1 *Network Infrastructures*

From Coordination to Alignment

1.1 Introduction

In the opening chapter, "Lifting the Veil," we briefly outlined our perspective of alignment between the technological and institutional dimensions of network infrastructures. We consider these networks to be socio-technological systems in which the technological dimension is subject to physical laws, whereas the institutional dimension of norms, laws, and organizations frames and organizes the production and consumption of the services delivered by network infrastructures. In what follows, we shall further elaborate on the issues of coordination and alignment mentioned in the opening chapter by discussing the special features of network infrastructures. What makes network infrastructures so special that coordination and alignment are considered crucial with respect to their performance?

Analytically, we represent the two dimensions of our framework as layers, starting at the top with the generic technological architecture that is interdependent with macro-institutions. The intermediate layer is concerned with technological design specific to a time and place that is interdependent with meso-institutions. The lowest layer of specification in our alignment framework refers to the technological operations that are interdependent with micro-institutions, which coordinate transactions. This is the layer at which the concrete services should be delivered by the network infrastructure as expected by society.

In this chapter, we shall use examples from road transportation infrastructure to illustrate these two dimensions and three layers. The technological architecture consists of a network of roads (highways, urban roads, secondary roads in the countryside) connected by crossings, tunnels, bridges, and the like. Passenger cars, transportation trucks, and buses all use the roads simultaneously. A number of means of technological communication are used to guide the drivers of these vehicles – also cyclists and pedestrians – such as traffic lights, road

https://doi.org/10.1017/9781108962292.004 Published online by Cambridge University Press

signs, and information panels. The technological architecture is interdependent with the macro-institutions of norms, laws, and regulations providing general rules regarding how to behave on the roads (e.g., driving on the right side, priority of traffic coming from the right).

Technological design at the second layer of the road network example is concerned, for instance, with the creation of specific lanes on the roads that are only available for public transport buses. This is the case when government policies aim to stimulate the use of public transport by making it more efficient compared to private transport; the creation of specific lanes in crowded areas offers buses the opportunity to bypass private cars that are locked in traffic jams. In terms of contextual technological design, the lanes need to be equipped with specific technical components such as traffic signs and radar systems. This is necessary for monitoring and controlling the use of specific lanes by dedicated vehicles only. The corresponding meso-institutions (e.g., traffic control authorities) translate the general macro laws and regulations into more specific ones designed for the traffic in these designated lanes; they monitor the use of the lanes and intervene when necessary.

The third layer of our framework concerns the technological operation of the road network infrastructure. For instance, in order to increase safety, both the infrastructure and vehicles have been equipped in recent years with driver assistance technologies that help drivers avoid drifting into adjacent lanes or making unsafe lane changes. Or the vehicle might brake automatically if a car ahead stops or slows down suddenly. These technologies at the operational layer use a combination of hardware (sensors, cameras, and radar) and software to help vehicles identify certain safety risks, so the driver can be warned. The operational technology is interdependent with the micro-institutions that coordinate the transactions at the micro layer. In the institutional dimension, transactions should be organized through micro-institutions (such as contracts, vertically integrated firms, public–private partnerships), which fulfill the technological safety requirements. This will be further discussed below, and more specifically in Chapter 7 of this book.

These examples highlight some of the core characteristics of the road network infrastructure. Underlying these characteristics, societal values such as the general need for mobility, efficiency, and safety permeate the three institutional layers already identified, largely

https://doi.org/10.1017/9781108962292.004 Published online by Cambridge University Press

determining the kind of services society expects to be delivered by the network and orienting technological choices accordingly. In the technological dimension, different complementary components are involved that require tight coordination, in order to perform their functions well. Strong interdependencies exist between the components of the technological dimension and the institutional dimension at all three layers of our framework. A central hypothesis underlying this framework is that alignment between the two dimensions is crucial for network infrastructures to deliver expected services.

In Section 1.2, we discuss the specific features of network industries from this alignment perspective. The line of reasoning starts with the desired performance of the infrastructures: What are the services societies expect the socio-technological system to deliver? Network infrastructures have to provide clean water, remove waste, support safe transport, assure a reliable electricity supply, etc. In Section 1.2.1, we discuss how the expected network services are related to the values of the society in which the network infrastructures are embedded. The values of society regarding access to services such as water and electricity, as well as safety, efficiency, and privacy, differ according to time and location. Consequently, the content of the services that network infrastructures are supposed to deliver is not absolute, but relative, often having to meet different, even conflicting expectations.

In Section 1.2.2, we discuss the specific features of network infrastructures resulting from the close complementarities of their constituting components. With respect to the technological dimension, network infrastructures consist of nodes (for instance, train stations, gas production plants, traffic lights) and links (railways, pipes, roads). We shall argue that due to the technical complementarities of these nodes and links, adequate coordination is required in order to make the infrastructure deliver the expected services. Such coordination takes place through specific arrangements such as electronic devices in vehicles with driver assistance technologies. In such vehicles, sensors receive signals, for instance, about the distance from other vehicles on the road, which are transmitted to the brakes or accelerator of the vehicle by means of an electronic device. The same applies for the institutional dimension: specific nodes (laws, rules, organizations) and links (driving codes, traffic regulators) should be adequately coordinated through organizations such as the National Highway

Traffic Safety Administration (NHTSA) in the United States.[1] We explain in this chapter the importance of coordination of the components in each dimension (technology and institutions) and stress the central role of coordination arrangements.

In Section 1.3, we discuss the question of alignment in relation to the specific characteristics of network infrastructures. Alignment issues arise because of the interdependence between the technological and institutional dimensions. In this respect, they are closely related to the coordination issues already discussed; the coordination arrangements on both the technological and institutional sides must be matched. The alignment question refers to the compatibility (or incompatibility) between the characteristics of coordination arrangements. In this chapter, we shall more closely delineate the intertwined relations between coordination and alignment, and raise questions about their matching. The following chapters will provide specific insights and possible solutions to these issues.

