners Dilemma.
- Another important class of coordination problems arise in what we call *Disagreement Games* where there are two Nash equilibria each of which is Pareto-efficient, so that they cannot be Pareto-ranked, and players disagree about which Nash equilibrium they would like occur. These are like the Planting in Palanpur game but with the additional challenge stemming from a conflict over which Nash equilibrium will be implemented.

We start with an even less challenging game in which players' self interests lead them to a Pareto-efficient Nash equilibrium.

Checkpoint 1.13: Guilty prisoners

a. Referring to Figure 1.13 a, imagine that both Bob and Alfredo have become ethical and now would feel guilt if by defecting when the other cooperated. Should they do this, their guilt results in a subtraction from the payoff points shown in the figure.

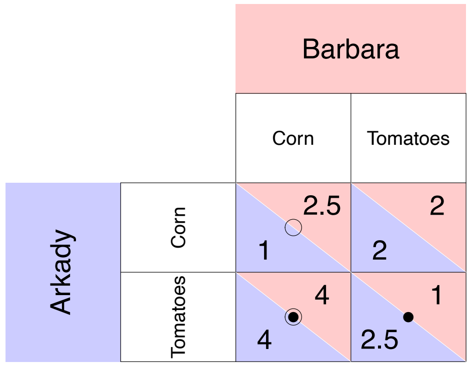


Figure 1.15: **An Invisible Hand Game with best responses indicated by circles and dots.** Arkady's payoffs are listed first in the bottom-left corner. Barbara's are listed second in the top-right corner. The game captures Adam Smith's ideas of *specialization* and *gains from trade* (that is, the opportunity to obtain economic rents from trade).

b. What is the smallest value of this guilt that each feels that would make Defect no longer be the dominant strategy?

c. If Alfredo but not Bob acquired this "defection guilt" so that for him defecting on a cooperator was no longer a best response, but Bob continued with the values (and payoffs) in the figure, is there a Nash equilibrium of the game, and if so, what is it?

1.11 Coordination successes: An Invisible Hand Game

INVISIBLE HAND GAME An Invisible Hand Game has a single Nash equilibrium that is Pareto efficient.

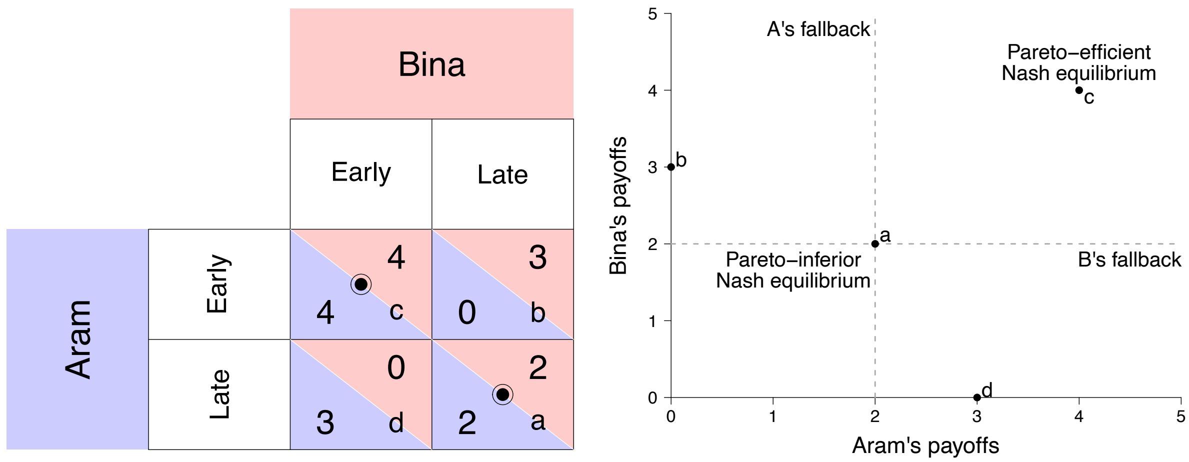
The characteristic of what we call an Invisible Hand Game is that it has a single Nash equilibrium that is Pareto-efficient. Apologies to Adam Smith: our game is much simpler than Smith's reasoning and Smith did not use ideas like Pareto efficiency. But our **Invisible Hand Game** illustrates Adam Smith's core insight that through the competitive buying and selling of privately owned goods on competitive markets, self-interested people can achieve outcomes to the benefit of all at least some in conditions (that we spell out in Chapter 14).

Consider a 2-by-2 game with two players, Arkady and Barbara, both farmers. Each player can choose one of two strategies: planting corn or planting tomatoes. The payoffs that they assign to the various outcomes of the game are provided in the matrix in Figure 1.15, which we call the Corn-Tomatoes game.

The payoff matrix reflects two facts about the problem that the two farmers face.

- Either because of their skills or the nature of the land they own, Arkady is better at growing tomatoes; Barbara is better at growing corn
- They both do poorly when they produce the same crop because the increased supply of whichever good it is that they both produce drives down the price.

The Nash equilibrium of the Corn-Tomatoes game is (Tomatoes, Corn), that is,



(a) Assurance Game with labeled payoffs

(b) Payoffs plotted against each other

Figure 1.16: **Planting in Palanpur: An Assurance Game.** Aram's payoffs are listed first in the bottom-left corner. Bina's payoffs are listed second in the top-right corner. Aram's best responses are shown by the solid point and Bina's are shown by the hollow circle. The Nash equilibria of the game are (Plant Early, Plant Early) and (Plant Late, Plant Late), with payoffs (4, 4) and (2, 2). The Plant Early Nash equilibrium is Pareto-efficient. The Plant Late equilibrium is not. In the right-hand panel, the payoffs are plotted against each other. Aram's payoffs are plotted on the horizontal axis, increasing as you move rightward. Bina's payoffs are plotted on the vertical axis, increasing as you move upward.

Arkady plants tomatoes, and Barbara plants corn, at which the players receive payoffs (4,4). (Tomatoes, Corn) is Pareto efficient as there is no alternative outcome which is Pareto superior to it. This is the best they could do. There was no need for them to explicitly agree on how to coordinate to achieve this result.

Just as in Adam Smith's reasoning about his invisible hand, Arkady and Barbara, are in a situation in which by simply following their self-interest they coordinate to their mutual benefit.