Last, in Section 1.4, we posit this alignment framework in relation to the contributions of microeconomics, New Institutional Economics, and the approaches of socio-technological systems to the understanding of network infrastructures. We shall conclude that these contributions provide valuable insights into specific characteristics of network infrastructures, but that the core issues of the interdependence between technologies and institutions, and the related questions of coordination and alignment, are underdeveloped. Bridging this gap by developing theoretical concepts appropriate to the analysis of network infrastructure and building a coherent approach to the alignment issue is the main goal of the next chapters. In what follows, we put together the pieces of the puzzle, which are summarized in Section 1.5.

1.2 Features of Network Infrastructures

In this section, we discuss the characteristics of network infrastructures. What makes them so specific from our alignment perspective? What are the elements a theoretical framework should capture when analyzing network infrastructures? As a first distinguishing feature, network infrastructures provide services that are strongly related to societal values. Second, the provision of these services is characterized

[1] See Chapter 7 of this book for details.

by strong complementarities and the need for tight coordination among and between nodes and links.

1.2.1 The Value-Loaded Provision of Infrastructure Services

Infrastructures are the backbones of the economy, providing support to the delivery of services expected by citizens in a certain society at a certain point in time. We argue that societal values play a key role in establishing which services are considered essential in a society. For instance, the value of safety rather than speed can be prioritized in relation to the services that road infrastructures are expected to deliver. Furthermore, values and norms also impact the way society wants the technological and institutional components to be coordinated. For example, should the technology be fully coordinated through algorithms,[2] or should human agents also have a role to play? Should the institutional coordination be arranged through contracts and competitive market relations, or should a central government institution be in charge?

Expected Services: The Role of Values

In this section, we first briefly point out the existing variety of concepts used to capture the specific services that network infrastructures deliver. After having discussed some different approaches, we explain why we prefer to use the concept of "expected services" in this book.

Infrastructure services are services of specific interest, i.e., services that public authorities classify as essential for a society and its citizens. Following the liberalization of infrastructures in the European Union and elsewhere, with public transport, roads, energy, telecommunication, water, and the like, increasingly provided by private firms, a distinction has been made between two types of essential service: services of general interest and universal services (Finger and Finon, 2011). In the case of the latter, governments formulate universal service obligations, which are considered to be a form of consumer protection. For each specific service provided by the infrastructure, obligations are defined concerning its affordability (are all citizens able

[2] An algorithm is a list of steps to follow in order to solve a problem. Algorithms in computers can perform calculations, data processing, and automated reasoning. This is the case in, for instance, automated vehicles, where the technical components act on the basis of an algorithm (see Chapter 7).

https://doi.org/10.1017/9781108962292.004 Published online by Cambridge University Press

to afford the consumption of a service such as drinkable water or train transport?), the quality of the service (does the provision satisfy standards of safety, security, and punctuality?), and its accessibility (do citizens in the countryside also have access to telecommunication, education, and transportation services?). The monitoring and control of the universal service obligations are mostly put in the hands of sector-specific regulators. In the case of services of general interest, broader public policy objectives are included that go beyond citizen protection, such as general security of the energy supply, road safety, environmental protection, sustainable development, among others.

In our perspective on network infrastructures, universal services and services of general interests become essential when government formally recognizes them as a basic right for its citizens and the failure to deliver such services would result in potential risks to the public, the economy, or society. What is considered in the general interest, a basic need, and what is considered a potential risk depends on the societal values of that specific time and place. This can be illustrated using the services provided by road infrastructure.

As already mentioned, the road network infrastructure is a socio-technological system that performs important functions in society; the function of transport and mobility is considered essential from a social as well as from an economic point of view. The road system includes not only roads, bridges, and tunnels but also vehicles and their drivers. The road network infrastructure provides services to drivers of passenger vehicles, trucks, and buses, and also to cyclists and pedestrians. These services should result in safe, efficient, and convenient transportation from A to B. In that respect, the road network infrastructure is a socio-technological system, in which different components have to be coordinated to deliver the expected service.

How important the transportation function is, how and what technology can be used to fulfill the function, and what room is left for people's own responsibility depends on the societal values in place. Gerxhani and Van Breemen (2019) pointed out that societal values affect people's preferences and attitudes, but also encompass broader issues. Through processes of socialization within families, communities, and working environments, societal values become individual social values to be defined as "people's generalized beliefs regarding the desirability of conducts or end-states" (Gerxhani and van Breemen, 2019: 262). Examples of relevant values that become individual social values are "prosocial" and "proself" values, which reflect concern for others' welfare, or, respectively, for self-interest. The former motivate cooperative behavior, the latter motivate the accumulation of personal wealth. "Values have been shown to be vital in guiding evaluation of alternatives and shaping behavioral choices" (Gerxhani and van Breemen, 2019: 263).

People in society may consider it to be important that individuals who make use of the road network are personally responsible for their own safety and that of others. People in society can attribute high value to the personal freedom of road users to decide for themselves with regard to appropriate speed, acceptable maneuvers to pass other cars, etc. People in society may value differently the role of government in setting rules for the users and suppliers of infrastructure services. Likewise, people may value differently the extent to which technology should automatically control users regarding speed and maneuvers, in which case individual responsibility is absent. Values are relative and change over time.

Enlightening in this respect is a document from the National Highway Traffic Safety Administration (NHTSA) in the United States, which shows how, in the 1950s, the Big Three (Ford, Chrysler, and General Motors) competed with each other not so much on safety features but, first and foremost, with big, shiny impressive automobiles. Safety became more and more of an issue after activists such as Ralph Nader (1965) made public how unsafe the cars were, despite the fact that the technology was available to make important improvements.[3] The situation with regard to European car manufacturers was similar. Between 1950 and 2000, more attention was gradually paid to safety (and also convenience) features such as cruise control, seat belts, and anti-lock brakes. Government regulation no doubt played an important role in this. Later, new features were added such as electronic stability control, blind spot detection, lane departure

[3] Ralph Nader became a consultant to the US Department of Labor in 1964, and in 1965 he published Unsafe at Any Speed, which criticized the American auto industry in general for its unsafe products and attacked General Motors' Corvair automobile in particular. The book became a bestseller and led directly to the passage of the 1966 National Traffic and Motor Vehicle Safety Act, which gave the government the power to enact safety standards for all automobiles sold in the United States (www.britannica.com/biography/Ralph-Nader; last accessed May 2, 2019).

warning, and forward collision warning, followed by adaptive cruise control and self-parking assistance. The fully self-driving vehicle is nowadays tested not only in specific designated areas but on public roads as well (see Chapter 7 for details).

Next to safety, values surrounding environmental protection have become increasingly important with regard to the services provided by road network infrastructures. Car manufacturers have focused in recent years on the development of less polluting engines, lighter vehicles, and more efficiency. More sustainability and recyclability have become important elements of marketing strategies. These developments show that values that influence policies and the behavior of private agents change over time. These changes are initiated by technological developments, as well as changing norms and values of citizens and related government policies.[4]

Another interesting development in the values related to road network infrastructure is the valuation people attach to having a privately owned vehicle. The introduction of self-driving vehicles is expected to strongly influence the way people make use of passenger vehicles; it is expected that the transportation market will develop in the future into a market with a differentiated offer of transportation services (Arbib and Seba, 2017). People will value a safe, efficient, and convenient mode of transportation from A to B. Whether that service is provided by a privately owned vehicle or by private or state-owned firms offering transportation services, or whether the service is offered as a scheduled service or as a customized service will increasingly become a matter of consumer choice. This development illustrates how technological developments in passenger vehicles not only impact the technological components of the infrastructure but also influence societal values and the related services society expects the road infrastructure to deliver (see Chapter 7 for details). Clearly, government policies play an important role; new technologies can be stimulated or hindered and changing values can be appreciated or disapproved through different kinds of policy measures.

The lesson to be learned from this is that societal values such as safety, efficiency, privacy, etc. are incorporated into "services of general interest" and "universal services." By means of general policies,

[4] The change in energy policy as part of energy transition set out in Chapter 5 also illustrates this point very well.

the former are further specified to provide criteria ("standards of judgment"; Bush, 2009) for evaluating the services provided by the network infrastructures. Likewise, new technologies are evaluated by institutional entities such as parliaments and governmental agencies. These values are also specified, in order to provide universal service obligations for firms that provide infrastructure services to the market.

However, the translation of values into specific "standards of judgment" is not always straightforward. For instance, the evaluation of a specific new technology such as the introduction of self-driving vehicles can be problematic when conflicting values are at stake. In the case of road network infrastructures, values of efficiency (speed) can conflict with safety or environmental protection. To solve such value conflicts, a ranking system needs to be established, implying a judgment about the importance of each value involved. A system of judging and ranking values implies that values are subject to assessment, and that political institutions have to debate and decide about their ranking, and about the positive and negative role they play in society (Bush and Tool, 2003; Correljé et al., 2014). This does not mean that values change only through explicit political action. On the contrary, many of the values change "spontaneously" through anonymous interactions. Although they often happen in an incremental way, revolutionary value changes are also possible, for instance, due to a sudden crisis in the provision of an essential service. If research shows how strongly the air is polluted within a range of one kilometer from a highway, and that people suffer from relatively more serious diseases, or that the accident rate in a specific trajectory is relatively high, then the value of efficiency or speed of the road system can rather abruptly be replaced by environmental values or road safety standards.

In short, we consider that what makes services associated with specific network infrastructures perceived to be "essential" is deeply rooted in societal values. Which values in society are important and which services are essential is largely established or consolidated through political institutions. Because values differ over time and from place to place, the content of essential services is not absolute but relative. That is why in this book we shall mostly refer to "expected services" rather than "essential services," "public services," "utilities," and other terminology intending to capture the type of services that network infrastructures are expected to deliver. Values are subject to assessment, can be conflicting, and are ranked through political institutions. Consequently, the performance of network infrastructures with respect to the provision of expected services ought to be monitored, and when standards or universal obligations are not met intervention is needed.

1.2.2 Complementarities between Constituting Components

In this section, we explore in more detail the "network" character of infrastructures, the components and links making up both the technological and institutional dimensions. Central to a network is the concept of complementarity, which points to the need for tight coordination between complementary nodes and links in the networks, a crucial requirement to be fulfilled if expected services are going to be delivered and critical functions safeguarded. In other words, we substantiate the concept of "network" through the notion of complementarity, so as to better pinpoint the nature of network infrastructures in terms of the need for coordination and alignment.

Indeed, a defining feature of network infrastructures is the complementarity between nodes and links; a good or service provided by a network requires at least two nodes with a link between them (Economides, 1996). A node has no value without the other nodes and links; they are complementary to each other. From our alignment perspective, we first focus on technological complementarity, with specific attention paid to the underlying coordination required among the different technical components. We consider a tight technological coordination to be crucial for the network to perform its critical functions well, which we shall elaborate on further in Section 1.2.3. Second, we analyze the institutional dimension of complementarity and the need to coordinate the institutional components adequately. Third, we introduce the core of our perspective, which is about the alignment between the coordination arrangements along the technological and institutional dimensions of network infrastructures.

Technological Complementarities

In the case of road networks with vehicles using driver assistance technologies, or in the case of controlled entry to highways in situations of congestion, technical components need to be installed both as part of the infrastructure and inside the vehicle. Since the components of the infrastructure are in one way or another connected through a

https://doi.org/10.1017/9781108962292.004 Published online by Cambridge University Press

physical network, they cannot be operated independently from each other. The complementary components can adequately fulfill their functions only when they are tightly coordinated. The concept of technological complementarity points to a strong need to coordinate the various complementary functions performed by the technical components. For instance, sensors along highways are implemented to detect the number of vehicles, their speed, and the distance between them. That information has to be combined with other technical components, such as the signaling system at the entry points of the highway, in order to instruct vehicles in areas of congestion regarding whether they are allowed to enter or not. To fulfill these functions, the activities of the different technical components need to be well coordinated through specific arrangements such as electronic devices. This type of configuration of technological devices[5] will be discussed further in Chapter 3.

In short: the technological system consists of complementary components that fulfill specific functions and perform specific activities, which need to be tightly coordinated by technological arrangements.