Checkpoint 1.14: Invisible Hand Game

Which entries in the payoff matrix would you have to compare in order to show the following:

a. They each do better when Arkady specializes in tomatoes and Barbara specializes in corn then vice versa.
b. They each do worse when both produce the same crop.
c. Growing corn is Barbara's dominant strategy
d. Arkady growing tomatoes and Barbara growing corn is the dominant strategy equilibrium.
e. Explain why the Nash equilibrium of the game is Pareto efficient.

1.12 Assurance Games: Win-win and lose-lose equilibria

Return to the farmers in Palanpur. There are two Nash equilibria in this game, one in which both participants Plant Early and one in which both Plant Late. The best response to the other farmer planting early is also to plant early, while the best response to the other farmer planting late is also to plant late. The outcome where both farmers plant early is Pareto-superior to the outcome when both farmers plant late.

The players do not have any conflict of interest: both would share equally in the gains to cooperation, should they find a way to coordinate on planting early. The problem for the real life farmers of that village is that they are stuck in the Pareto-inefficient Nash equilibrium of what is called an Assurance Game. Their challenge is how move to the Pareto-superior Nash equilibrium.

This could happen if all the participants had confidence (were *assured*) that the other participants also play the strategy yielding superior outcome. This is why it is an "Assurance Game."

Figure 1.16 is the payoff matrix for two players, Aram and Bina, choosing when to plant their millet in the village of Palanpur, India. (It is the same as the earlier figure about the two farmers, except that we now have numbers representing the farmers' payoffs). Coordination failures arise in the Assurance Game because of *positive feedbacks*: if one plants late the more is the incentive for the other to plant late, and vice versa. The strategies are **strategic complements**.

Checkpoint 1.15: Graphing Palanpur

a. Using the graphical method for identifying Pareto-efficient outcomes as shown in Figure 1.16, show which outcomes in the Palanpur game are Pareto-efficient.
b. Can you explain why **a** and **c** are Nash equilibria?

ASSURANCE GAME An Assurance Game is a 2-person, symmetric, strategic interaction with two strict Nash equilibria, one of which is Pareto superior to the other.

STRATEGIC COMPLEMENTARITY Strategic complementarity exists when a) A strategy is a strategic complement to itself: The payoff to playing a particular strategy increases as more people adopt that strategy as a result of some form of positive feedbacks, or b) One strategy and another are strategic complements to each other. In this case, for two activities A and B, the more that A is performed, the greater the benefits of performing B, and the more that B is performed, the greater the benefits of performing A.

Assurance game and strategic complementarity

Social media, dating platforms, and other matching services are examples of strategic complementarities. They are more valuable to everyone if many people participate.

Strategic complementarity exists when either of two conditions hold.

1. *A strategy is a strategic complement to itself*: The payoff to playing a particular strategy increases as more people adopt that strategy as a result of some form of *positive feedbacks*. Plant Early in Palanpur is an example Dating platforms are another. The strategy could be "Open a dating app account." The positive feedback arises because the more other people that are using the dating app the more people you will "meet" (which is better for you and better for them). Tinder, Bumble, OkCupid, Hinge, Grindr, and other dating apps are social platforms illustrating what are called **network externalities** or *network external effects* which occur when the benefits to members of a social or physical network increase when more people join

the network. In this case the strategy of joining the network is a strategic complement to *itself*.

2. *One strategy and another are strategic complements to each other.* The payoff to playing one strategy (say, A) is greater the more people adopt the other (B). In this case we say that strategies A and B are *strategic complements*. An example is the Invisible Hand Game shown in Figure 1.15. The payoff to Arkady from planting tomatoes is greater if Barbara plants corn (instead of tomatoes), and the payoff to Barbara from planting corn is greater if Arkady plants tomatoes (instead of corn). Growing corn and growing tomatoes are strategic complements.

We predict and evaluate the possible outcomes of the Planting in Palanpur Game using the concept of best response (using the dot and circle method introduced earlier). We see that the game has two Nash equilibria (Early, Early) with payoffs (4, 4) and (Late, Late) with payoffs (2, 2). (Early, Early) is Pareto-superior to (Late, Late) and it is Pareto-efficient because no alternative outcome is Pareto-superior to (Early, Early).

Even though there is a Pareto efficient Nash equilibrium, a population – like the people of Palanpur – may get stuck in the Pareto inferior Nash equilibrium. That does not guarantee players will actually play it. From the Assurance Game we have learned two things applicable across many kinds of social interaction:

- The fact that a Pareto efficient outcome is a Nash equilibrium does not mean that it will be the one we observe; getting there is not assured, and

- In cases where there is more than one Nash equilbrium, we need more information than is provided by the Nash equilibrium and Pareto efficiency concepts to make a prediction about the strategy profiles we will see in practice.

The second requirement to make a prediction – called equilibrium selection – becomes a serious challenge is cases where, unlike the Assurance Game, the players *disagree* about which equilibrium they would like to occur. This is the case in the next game in our catalogue.

Checkpoint 1.16: Assurance Game

Which payoff table entries would you have to compare in order to show that:

a. Planting early is Pareto efficient.
b. Planting late is a Nash equilibrium.
c. The best response to the other planting early is to plant early.

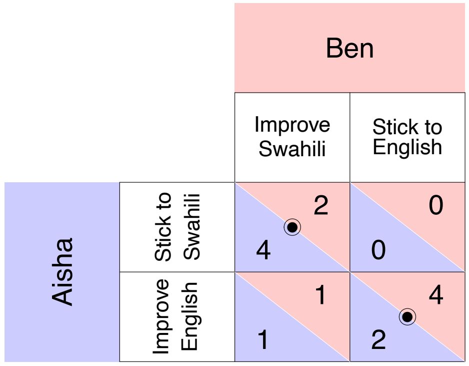


Figure 1.17: **A Disagreement Game**: The players need to coordinate on an equilibrium, but each prefers one equilibrium to the other, so there is a *conflict of interest*. If they fail to coordinate on one of the Nash equilibria because of the conflict of interest, the outcome will be a *coordination failure*.

1.13 Disagreement Games: Conflict about how to coordinate

An example of what we call a **Disagreement Game** is illustrated in Figure 1.17. In a Disagreement Game there are two Pareto-efficient Nash equilibria and the players are in conflict over which Nash equilibrium each prefers. But both of the players prefer both of the Nash equilibria to the *other* outcomes. The players' problem is to find a way to *coordinate* on one of the Nash equilibria to ensure that *no* coordination failure results. They do *not* want to end up at an outcome where both of them do worse than at one of the Nash equilibria.

Consider two players, a home-language Swahili-speaker (Aisha) and a home-language English-speaker (Ben) who have recently met. Each person can speak the other language, but prefers to speak their home language. They share many common interests but do not communicate as well as they would like. Each has two strategies: Stick to your home language or Improve the other language.

Among the possible outcomes are that he could learn better Swahili and they could routinely converse in that language; or she could learn better English and they could converse in English. They do not need to both be fluent in both languages.

So for Aisha, if Ben becomes fluent in Swahili, then her best response is not to take the time and trouble to improve her English. For Ben, similarly, if Aisha were to become fluent in English, then he would see little point in taking the Swahili courses.

The result is two Nash equilibria (Stick to Swahili, Improve Swahili) with payoffs (4, 2) and (Improve English, Stick to English) with payoffs (2, 4). The two Nash equilibria are both Pareto-efficient because there are no alternative outcomes which are Pareto-superior to these strategy profiles.

But, as shown in the payoffs in Figure 1.17, Aisha would prefer the (Stick to Swahili, Improve Swahili) Nash equilibrium and Ben would prefer the (Improve

English, Stick to English) Nash equilibrium.

The Disagreement Game is similar to the Assurance Game in that:

- There are two Nash equilibria
- Both players do better if they coordinate (that is, speak the same language at one or the other of these equilbria)

The Disagreement Game differs from the Assurance Game because:

HISTORY One of the first game theoretic studies of coordination problems – by David Lewis – was concerned with how we coordinate on a common language.

- Each player in the Disagreement Game prefers one of the Nash equilibria while the second player prefers the other, while both prefer the Pareto-superior Nash equilibrium in the Assurance Game, so as a result
- the players in the Disagreement Game have a *conflict of interest* concerning which equilibrium gets selected.

Disagreement Games highlight how there can be social interactions with multiple Nash equilibria, each of which is Pareto-efficient, but there may be no 'middle ground' to coordinate on and as a result conflict over who gets to benefit the most is unavoidable. Both players in the Disagreement Game would both be worse off out of equilibrium than at one of the Nash equilibria in the game. They have a common interest in coordinating somehow as opposed to *not* coordinating; but their interests conflict in *how* they coordinate.

Checkpoint 1.17: Language Game

Label the outcomes of the Language Game (Figure 1.17) like we did for the Prisoners' Dilemma Game in Figure 1.13. Plot the outcomes using axes with the players' payoffs, and determine which outcomes are Nash equilibria and which are Pareto-efficient.

1.14 Why history (sometimes) matters

As we have seen from Disagreement Games and Assurance Games, strategic complementarities in games may give rise to more than one Nash equilibrium. When this is the case we cannot say which Nash equilibrium is our prediction of how the game will be played. The best the Nash equilibrium concept could do is to say that the outcome of the game is likely to be one of the (perhaps many) Nash equilibria.

We need more information to make a prediction. Think about the Palanpur game, and imagine that all you know is the payoff matrix (not how the farmers played the game in recent years). Though you would be on solid ground predicting that it is likely that you'd see both farmers planting either early or late, you would not have much confidence in which it would be.

But now suppose you were told that last year they planted late. Then, unless they had discovered some way to coordinate a switch to planting early, you would be correct when you predicted that they would both be planting late this year too.

PATH DEPENDENCE A process is path dependent if the most likely state of something this period depends on its state in recent previous periods.

When history matters in this sense, we say that outcomes may be **path-dependent.** When the outcome of a game is path-dependent, knowing the recent history of a social interaction is valuable information to predict which equilibrium will occur. So, quite different equilibrium outcomes – poverty or affluence, for example – are possible for different groups of players with identical preferences, technologies, and resources but with different histories. This is how "history matters."

POVERTY TRAP A poverty trap occurs when identical people in identical settings may experience either an adequate living standard or poverty, depending only on chance events of their histories. Poverty in this case is a result of a person's circumstances, not personal attributes.

The Palanpur payoff matrix describes a **poverty trap**. A poverty trap occurs when identical people in identical settings may experience either an adequate living standard or poverty, depending only on chance events of their histories, for example were their parents rich or poor, or were they citizens of Norway or Nigeria. The possibility of poverty traps alerts us to the fact that people may be rich or poor not because of anything distinctive about their skills, hard work or other personal attributes, but because of the situation they find themselves in. Poverty may be inherited as it is in Palanpur not by anything that parents pass on to children but instead by the inheritance of a common history.

GROUP INEQUALITY Economic differences between sets of people distinguished by some common attribute – men and women and people of different nations, ethnic or racial groups – are called group inequalities.

The same is true about other aspects of how we interact in society, for example in the ways our lives may be highly segregated in interacting with people who differ in the groups with which they are identified, whether that be ethnicity, or religion, or even loyalty in sports teams.

Checkpoint 1.18: Drain the meadow: Name that game

a. Write down a payoff matrix for Hume's "drain the meadow" game, with the two actions open to farmers Adams and Brown being "drain" and "do not drain," and assuming that the value of the drained meadow (to each farmer) is 5, the value of the undrained meadow is 3, and if the two farmers jointly work on the draining it costs them 1 each, while if a single farmer does the draining alone it costs him 3.
b. What kind of game is this? Explain how it might be solved if there were just two farmers, and why with many farmers (as Hume wrote) it would be "difficult and indeed impossible" for them to agree on a common course of action and avoid in a coordination failure.

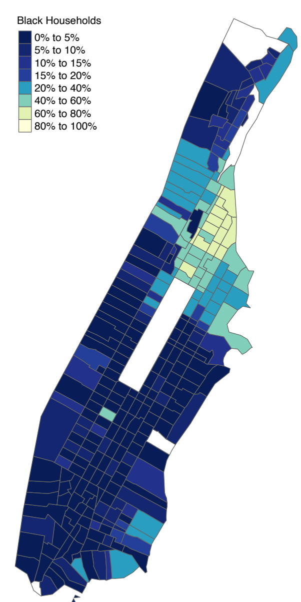


Figure 1.18: **Segregation in Manhattan, the central borough of New York City in 2018.** The rectangle without shading is Central Park Source: Sethi and Somanathan (2009) using updated 2018 block data from the American Community Survey, 2018.