Institutional Complementarities

The different components in the institutional dimension of the network infrastructure fulfill complementary functions as well. Norms provide rules of behavior, which are not explicitly formulated in laws and regulations, but which reside in people's souls and minds. Laws and regulations are formal institutions that are registered. Organizations are institutions that structure behavior within specific boundaries (for details of the institutional components, see Chapter 2). The concept of institutional complementarity points to a strong need to coordinate the different institutional components within and across the different layers of our framework. In relation to road infrastructure, the norms imposed on drivers about their responsibilities and/or the laws and rules about obeying traffic signs need to be tightly coordinated and strictly implemented; otherwise, none of the institutional components will adequately fulfill its function.

Coordination, Monitoring, and Control

Proper coordination of the complementary components along both the technological and institutional dimensions of the network is therefore

[5] We use the terms "components" and "devices" interchangeably in this book.

vital for the provision of the expected services. Coordination is the act of creating and organizing the conditions to get different people or components to work together to achieve required effects or fulfill desired goals in a network. Different activities of the components and links of a network have to be adequately tuned and adjusted, in order to make the complementary components all contribute effectively to the provision of expected services.

Adequate coordination, that is, the level of coordination needed to achieve a specific goal, requires monitoring and control. Control is about configuring the components so that they are in accord with the desired performance of the network. Monitoring of the network can be based on the performance of a network in terms of safety or efficiency. When monitoring signals that the performance does not meet specific standards, resulting for instance in congestion on the road, a technological coordination arrangement could automatically lead to adjustments, for example, traffic lights could start to function at the entry points to the road. However, usually a combination of technological and institutional devices and entities is responsible for the monitoring and control; agents responsible for assessing the situation and organizing the control make use of technological devices requiring tight coordination between both the technological and the institutional dimensions of the transportation system.

Monitoring and control come at a cost. Rigid bureaucratic control mechanisms can incentivize opportunistic behavior, whereas control mechanisms based on shared norms can facilitate less costly cooperative behavior (Gerxhani and van Breemen, 2019). Monitoring and control also become more complex and costly when a variety of technical components and many parties are involved. Furthermore, coordination becomes more complex and costly when the system has to deal with high environmental and behavioral uncertainties. It can be assumed that societies tend to be prepared to pay for high monitoring and control costs when important services are involved which affect highly ranked societal values. It is important to take these issues into consideration when explaining concrete coordination arrangements in different infrastructures in different contexts.

1.3 Network Infrastructures: Alignment at the Core

In this section, we first consider contributions to the analysis of network infrastructures in microeconomics, "New Institutional Economics," and

https://doi.org/10.1017/9781108962292.004 Published online by Cambridge University Press

"Socio-Technical Systems" (STS)[6] approaches. We will conclude that from our alignment perspective important elements are missing in these analyses. We then make suggestions about concepts and relations on which to build an analytical framework capable of dealing with the core issues of coordination and alignment.

1.3.1 Insights from Microeconomics, New Institutional Economics, and STS Approaches

Microeconomics

In the microeconomics literature on infrastructures, a distinction is made between the construction of the infrastructure, the production of its services, and the consumption of the services (Kessides, 2004).

With respect to construction, the generally large capital investments required upfront, the extended construction period, and the high level of sunk investment[7] increase risks and require a long-term horizon. This can discourage private investment and once the investments are made new entries might be dissuaded from joining. More specifically, private investors may then ask for guarantees in the form of long-term contracts, including delivery conditions and price controls, and may try to erect forms of barriers to entry. Alternatively, these risks may explain why very often such investments are made by public authorities, even in the context of so-called public–private partnerships.

With respect to the production of services, the characteristic of increasing returns to scale is prominent, leading to market concentration or even to natural monopoly. This leads to policies of "competition for the market" instead of "competition in the market," or alternatively the decision to transfer the production of services to state monopolies. In the microeconomics literature, the increasing returns to scale are connected to the cost and pricing issue. Helm (2009b: 315) points to the "wide gulf between average and marginal costs" in the delivery of infrastructure services. Once sunk investments

[6] In this section, we refer to the literature known as the "Socio-Technical Systems" approach. In the rest of the book, we refer to "socio-technological systems," since technology has a broader meaning and allows for the consideration of values, a key point in our view.

[7] Sunk investments are specific investments that, once undertaken, result in their value in alternative uses being substantially below investment costs (non-redeployability).

have been made, the marginal costs are close to zero, unless congestion is reached. Private firms try to solve the pricing issue by means of long-term contracts, in which they bind customers to buy their services for a long period and pay a price that covers average costs. An alternative solution is to rely on public investments. For instance, for a long time in Europe many countries opted for public ownership and socialized costs through taxes or a combination of taxes and user fees.

With respect to the consumption of infrastructure services, both public good characteristics and externalities play a role. In the case of pure public goods, the consumption of services is non-rivalrous and non-excludable, causing free rider problems. Infrastructures such as road networks often provide so-called impure public goods. In the case of transportation networks, when roads have sufficient capacity in relation to demand, all users of the road consume the service without rivalry. However, beyond a specific number of users the road capacity reaches its limits and congestion occurs. Consumption becomes rivalrous and the marginal cost positive. Infrastructures are typical examples of such impure public goods.[8]

Related to the consumption of infrastructure services are the so-called network effects (Economides, 1996). A network effect exists when the consumer value increases with the number of users. This is clearly the case in telephony and social media. The more users that are connected to these communication networks, and the more services individual users can derive from them, the higher their consumer value. Economides (1996) elaborates on the consequences of complementarity and network effects in terms of strategies of firms, types of markets, and the role of competition authorities.

Other issues raised in microeconomics are mainly related to questions of price and affordability. Because the services of network infrastructures are mostly considered "essential" for the users and therefore should be available to everyone at affordable prices, microeconomics suggests applying price discrimination and price regulation.