1.15 Application: Segregation as a Nash equilibrium among people who prefer integration

Segregated communities – whether on grounds of ethnicity, race, religion, or class – often cultivate inter-group prejudice and hostility and are the basis for

(a) Household B has one neighbor of each group, their *ideal* situation, is *not dis-satisfied*, and will play Do Nothing

(b) Household B has two neighbors of their own group, is *not dis-satisfied*, and will play Do Nothing

(c) Household B has two neighbors of the other group, is *dissatisfied*, and will play Signal Dissatisfied

Figure 1.19: **The preferences of a household depending on the kinds of neighbors that surrounds it**. Household B will either be satisfied or dissatisfied depending on the groups of neighbors they have. B's choice to play Signal Dissatisfaction or Do Nothing therefore depends on the composition of their immediate neighborhood.

systematic denial of equal dignity to all citizens. The correlates of segregation typically include systematic deprivation of adequate schooling, health facilities, personal security, and other necessities of life to a politically subordinate demographic group. Racial segregation in New York City is illustrated in Figure 1.18.

Segregation often results from deliberate policies of discrimination by governments, banks, and home owners. Examples are the apartheid system of enforced racial separation in South Africa that persisted until 1994 and legally mandated housing segregation in the U.S. – the so-called racial covenants that were finally outlawed in 1968. Deliberate attempts to sustain segregated communities continue to the present; in the U.S. for example in state laws – "single family zoning"– that effectively make it impossible to build inexpensive housing in high income neighborhoods.

But segregation can also result from the uncoordinated decisions of people who would actually prefer to live in integrated communities. This counter-intuitive result illustrates the use of the Nash equilibrium concept.

It underlines the lesson already learned from the interaction among the Palanpur farmers. The lesson is that there may be more than one Nash equilibrium – one Pareto-superior to the other – and a society can find it difficult to escape the inferior equilibrium. The example of segregation is also a reminder – like the case of the over-fishing Nash equilibrium – that the fact that an outcome is a Nash equilibrium does *not* mean that it is something that the players would choose, if they could coordinate and decide jointly on the outcome.

The set-up of the model

Here is a model. There are two groups of people, Greens and Blues, and they live in homes arrayed around a circle representing a neighborhood. The homes are identical except that they may differ in the group-identities of the immediate neighbors. The neighborhood is the circle as a whole. A household's immediate neighborhood is made up of the two households on either side of it.

FACT CHECK Two decades into the twenty-first century in Seattle, Washington (USA), unenforceable but still on the books "racially restrictive covenants" covering more than 20,000 homes prohibit sale or rental to particular groups. One stipulated that, "No person or persons of Asiatic, African or Negro blood, lineage, or extraction shall be permitted to occupy a portion of said property."[12]

If a citizen would like to live at some other location around the circle, they can switch with some other person currently occupying that position, as long as the other person is willing. The homes just change occupants with no money changing hands. We would like to know what the neighborhood will look like when all the switching that people can do has been carried out, so that the neighborhood's composition *stops changing*.

When *no citizen is able to benefit* by switching a home the distribution of homes among the groups around the circle is a Nash equilibrium.

Greens and Blues have identical preferences and they care only about the group identity of their two immediate neighbors. All people in the neighborhood would prefer to have one neighbor of each group, as is shown in Figure 1.19. But they are "satisfied" as long as they either have an immediate neighbor of each group or if both are of their own group. People are "dissatisfied" if both immediate neighbors are of the *other* group An ideal neighborhood, then is shown above in Figure 1.20 c: Each person has one neighbor of each group.

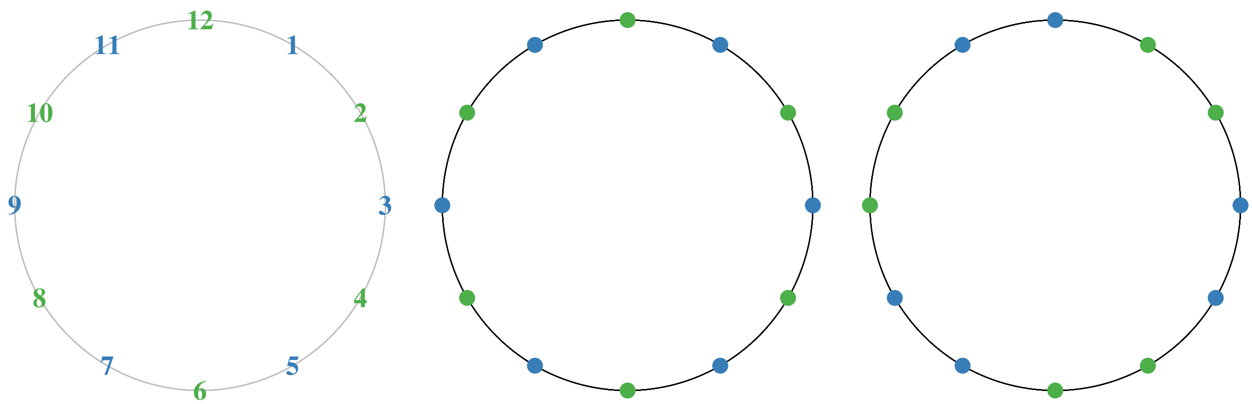
FACT CHECK Between-nation group inequalities are much greater than within nations inequalities as you can see from this interactive data set. https://jackblun.github.io/Globalinc/ which is a link to CORE global income distribution data. Another example of group inequality: the most educated member of upper caste Brahmin households in India has on average twice the years of schooling as the most educated in lower caste households. And in all castes women have much less schooling then men. Membership in all of these groups – castes, genders, nations – are accidents of birth.[13]

People have two strategies: "Do Nothing" or "Signal Dissatisfaction." Signalling dissatisfaction means being willing to switch positions with another person – anywhere in the neighborhood – who has also "signalled dissatisfaction." People are willing to switch only if they prefer the new location to their old location. For this reason people of either group will *never* switch with a person of the *same* group. This is because, for example, if a Green is dissatisfied with her current location and would like to move, all other Greens would be equally dissatisfied were they to take her position, so no other Green would agree to a switch. So all of the switches will be with different groups: a Green will switch with a Blue, but a Blue will never switch with a Blue and a Green will never switch with a Green. This means that switches will change three things:

- *both switchers' own immediate neighbors*: for the two who switched, they now have immediate neighbors that differ in their group identity from before the switch (this is the reason for the switch)
- *the switcher's new immediate neighborhood*: those on either side now experience having a neighbor of a new group identity given the switcher's arrival and the previously dissatisfied person's departure
- *the switcher's old immediate neighborhood*: those who were previously on either side of the switcher have a neighbor of a new group given the arrival of the person with whom the previously dissatisfied person switched.