With respect to externalities, both positive ones (for instance, the effects of the road infrastructure on economic growth) and negative ones (for instance, the effects of the usage of the roads on air pollution)

[8] See also club goods, which are considered to be impure public goods because of the limited access allocated to an exclusive small membership aimed at avoiding congestion.

are present and are not internalized, or very partially so, in the prices of goods and services. According to the rich literature on this subject, either private contracting (Coasean solution) or public intervention through taxes or subsidies (Pigovian solution) are possible. Both solutions are not always completely effective, and if essential services and societal values are endangered, governments can decide to completely eliminate the externality by forbidding the activities. For example, restrictions may be placed on traffic during peak hours generating high pollution. However, all such solutions require institutions to establish and implement the underlying norms and rules, an issue rarely discussed in the standard microeconomic approach to infrastructures besides reference to government monitoring (through taxes, regulation, etc.).

Insights from New Institutional Economics

Institutional economics focuses on issues of governance: the efficiency of different modes of organization is assessed in a comparative way. Transaction cost economics, positive agency theory, and the economics of property rights are prominent theories in New Institutional Economics (Williamson, 2000). With respect to the domain of infrastructures, Williamson (1999) paid explicit attention to the transaction cost minimizing function of "public ordering" through regulation and public bureaucracies. Williamson explained that in cases of natural monopolies and specific transactions which implicate the security of the state (such as in foreign affairs), contracting out poses great difficulties that translate into high transaction costs. Regulation or coordination of the transaction through government entities then become efficient solutions according to a transaction cost minimizing perspective. Principal-agent theory addressed the incentive issues required to align the interests of parties involved in a transaction, whereas the economics of property rights dealt with questions of ownership in infrastructures (Groenewegen, van Spithoven, and van der Berg, 2010).

The concept of transaction costs is further developed and applied in the world of public contracting, particularly when it comes to building, operating, and managing network infrastructures. Savedoff and Spiller (1999: 6) discuss the characteristics of infrastructures, emphasizing how "prevalence of sunk costs, economies of density/or scale, and massive consumption, lead to the politicization of utility pricing." As a consequence, infrastructures are characterized by their tight embeddedness in political choices about governance and pricing. The

https://doi.org/10.1017/9781108962292.004 Published online by Cambridge University Press

economies of density refer to the fact that in a given distribution network, an increasing number of household connections to the infrastructure reduces the network's average costs. This drives the market structure toward an ever-decreasing number of suppliers. Savedoff and Spiller argue that the characteristics of sunk costs, density, and massive consumption allow governments to behave opportunistically toward the investing company: "For example, after the investment is sunk, the government may try to lower prices, disallow costs, restrict the operating company's pricing flexibility, require the company to undertake special investments, control purchasing or employment patterns, or try to restrict the movement or composition of capital" (Savedoff and Spiller, 1999: 7).

Next to governmental opportunism, Spiller also points to third party opportunism, which is undertaken by parties that are not directly involved in the contract between, for instance, a governmental agent and a private operator. Third parties such as employers' associations, labor unions, environmentalists, representatives of regional interests, and the like "may have incentives to challenge the 'probity' of the public agent involved in the transaction, even if the transaction is being undertaken in an honest way" (Spiller, 2009). In contrast to the Chicago School of regulation (Stigler, 1971), in which rent seeking is at the core, Spiller and others emphasize the institutional aspects that impact on the nature of regulatory institutions. In doing so, they provide deeper insight into the causes of regulatory risk (governments opportunistically changing the regulation rules) and regulatory capture (private parties decisively influencing a regulation that is not beneficial to them), making an important step in the direction of a more inclusive approach to the institutional dimension involved in the functioning of network infrastructures.[9]

Insights from Socio-Technical Systems Approaches

In microeconomics and New Institutional Economics, technology is most of the time ignored[10] or plays a very limited role. Regulatory issues

[9] Spiller pays much attention to different types of opportunism related to issues of "public contracting" in industries such as infrastructures; in that respect, his approach also differs from the incentive approach to regulation, as developed among others by Laffont and Tirole (1993).

[10] There are important exceptions, such as Economides (1996) from a micro-neoclassical point of view, or Shelanski (2007) from a new institutional point of view. See also Chapter 2.

are typically discussed without sufficiently taking into account the specific technological characteristics of different network infrastructures.

In the literature on Socio-Technical Systems (STS), the technical dimension is explicitly taken on board. However, we shall argue that although technology and institutions are clearly central in STS approaches, they do not adequately analyze their interdependence as outlined in our alignment perspective.

Socio-Technical Systems are built up from technical artefacts and institutional artefacts. Technical artefacts are physical artefacts with a technical function given to the artefact with human intentionality (Bauer and Herder, 2009; Kroes et al., 2006). This dual character of technical artefacts, with a physical structure and a technical function, is the first building block of a socio-technological system. In infrastructures, different technical artefacts fulfill different functions, which ought to be well coordinated, as explained in Section 1.2.2.

In STS, the technical characteristics are complemented by social components: values, norms, and laws that structure the behavior of the agents. The key structural features of STS (Bauer and Herder, 2009; Künneke, 2008) are that the technical and social parts are intertwined (for instance, when values about safety change, the technical components of the road need to be adapted, such as the introduction of electronic speed control); that each consists of multiple layers (norms about safety or privacy are formulated at the highest layer of abstraction and need to be translated into concrete laws, regulations, protocols, and standards for the agents at the micro layer); and that each layer refers to a different time scale (changes in the layer of norms take much longer than changes at the layer of protocols). These features and also the layers identified in Williamson (2000: 597) connect well to our own alignment framework, as we shall detail in the following chapters.

Much of the literature on STS relates to the so-called blueprint paradigm: the questions about technical artefacts and their coordination are formulated in terms of "optimization under constraints." An objective is formulated, the environmental constraints (technical, physical, and institutional) are identified, and instruments to deal with them are chosen. This connects well to the world of systems engineering further discussed in Chapter 3 of this book. In that respect, the choice of referring to "technical" rather than to "technological," as we do in our perspective, is significant. Socio-Technical Systems are concerned with the design of the technical elements of a system, with the aim that all the elements in combination and interaction fulfill a specific purpose, a specific function. The design and engineering approach also holds for the social part of the system. Most involved disciplines (engineering, economics, law, public administration, and management science) assume, often tacitly, that effective solutions to the design issues can be found and implemented. In this approach, the network infrastructure is typically perceived as a blueprint: the design of the technical as well as the social part is a matter of engineering, submitted to scientifically established physical and social laws and regularities. In this blueprint paradigm, STS such as electricity, transport, and water systems can and should be comprehensively designed and controlled.