A segregated Nash equilibrium

We begin with 6 Greens and 6 Blues occupying alternating positions in the 12 "houses" at the locations on the circle that are numbered as if from time on a clock (so, 12 is the top). The twelve homes on the circle are "the neighborhood." We call the assignment of different groups to the the homes around the circle in Figures 1.20 b and c: an *allocation*. An allocation in this game



(a) The circle as a clock

(b) The baseline

(c) The citizens' ideal integrated neighborhood

Figure 1.20: **The neighborhood and the citizen's ideal integrated outcome** Panel a is the "geography" of the neighborhood, showing that, for example, the citizen at position 2 on the circle has two immediate neighbors, the people at positions 1 and 3. Panel b shows that the person at position 2 is a Blue and her two immediate neighbors are both Greens is just a starting point at which the neighborhood is as integrated as possible in the sense that the two immediate neighbors of each citizen are of the *other* type. Panel c shows the distribution of types across locations that the citizens prefer: each citizen has one immediate neighbor of each type.

is an assignment of homes to the groups at a given stage of the game. The allocation before the game starts is the *initial* allocation. The allocation after the game ends is the *final* allocation.

The game proceeds as follows.

At each step, each of the 12 people plays either Do Nothing, or Signal Dissatisfaction. Their choice of a strategy is known to all other players.

Then, one of the twelve citizens is randomly selected and given the opportunity to make a switch if she wishes and can find another person willing who has also signalled dissatisfaction and is willing to make a switch.

At step 1, for example, we might ask the Green at position 10 o'clock in Figure 1.22 a if she would like to switch. She would, because both of her neighbors are Blues (she signals "D" as in the figure). Whether she is able to make a switch depends on whether there are others who have chosen the strategy Signal Dissatisfaction. Because everyone else is also dissatisfied, she has many choices.

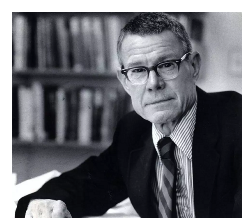
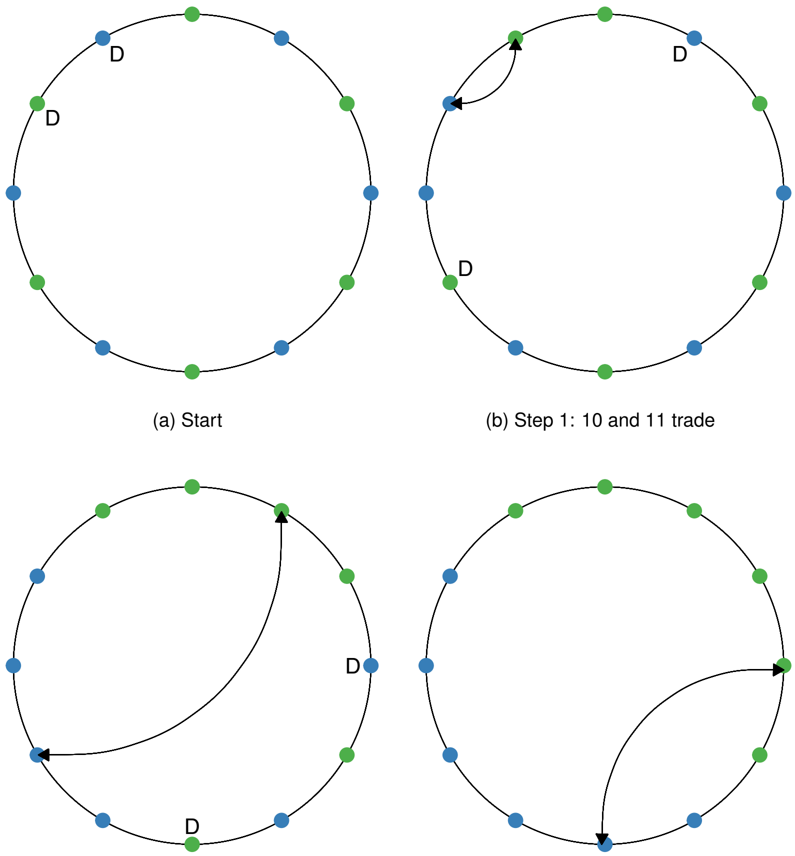


Figure 1.21: **Thomas Schelling (1921-2016)** was an American economist who won the Nobel Prize in economics in 2005 for his contributions to our understanding conflict and cooperation in what are now called "non-market social interactions" that go beyond simple exchanges of the type typically taught in economics courses. He sought to establish "an inter-disciplinary . . . theory of bargaining . . . that could be useful to people concerned with practical problems." The model of segregation here is based on his work.[14]

Suppose she switches with her friend and immediate neighbor, the Blue at position 11 (who is also signaling "D"), shown by the colors of position 10 and 11 changing from Panel a Start to Panel b Step 1. The two people are still friends and neighbors, but each now also has a same-group neighbor on the other side.

Suppose, next, that it is the Blue at position 1 who is picked to stay or switch. If he plays "Signal Dissatisfaction" (D), he could switch with his friend and at position 8. We continue this process until either no one is dissatisfied, or if someone is dissatisfied, there are no others playing the strategy Signal Dissatisfaction with whom a switch is possible. This process could continue as shown in the figure, resulting at the end of 3 steps in the completely segre-



(a) Start

(b) Step 1: 10 and 11 trade

(c) Step 2: 1 and 8 trade

(d) Step 3: 3 and 6 trade; Complete Segregation

Figure 1.22: **From integration to a segregated Nash equilibrium**. The figure shows one of many possible progressions from an integrated non-equilibrium situation to an entirely segregated Nash equilibrium. Panel a shows the starting point from the previous figure. In step 1 the Green at position 10 and the Blue at position 11 switch positions, shown by the double headed arrow, and resulting in the neighborhood shown in Panel b. The remaining panels show the successive steps to the final fully segregated Nash equilibrium. A "D" next to a household's position indicates that that household is signalling dissatisfaction and will switch places with the other household signalling dissatisfaction in the next stage.

gated neighborhood shown in Figure 1.22.

At step 4 (not shown), each of the 12 would choose the strategy Do Nothing, because 8 of them have members of their own group as neighbors only and the other four have one neighbor of each group. So no one is dissatisfied. As a result there we observe no further moves: the allocation is *stationary* (meaning unchanging). It is a Nash equilibrium.

Avoiding outcomes that nobody prefers

The conclusion is *not* that complete segregation will necessarily be the result. This is true for two reasons.

- *There is also a Nash equilibrium that is integrated rather than segregated.* In Figure 1.20 c, the allocation has each person's immediate neighborhood composed of both groups. You can confirm that, like the completely segregated allocation, this integrated allocation is also a Nash equilibrium: every citizen has their ideal immediate neighborhood, so no citizen is dissatisfied

EXAMPLE As a practical matter, they would not have to implement their ideal integrated solution. Taking account of some of the policies promoting segregated communities (that are not in the model), for example, they could vote to repeal "single-family zoning" regulations that, in the U.S., promote more homogeneous neighborhoods by limiting the kinds of housing (including lower cost housing) that can be constructed in the neighborhood.

and each are best responding with Do Nothing. This allocation could have come about by the same rules of the game that resulted in complete segregation. It was just a matter of chance whether the ideal or fully segregated neighborhoods occurred. This is an example of implementing a desirable allocation within given set of rules of the game

- *The citizens could play the game cooperatively rather than non-cooperatively*. If the citizens had realized that playing the game non-cooperatively could lead them to a complete segregation outcome that nobody wanted, they could have acted cooperatively – that is jointly agreed – to implement their ideal allocation. This is an example of implementing a desirable allocation by changing the rules of the game: agreeing to act jointly was not an available strategy in the non-cooperative variant of the game above.

The outcome in the segregation model shares three features with a game representing what would appear to be a very different situation: Planting in Palanpur.

- *A Pareto-inferior Nash equilibrium:* There is a Nash equilibrium – planting late and a segregated community – in which everyone is worse off than they could be at some other allocation.
- *A path-dependent outcome*: History matters because an outcome that is preferred by all participants is also a Nash equilibrium, so if the preferred outcome were to occur, it could persist.
- *A change in the rules of the game can avoid the inferior outcome:* By coordinating their actions – changing the interaction to a cooperative game – they could escape the Pareto inferior outcome

In these three respects the two interactions – when to plant and where to live – are *not unique* or even unusual in these three respects.

Checkpoint 1.19: Segregation as a Nash equilibrium

a. Show that the segregated neighborhood in Figure 1.22 is a Nash equilibrium.
b. Explain why the 'ideal neighborhood' in Figure 1.20 is also a Nash equilibrium.
c. Show that in Figure 1.22 if Step 3 had been different the equilibrium allocation could have been the citizens' ideal integrated allocation. Which Step 3 switch would have accomplished this result?
d. Suppose that the game was changed slightly so that a dissatisfied person knows only the "dissatisfaction status" of her immediate neighbors. Show that, starting with the alternating groups status quo (Panel a Start, in Figure 1.22) that the neighborhood would then evolve to the ideal distribution.
e. Suppose that in the fully segregated neighborhood case citizens decided to

have a binding referendum to implement the ideal neighborhood (requiring whatever moves are necessary to bring that about), but because the question of where you live is a sensitive one, they adopt the rule that unanimous approval of the referendum is required for it to be implemented. Would it be implemented?

1.16 How institutions can address coordination problems

Game theory has given us a catalogue of coordination problems: Prisoners' Dilemmas, Invisible Hand Games, Assurance Games, and Disagreement Games (there are many more!). Knowing how the structure of these games differ will help to diagnose the nature of a coordination problem and to devise policies and constitutions – changes in the rules of the game – to avoid a coordination failure.

Changing the rules of the game

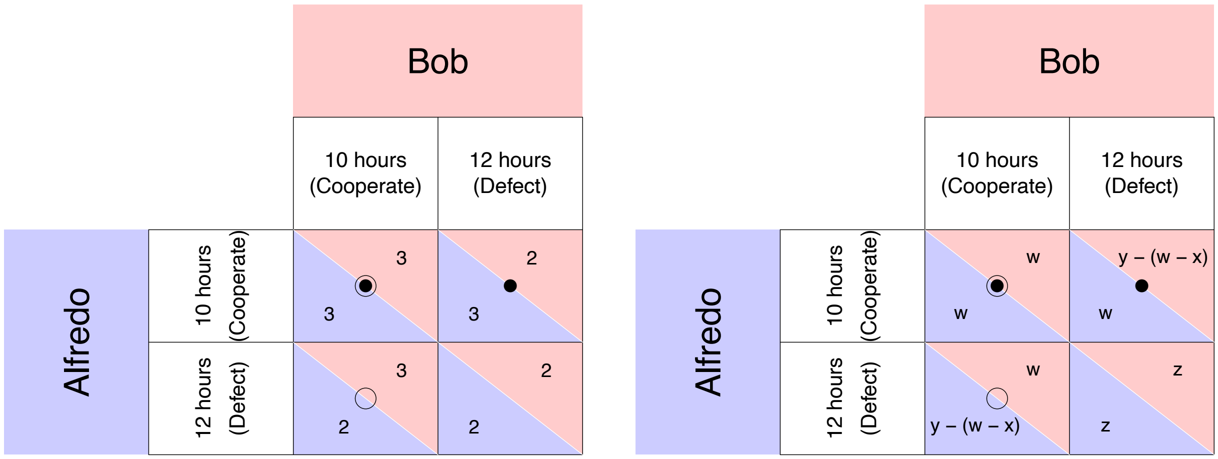
This is an example of using the concept of equilibrium to understand how to change an undesirable outcome. The idea is simple: a change in the rules of the game can eliminate an undesirable Nash equilibrium, so that it is no longer our prediction of how the game will be played. Instead it may be possible to make some preferable strategy profile a Nash equilibrium which then could be the predicted outcome of the game.