In the literature on STS, the blueprint design paradigm is contrasted with the so-called process paradigm. According to this paradigm, the focus in complex, adaptive systems should be on the evolutionary process, in which the technical and institutional dimensions co-evolve. Murmann (2003), Saviotti (2005), Bijker, Hughes, and Pinch (1987), and Vazquez, Hallack, and Perez (2018) explain, among others, how all artefacts in STS are constituted in an interactive process. Such a process cannot be designed, let alone be controlled and engineered toward a specific ex-ante formulated purpose. In the literature of the STS process paradigm, detailed descriptions of specific cases are analyzed to demonstrate how technology and institutions co-evolve in an incremental, but sometimes also revolutionary way. Based on detailed case studies, path dependencies, sub-optimal solutions, and largely unpredictable developments are claimed to be core characteristics of STS.

In a noteworthy publication, Vazquez, Hallack, and Perez (2018) provide a detailed simulation of the co-evolution of technology and institutions in the electricity sector. They are concerned with the evolution of rules in the regulatory framework: how does such a framework emerge from the complex interaction between rule makers and industry? Based on the work of Dosi (1982), Williamson (1998), Künneke (2008), and Ostrom (2009), among others, these authors develop a framework in which technology and institutions interact at different levels. In that sense, they connect well to the process approach of STS and take the process paradigm of STS discussed in this section a step further into formalizing and simulating the dynamics.

What Is Missing?

From the discussion in this section, we conclude that microeconomics, institutional economics,[11] and the STS approach do provide useful insights into aspects of the nature of network infrastructures. Indeed, insights into the monopolistic tendencies, the regulatory issues, the question about efficient governance, and the design and evolutionary aspects certainly prove to be useful for the understanding and designing of network infrastructures. However, none of them provides a satisfying framework, with related concepts, to deal with the core of our alignment issue: the interdependence between technology and institutions. We are primarily interested in dealing with this issue by adopting a comparative static perspective. We do not intend to analyze the interaction between technology and institutions out of which coordination arrangements and modalities of alignment would emerge. We rather focus on the interdependence between technology and institutions at a certain point in time, and aim to deepen our understanding of the (mis)alignment between their respective coordination arrangements. We hypothesize that such an understanding will support a more effective design of STS, helping to formulate better policy recommendations, and ultimately opening ways to better capture the dynamics of STS (see Chapter 8 of this book).

1.3.2 About Coordination and Alignment

In this section, we further explore the core of our alignment framework: the coordination of the technological components within each layer of the technological dimension on the one hand, the coordination among the institutional components within each layer of the institutional dimension on the other hand, and the alignment between these two sets of coordination arrangements within each layer of our framework. In order to better understand these issues of coordination and alignment, we first discuss in more detail the four critical functions we identified in the opening chapter of this book. The four critical functions of system control, capacity management, interconnection, and

[11] In the other school of institutional economics, often called American or original institutionalism with authors like Veblen, Ayres, Galbraith, and Myrdal, technology plays a more prominent role, but the way it is analyzed is more at the general level and not adequate for the type of alignment questions we address in this book. See also Chapter 2 of this book.

interoperability have to be fulfilled; otherwise, the expected services cannot be provided by the infrastructure. We hypothesize that adequate coordination and alignment are needed, in order to fulfill these four critical functions and to enable network infrastructures to deliver expected services.

Once we have clarified what these four critical functions mean and their main characteristics, we dig deeper into the issue of coordination and alignment, which is crucial for the fulfillment of these functions. We explore the arrangements that coordinate the technological components on the one hand and the institutional components on the other hand. Further, we raise questions about these two types of arrangements and the alignment between them: how to analyze their compatibility, how to define alignment, how to create it, and when a misalignment has occurred how to restore it. The answers to these questions are provided in the coming chapters.

Four Critical Functions

As should be clear by now, we are primarily interested in those technological and institutional coordination requirements of network infrastructures that are needed to support the complementary activities between the various nodes and links which allow the provision of expected services to their users. We specify these requirements for the four critical functions briefly mentioned in the opening chapter: system control, capacity management, interconnection, and interoperability (Finger, Groenewegen, and Künneke, 2005; Künneke, Groenewegen, and Ménard, 2010).

System control: network infrastructures constitute systems that deliver a specific good or service of an expected quality. "*System control pertains to the question of how the overall system (e.g., the flow between the various nodes and links) is being monitored and controlled and how the quality of service is safeguarded*" (Finger, Groenewegen, and Künneke, 2005: 241). In the case of the road network infrastructure, providing a service of a specific quality concerns safe, efficient, and convenient transportation. Such a service is delivered by an overall system of roads, traffic lights, vehicles, laws, regulations, drivers, cyclists, pedestrians, etc. All these components contribute to the essential service, of which the quality depends on the quality of the components (such as the quality of traffic authorities, the capability of the drivers, the accurateness of specific sensors) and

https://doi.org/10.1017/9781108962292.004 Published online by Cambridge University Press

on the coordination among them. In order to perform adequately, the technological dimension of the network infrastructure should be operated according to certain technical requirements. In the case of road infrastructure, the quality of the pavement, the existence of shoulders along the road, the presence of notice boards informing users about accidents, etc., are examples of technical components that contribute to the provision of safe services. With respect to the institutional dimension, the essential services benefit from clear, consistent rules, and competent institutional entities that test vehicles and drivers' competences.

In other network infrastructures such as telecommunication, electricity, railroad, and bus transport, the liberalization of the last decades of the twentieth century has resulted in the entry of competing private firms into the industry, whereas the road network mostly remains the property of public authorities or regulated enterprises. Under these conditions, agents belonging to these different and partly competing entities have an incentive to pursue their own strategic objectives, which can conflict with the need to maintain the quality of service of the entire infrastructure (Kessides, 2004). With growing fragmentation of the technical system due to unbundling, outsourcing, and the like, there is a growing need from the perspective of system control to monitor and control the activities of all agents involved. In Part II of this book, we shall discuss several examples of how the function of system control can be best safeguarded.