A common approach to averting coordination failures is in a Prisoners' Dilemma to devise policies or institutions that transform the payoff matrix so that the game is no longer Prisoners' Dilemma. There are two possibilities to consider:

- Change the Prisoners' Dilemma to an Assurance Game
- Change the Prisoners' Dilemma to an Invisible Hand Game

Changing the Prisoners' Dilemma game into an Assurance Game means making the mutual cooperate outcome a Nash equilibrium (it was not in the Prisoners' Dilemma) even if mutual defect remains a Nash equilibrium. In Section 5.12 we show that one way this can be accomplished is to let the the same players interact many times in what is called a repeated game. In this set-up cooperating to restrict fishing can be sustained as a Nash equilibrium because those who defect – over-fishing and exploiting the cooperation of others who fish less – can be punished in future interactions.

The second option, the one we will explore here, is more ambitious: converting the game from a Prisoners' Dilemma to an Invisible Hand Game. To see how this might work, remember that the coordination failure that results in the Prisoners' Dilemma is a consequence of the fact that in that players take actions that inflict costs on others – negative external effects – that are not part



(a) Liability rule with numbers

(b) General liability

Figure 1.23: **Fishermen's Dilemma with a liability rule**. Players can implement a desired outcome by transforming property rights using a liability rule (the harm a player does to another player is deducted from their payoff). This payoff matrix is based on Figure 1.14 modified by the liability rule. Alfredo's payoffs are listed first in the bottom-left corners and shaded blue. Alfredo's best responses are shown by the solid point. Bob's are listed second in the top-right corners and shaded red; his best responses are shown by the hollow circle.

of their thinking when they decide what to do.

To see that internalizing these external effects can address the coordination failure, we examine the implementation of a liability rule in the Fishermen's Dilemma. Tort is a branch of law dealing with damages inflicted by one person on another (or another's property). Among other things, tort law establishes the responsibility – called the *liability* – of the person inflicting the damages to compensate the harmed individual. The requirement to compensate the harmed individual internalizes the external effect.

Compensating for external effects by liability law

How would a liability system work in the Fishermen's Dilemma? Look again at Figure 1.14. We will let the payoff numbers measure something – like kilos of fish caught – that can be transferred from one fisher to another.

Suppose Alfredo and Bob decided to jointly adopt "Cooperate" (fish 10 hours) as an agreement. In their agreement, they also choose to adopt a *liability rule* requiring compensation be paid to the other party if one's actions result in lower payoffs than would have occurred had the agreement to cooperate (fish only 10 hours) been observed.

With the liability rule the following will happen:

- If both Cooperate as they have agreed, then they both get 3 as before
- If Alfredo Defects on Bob (plays Fish 12 hours), Alfredo initially gets 4
- . . . but because of the liability rule, then Alfredo must compensate Bob for the costs that his defection inflicted on Bob, who got a payoff of 1 rather than 3 (the payoff Bob would have obtained had Alfredo not violated the agreement).

- So Alfredo pays Bob 2 who ends up with 3; Alfredo ends up with 2.

We can use these changes to the payoffs to construct a transformed payoff matrix. The transformed payoff matrix for Alfredo's and Bob's payoffs is given by the entries in Figure 1.23.

Did the change in the rules of the game work? Put yourself in Alfredo's position, contemplating defecting on Bob. If he honors the agreement and fishes 10 hours, like Bob he gets 3. If he defects and fishes 12 hours he ends up with 2 after having paid Bob the compensation required by the liability rule. So Defect is no longer a best response to Bob playing Cooperate; Bob will honor the agreement. If Bob were to consider defecting on Alfredo, he would reach a similar conclusion.

M-CHECK Because $y - w + x < w$, Cooperate is now a best response to Cooperate and (Cooperate, Cooperate) is a Nash equilibrium. Cooperate is also a best response to Defect (because $w > z$), so Cooperate is the dominant strategy with the liability rule in force, and (Cooperate, Cooperate) is the dominant strategy equilibrium. Note too that $x + y < 2w$, which is important for the liability rule to work. We explain how the structure of liability rules in the Mathematical Appendix.

From Figure 1.23 using the circle and dot method you can see that (Cooperate, Cooperate) is now the only Nash equilibrium. Redefining property rights – to take account of liability for damages – can implement a Pareto-efficient outcome by inducing each player to account for how his actions affect the other player. By redefining property rights to include the liability of the damages (external effects) that one inflicts on others, we have transformed the game to an Invisible Hand game.

Checkpoint 1.20: Limited liability by the numbers

Use the model of the liability rule in Figure 1.23 to complete the following tasks.

a. Re-draw the payoff table, but substitute in the values for x, y, w and z from Figure 1.12. **Hint**: The payoffs should only be 2s and 3s.
b. Solve your new game using best response analysis (the circles and dots method) to find the Nash equilibrium of the game. What is it? Explain.
c. Does either player have a strictly dominant strategy? Is there a dominant strategy equilibrium? Explain.

1.17 Game theory and Nash equilibrium: Importance and caveats

We have started off this introduction to modern microeconomics with game theory. The reasons are that

- Many important economic relationships – in labor markets, families, credit and financial markets, between citizens and governments, among neighbors, between nations seeking to address climate change, and many more – are strategic, and require the tools of game theory.

- Focusing on people as actors often in conflict with each other, but also sharing common interests, is essential to economics as a social science, and game theory allows us to do this.

- How these interactions work out depends on the institutions that regulate them, and game theory allows us (even requires us) to be very specific about the varieties of possible rules of the game under which we now interact, and how we might change these rules for the better.

For game theory the Nash equilibrium is a key economic idea and it provides a way to answer the question: what will be the outcome if each of the actors adopts a strategy that will not lead any other actors to change what they do?

In many situations the Nash equilibrium among players who independently pursuing their individual interests provides a good prediction of what we observe in the real world. But not always. We will consider three caveats.

- *Individualism*: Overlooked opportunities for collective, not just individual best responses to the strategies of others.

- *Equilibrium selection*: The need for a method to predict outcomes when there is more than a single equilibrium

- *Dynamics*: We are interested in what happens out of equilibrium, in part because we need to know an equilibrium will come about.

The predicament of the Palanpur farmers illustrates the first caveat. If the farmers could have *all* agreed to plant early – as would be the case in a cooperative game – then they could have easily solved their planting late problem. However, following what one the farmers said (we quoted him in the introduction), we assumed in modeling their situation that an agreement among the entire set of players not possible.

But in using the Nash equilibrium concept we went to the other extreme: we assumed that an outcome would be an equilibrium (meaning undisturbed by any players changing their strategy) as long as *no single individual (acting alone)* could do better by altering strategy. But perhaps two or three jointly deciding to plant early could have done better.

A more adequate equilibrium concept would take account of collective best responses where there is reason to think that small groups might be able to decide to act together even when the entire population could not jointly coordinate.

The second caveat – about equilibrium selection – is not a criticism of the Nash equilibrium concept itself. Instead, it is a reminder that the Nash equilibrium concept by itself is *insufficient* to make predictions in cases where there are two or more Nash equilibria, as in the Planting in Palanpur Game, the Language Game, and the model of residential segregation. In these cases, knowledge of the recent past play of the game would be an important part of making predictions based on the Nash equilibrium.

We will return to the third caveat – about understanding what happens out of equilibrium – when we consider such questions as firms decisions whether or not to enter an industry and how a group of buyers and sellers could get to a competitive equilibrium (Chapters 9 and 14)

1.18 Application: Cooperation and conflict in practice

If all that is needed to address a coordination failure is to require that people pay the costs that their actions impose on others then why are coordination failures so common?

Over-exploitation of fisheries is an international problem that humans as a world community have failed to solve. Many over-exploited fisheries will not recover for a long time. But local communities and groups of fishermen have found ways to combat over-fishing, and we can learn from what they do. Many groups – from farmers to fishermen – face equivalent problems world-wide. These outcomes provide a concrete motivation to study the Prisoners' Dilemma Game and other coordination problems.

What we learn from these games is that an effective liability rule requires two things:

- The injured party or the courts have to have verifiable information (that is information sufficient to enforce the liability aspect of the property right) and
- There has to be a court or some other body willing to and capable of enforcing the contract.

When we turn from game theory to the study of real fishing communities we find that both conditions are unlikely to be met, which is why the over-exploitation of fisheries continues in many cases.

- *Limited information.* The lack of verifiable information is common in social interactions and this limits the policies that governments or private actors can design in response to the persistence of Pareto-inefficient Nash equilibria.
- *Conflicts of interest*. Governments may not have the capacity or the will to enforce the necessary policies especially in cases where doing this would impose costs on a powerful group. An example is the failure of many countries to address the problem of climate change, which is in part the result of the fossil fuel companies' opposition to putting a sufficiently high price on carbon emissions.

Fishing communities, of course, are not acting out a tragic script, as were the herders in Hardin's tale about the tragedy of the commons. They are not prisoners of the dilemma they face. Real fishermen are resourceful and seek solutions to the problem of over fishing.

- Lobstermen in the U.S. state of Maine limit how many lobster they catch using highly local restrictions on who can set traps where (the state government provides the legal framework for this).[15]
- Turkish fishermen allocate fishing spots by lottery and then rotate the spots so the distribution is fair.[16]
- The fishing community of Kayar in Senegal adopted the rule that only one trip to the fishing grounds per day is permitted (a bit like Alfredo and Bob limiting their hours of fishing) and appointed a committee to check that the rule was being observed. They also limited the number of boxes of fish that could be offloaded by a single canoe.[17]
- Shrimp fishermen in Toyama Bay, Japan have a rule that they offload their daily catch at the same time and place, so that the size of each boat's catch would no longer be asymmetric information.[18]

These rules and practices based on small local fishing communities are often disrupted by the entry of other groups whose members are not bound by the local rules. Conflicts of interest within the local community also sometimes limit the effectiveness of attempts to limit the catch. One reason is that restrictions on fishing are often supported as a way to raise the wholesale price of fish and hence the incomes of fishing families. But fish sellers – who buy the fish wholesale at the port and then sell to local consumers – would profit if they could pay less.

The rules regulating access to fishing that we see in existence are a small selection from a much larger set of rules that people have tried out at some point. What we see are the institutions that have succeeded well enough to allow the communities using them to persist and not to abandon their rules. The persistence of such rules does not require the rule to implement a Pareto-efficient outcome, it only requires that the rule be reproduced over time by people adhering to it. By this reasoning, even if the rules of the game do not implement Pareto efficient outcomes, we might expect a fishing community that has hit on a way of sustaining cooperation in the long run to do better in competition with groups that over-fish, and that successful groups may be copied by other groups.

Checkpoint 1.21: Institutions and Palanpur: Why history matters.

Supposing that the only voters involved in approving the Palanpur village council's decision to require planting early were themselves farmers, explain why they would unanimously support the measure. What would happen if after implementing the law requiring early planting one year, the next year the law was taken away?

1.19 Conclusion

The classical institutional challenge which we stated was "how to design institutions so that people could be left free to make their own decisions and at the same time avoid outcomes that were inferior for everyone?"

With the terms you have learned this can now be re-phrased "How can social interactions be structured so to avoid Pareto-inefficient Nash equilibria resulting from people's choice of their own actions?" The Fishermen's Dilemma is an example of a challenging coordination problem because an inefficient outcome is the unique Nash equilibrium. The negative external effect of overfishing in our model is intended as an analogy for coordination problems going far beyond the lake they share. The analogy includes the external effects of burning carbon-based fuels and the resulting change in global climate or the external effects associated with the spread of a pandemic.

To study a game and its likely outcomes and also how to improve these outcomes we have proceeded in three steps:

- First, use the Nash equilibrium concept to identify one or more likely outcomes of the game
- Second, use Pareto comparisons to identify outcomes that are "worse for everyone," and
- Third, devise changes to the relevant institutions – the rules of the game – or that would shift the population to a superior Nash equilibrium either preexisting (as in the case of Planting in Palanpur, or the segregation case) or novel (as with the transformed Prisoners Dilemma Game).

We have illustrated the third step by a legal remedy, the introduction of tort liabilities for damages in the Prisoners' Dilemma Game so as to internalize the external effects accounting for the coordination failure.

Making connections

Institutions and the rules of the game: To predict or explain the outcome of a social interaction, it is essential to know the "rules of the game" that determine who *knows* what and when, who *gets to do* what and when and as a result who *gets* what.

Equilibrium: Equilibrium describes an outcome that will persist until some aspect of the situation is altered as a result of externally caused changes. A Nash equilibrium is a special kind of equilibrium widely used in economics.

External Effects: People often take actions without considering the effects of these actions on others. The resulting external effects – positive and negative – pervade social interactions.