Capacity management: "*networks are scarce resources because the capacity of nodes and links is limited. Capacity management deals with the allocation of this scarce network capacity to certain users or appliances*" (Finger, Groenewegen, and Künneke, 2005: 241). For instance, it has to be determined which agents and vehicles will be allowed to use the road infrastructure and under what conditions. For passenger vehicles and trucks, technical standards are developed on the basis of which car manufacturer can get a license for the vehicle to operate on public roads. In order to reduce congestion, the limited capacity can be reallocated to a limited number of vehicles during rush hours. So the actual access to the roads is facilitated by specific regulations regarding the institutional dimension, which can specify how many vehicles are allowed to enter the network at a certain point in time. Another way to manage the limited capacity and resulting congestion is road pricing: high tariffs are, for instance, charged for using roads in city centers during rush hours (see, for example, the London Congestion Charge[12]).

Interconnection refers to the coordination of activities and services between different segments that perform similar or complementary tasks in an infrastructure network. Segments of networks need to be connected with each other, in order to guarantee the technical functioning and the delivery of expected services. For example, the provision of expected services by the road system may depend on the existence and quality of interconnection between the local, regional, national, and even international road networks. This also holds for the railroad and airline network infrastructures, and the interconnection between the different transportation networks. Other typical examples include the interconnection between different parts of communication networks, such as long distance lines and the last mile to the customer. Interconnection sometimes occurs beyond the boundaries of specific infrastructures. This happened in the transportation sector with the introduction of standardized containers, which allows for very fast and efficient intermodal traffic between road, shipping, or air traffic infrastructures.[13] The energy sector is another example, as electric power can, for instance, be used for transport mobility (electric vehicles), which requires establishing an interconnection between the transportation and electric power network infrastructures.

Interoperability refers to the requirements that components of infrastructure networks must satisfy, in order to support the complementarity between different nodes and links that structure the network. "*Interoperability is realized if mutual interactions between network components are enabled in order to facilitate systems' complementarity*" (Finger, Groenewegen, and Künneke, 2005: 240). For example, in the railroad sector, the specification of the tracks needs to be compatible with the requirements for locomotives and cars. "*In the aviation sector, airlines rely on specific navigation systems that guide planes to their destination without accidents*"(Finger, Groenewegen, and Künneke, 2005: 240). In the road network infrastructure, components such as

[12] Details of the London Congestion Charge can be found at https://tfl.gov.uk/modes/driving/congestion-charge; last accessed November 4, 2019.

[13] According to Levinson (2008), in his book on the emergence of containers and how it revolutionized the world economy, two dimensions of interconnection can be identified: "intra," which is within an infrastructure sector, and "extra," referring to different infrastructures.

the width and height of tunnels, or the carrying capacity of bridges, should be interoperable with the dimensions of the transport trucks. With the growth of automation of vehicles, interoperability between the technical components inside the vehicles and on the infrastructure becomes of crucial importance (see Chapter 7). Interoperability ensures that the elements of the network are compatible; technical norms, standards, and regulatory conditions of access are ways to secure interoperability. In this sense, interoperability is also of strategic importance. It determines the conditions of use as well as the rules for entry into and exit from a specific facility.

About Coordination

As outlined above, the complementary components that form network infrastructures require adequate coordination. In the technological dimension, coordination arrangements are often of a technological nature, but it is also possible that interventions are brought about by human beings who are assisted by technical devices. In the case of road networks, the monitoring of the congestion and the control of traffic lights that allow additional vehicles to enter the highway can be fully automated: monitored cameras provide computers with information which instruct the traffic lights. However, it is also possible that human beings, who at a distance monitor the traffic situation and can manually intervene when appropriate, assure coordination. Finally, the decision to fully automate traffic surveillance and control is largely a matter of cost, although values may also play an important role in this respect. As will be explained in Chapter 7, fully automated or self-driving vehicles can be equipped with technical components and algorithms that allow the vehicle to decide in case of threatening situations; for instance, to avoid a collision with another car and to leave the road and enter the cyclist lane. However, it may very well be that the owners and/or passengers of such vehicles prefer not to have fully automated technical components on board, but to have the opportunity to manually intervene. In other words, selecting the algorithm of the software can raise serious questions relating to societal values: who is responsible for the programming of automated vehicles, and should human beings not always have the technical possibility to intervene when the vehicle seems to take a decision that is unacceptable from a values point of view? These issues are discussed in Chapter 7. For the time being, we only wish to stress that different types of coordination can be chosen: completely technical and fully automated, or allowing human beings to intervene when they consider it appropriate, or a mix of these solutions, as when intervention by human beings is possible only under specific conditions.

In the coming chapters, we shall discuss in more detail the coordination issues within both technological and institutional dimensions, and along the macro, meso, and micro layers. We shall demonstrate that coordination can be of quite different natures. For example, coordination can be closed in the sense that one unique coordinating structure is allowed and no alternatives are accepted. Or the arrangement chosen might be centralized with no autonomy left for decentralized solutions. The opposite is also possible: from the 1980s onwards, we have seen a movement toward more open and decentralized arrangements in network infrastructures. At this point, what we consider to be of crucial importance is the acknowledgment that coordination in network infrastructures can rely on different types of solutions, along the technological dimension as well as the institutional dimension. However, of crucial importance for the provision of expected services is that the different modalities of coordination within the technological dimension be aligned with those within the institutional dimension.

About Alignment of Technological and Institutional Coordination

Alignment and misalignment refer to the compatibility (and incompatibility, respectively) between the characteristics of the technological and institutional coordination needed to safeguard the critical functions along each layer of our framework. Our concept of alignment finds its inspiration in the work of Oliver Williamson. He puts the concept of alignment at the center of transaction cost economics (TCE): alignment is about "matching," about a "fit" of the types of transaction and the structures of governance.

> Transactions, which differ in their attributes, are aligned with governance structures, which differ in their cost and competence, so as to effect a (mainly) transaction cost economizing result. (Williamson, 1998: 37)

For Williamson, the question of alignment matters because it has a direct effect on performance: in the world of TCE, misalignment results in relatively high transaction costs, which can be devastating for organizations operating in a competitive environment. It also means that Williamson restricts his analysis of the question of alignment to the third layer of our framework.

Our understanding of alignment is more general and concerns the compatibility of coordination along the three layers of our framework: between the technological architecture and the macro-institutions, between the technological design and the meso-institutions, and between the operational technology and the micro-institutions. However, consistent with the approach developed by Williamson, we also consider that alignment is central to performance. We actually hypothesize that whatever the layer under consideration, if there is misalignment between technology and institutions, then all or some of the four critical functions will not be fulfilled, causing the network infrastructure under consideration to underperform with respect to expected services.

Whether technological and institutional layers are "aligned" (or misaligned) can therefore be assessed by taking the expected performance of infrastructures into consideration. Signals of a potential misalignment can come from the users, who can show their dissatisfaction through different modalities, from media to public protest. Such signals of dissatisfaction can indicate a lack of coordination within either the technological or institutional dimension, or a misalignment between the coordination solutions along one or more layers. Another way to identify possible misalignments is to compare the performance of the network with the societal values, for example safety, privacy, reliability, or environmental sustainability. For instance, the National Highway Traffic Safety Administration (NHTSA) in the United States could report that the number of accidents on the highway has unacceptably increased and could propose a change in the law regulating the maximum speed, or they could propose a requirement for car owners to have their car checked every year. The responsible meso-institution monitors, investigates, and proposes measures for policy makers or the judiciary to adapt the rules. In line with the importance given in our approach to the role of values, the judgment about alignment is closely related to the societal values that shape expectations about the performance of infrastructures. Because values are time and place specific, there is no absolute degree of alignment or misalignment but only one in relation to these values, which implies that the operationalization of the concept of alignment is always contextual. In other words, what is considered to be (mis) alignment differs over time and from place to place.

1.4 Conclusions

In this chapter, we specified the nature of network infrastructures from our alignment perspective. We first paid attention to the expected services that network infrastructures intend to provide to society: they are the backbones of the economy and they deliver services essential to its citizens. We showed how the infrastructures and the services they are expected to deliver are embedded in societal values, which differ depending on time and place.

We then discussed the two dimensions of network infrastructures, the technological and institutional dimensions, and analyzed the characteristic of complementarity that underlies their components. Nodes and links have no function on their own, but only in complementarity with other nodes and links. Complementarities require tight coordination among the three layers of the technology as well as among the three layers of the institutions. Furthermore, we discussed in this chapter the core of our argument: that the modalities providing technological coordination on the one hand and institutional coordination on the other hand should be well aligned; otherwise, the fulfillment of the critical functions is endangered.

In the opening chapter of this book, we introduced the representation of network infrastructures as socio-technological systems, which are complex layered systems with interdependencies between technologies and institutions within each layer. The outline of our analytical framework was summarized in Figure I.1. In this chapter, we further elaborated on the dimensions, layers, and interdependencies involved in the different aspects of this framework. In doing so, we overviewed the contributions of microeconomics, New Institutional Economics, and the Socio-Technical System approaches to the characterization of network infrastructures. We concluded that they provide useful insights and concepts, but that we need to go further to better understand how network infrastructures operate, and under which conditions they can achieve the expected level of performance. To meet this challenge, we are making the choice to focus on the interdependencies between the technological and the institutional dimensions; on the critical functions as requirements for the system to provide the expected services; and on the necessity to align the coordination arrangements in both dimensions in order to fulfill these critical functions, without which the expected services cannot be delivered.

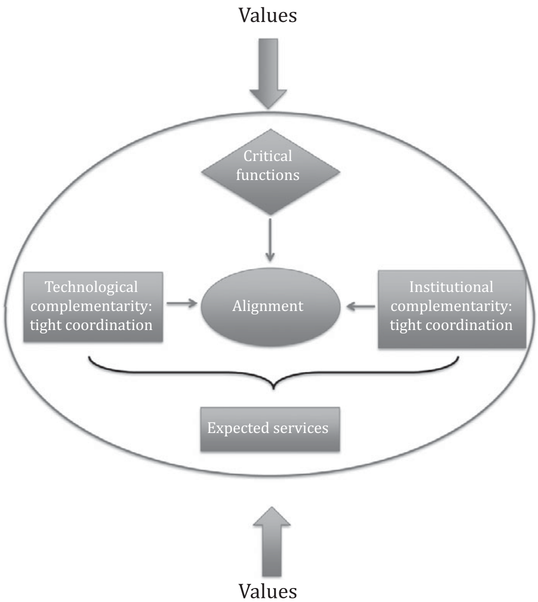

Figure 1.1 Features of network infrastructures: the alignment perspective

Figure 1.1 summarizes the key concepts and relationships of our alignment perspective. First, with respect to the performance of network infrastructure we focus on expected services. Second, with respect to each of the two dimensions, respectively technology and institutions, we focus on the complementarities of the components involved and on the need for their tight coordination. Indeed, the coordination of components within each dimension is crucial for the fulfillment of the critical functions. Third, we stress the importance of alignment between the coordination devices implemented within the two dimensions for the fulfillment of these critical functions. Last, we emphasize that each and all networks are entirely permeated by values in all their components and dimensions: the choice of services to be considered as essential, the choice of the technological components, the choice of the institutional components, the choice of the coordination arrangements, and the choice of the modalities of alignment are all embedded in societal values.

The main lesson to be drawn from this chapter is that infrastructures are very complex socio-technological systems, and that their analysis requires more subtle concepts than those provided by the existing paradigms. To go further in this direction, we suggested that the interdependencies between the technology and institutions can be best analyzed if three analytical layers are differentiated. We also suggested that attention should be focused on the coordination needed within each of the two dimensions of institutions and technology, and that the compatibility of the solutions thus implemented along the three layers we have identified should be explored.

In the following chapters of Part I, we shall further elaborate on the two dimensions of institutions and technology, on why they would be better understood if distinct layers are identified, and on how our concept of alignment can help to figure out the modalities through which network infrastructures become operational. Part II will promote the relevance of this alignment framework through the exploration of relevant characteristics of different network infrastructures, showing the explanatory potential of our framework without ignoring some of its limitations.