Design in socio-technical system development: 
three angles in a common framework

Pieter W.G. Bots

Faculty of Technology,
Policy and Management,
Delft University of Technology,
2600 GA Delft, The Netherlands

Cemagref, Montpellier Regional Centre,
BP 5095, 34033 Montpellier Cedex 1, France
E-mail: p.w.g.bots@tudelft.nl

Abstract: Large-scale socio-technical systems, such as infrastructures for transport, energy and telecommunication, are not designed and then constructed according to plan. Rather, they develop over a long period of time as a result of countless changes. Nonetheless, most of these changes have been produced by design. Three types of designs – system design, decision process design, and institutional design – are identified as pertinent to large-scale socio-technical systems, and characterised by applying a generic conceptual framework to a fictitious case. This characterisation provides some insights into the variety of design problems that must be addressed in the context of socio-technical system development.

Keywords: artefact; car sharing; CO₂ emissions; decision process; design process; flow; institution; regime; structure.


Biographical notes: Pieter W.G. Bots is an Associate Professor of Policy Analysis at Delft University of Technology, The Netherlands, but presently (11/2005–11/2008) on secondment as a Visiting Researcher at Cemagref in Montpellier, France. Trained as a Computer Scientist, he has taken to analytical techniques to support problem solving in multi-actor contexts. His present interests include the analysis, design and evaluation of policy- and decision-making processes, actor-oriented methods and tools for modelling and decision support, and the application of conceptual modelling techniques to represent and compare theories and models in the social and behavioural sciences.

1 Introduction

Emery and Trist (1960) introduced the term ‘socio-technical system’ to characterise the dynamic double bind between the social and the technical in the context of industrial organisation and development. It is this double bind – technology co-determining human
activities and human activities co-determining technology – that seems to confound system engineers who attempt to get a clear set of requirements and then create a technical artefact that does the job. The more intricately the artefact is linked with social processes, the more difficult it is to design.

The problem obviously increases with scale. Most of the literature on the design of socio-technical systems – see Mate and Silva (2005) for a recent overview – focuses on the scale of business processes, with a strong bias to information systems. These business processes and information systems are but components of the large-scale socio-technical systems that provide transport, energy or telecommunication services for a region, a country or more. Such large-scale socio-technical systems can not be designed – they develop – and yet they harbour a variety of design activities.

My objective in this essay is not to propagate one particular truth about design, but to clarify three types of designs that can be observed in the context of large-scale socio-technical system development. To achieve this objective, I have chosen a rather abstract vocabulary, which I define in Section 2. The characterisation of the three types of designs follows in Sections 3–5. I have opted for stylised fictitious design cases because these provide the most didactical illustrations. The cases have in common that they all address the problem of reducing the CO₂ emission by traffic. I start with a sketch of the types of systems that engineers from various technical disciplines might design to solve this problem. I then use the same approach to describe alternative designs of a decision process that, when executed, should lead to a policy to reduce CO₂ emission by traffic. Next, I outline different designs of an institution that would resolve the CO₂ issue. The characterisations provide an overview of essential similarities and differences between the three types of designs. In Section 6, I discuss the result and reflect on the extent to which systems, decision processes and institutions can be designed and in what way these designs are interrelated. I conclude by expressing the hope that distinguishing, as I do, between three types of designs will be helpful to design researchers who try to disentangle the variety of design activities in the context of large-scale socio-technical system development.

2 A conceptual model of design

Writing about design for an audience that largely consists of engineers, I am all too well aware that they know much better than I do on how systems in their particular technical domain are conceived, specified in detail, and eventually constructed. Moreover, every technical discipline has its own engineering vocabulary. This in itself is not a bad thing, but the disciplined mind tends to reject the otherwise perfectly acceptable views when they are phrased using one’s own pet words in a different context. As this is exactly what I may be doing in this essay, I kindly ask the readers not to halt at the first point where I give a meaning to a term that seems to be at odds with their own professional knowledge, but rather to make an effort to ‘translate’ my terminology to their own.

2.1 Design context and actor roles

The model of design that I will use as a conceptual framework is abstract so as to be very generic. The behavioural aspects of design are derived from a worldview that Jaeger et al. (2001) label as the Rational Actor Paradigm (RAP). Design is viewed as a purposeful
activity that involves various actors: individuals, teams of people, or larger organisations. These actors (inter)act in what I shall broadly refer to as a design context. Assuming that actors can think rationally and act purposefully, three main actor roles can be discerned:

- **Client.** Actors in this role have a **Client Problem (CP)** in the sense that they consider the present state of the world around them as unsatisfactory, or likely to become unsatisfactory in the foreseeable future.

- **Designer.** Actors in this role translate the CP into a **Design Problem Formulation (DPF):** an abstract description of the client problem in terms of means and ends. They then use what knowledge they have at their disposal to make a design: a representation of an artefact, the environment in which this artefact will be realised, and the actions this realisation requires. This artefact is such that, when realised, the CP will be solved.

- **Realiser.** Actors in this role execute the design by making the artefact it represents real. They do this by taking actions prompted by the design.

The reason to distinguish between actors and roles is that in some design contexts, several actors may figure in the same role and also that one actor may assume more than one role.

### 2.2 Structure and flow

Thinking about design has led me to a particular intuition about artefacts: any artefact is something static that guides something dynamic. I will use the term **structure** to denote the static and **flow** to denote the dynamic. The idea is that structure and flow are two essential aspects of an artefact: the structure is the aspect that is realised by the realiser, the flow is the aspect that meets the client’s needs. A design describes the structure as well as the flow to express the function of the artefact, but only the structure is realised. Ropohl (1999, p.63) links structure immediately to function, but I prefer to keep flow and function as separate concepts, because some of the flows that occur once the artefact has been realised in its environment may not be intended by the designer.

In a complex artefact, the duality of structure and flow can be observed at different levels. Structures at different levels may produce flows simultaneously (vehicles riding on a road, rainwater draining away through the porous pavement, runoff flowing through sewer pipes to a water treatment facility) or sequentially (concrete cast in moulds to form sewer pipes, a tightly scheduled workflow to lay the pipes and the various pavement layers). More precisely, a flow is co-produced by a structure and its environment (a traffic flow only occurs when there are people who want to travel, the drainage flow only occurs when it rains).

Each of the structures (road network, road surface, mould, sewer pipe network, project schedule) has been designed and realised as an artefact. Some structures (mould, project schedule) are ‘transient’ in the sense that they guide a temporary flow that produces a new structure, after which the transient structure is obsolete. To function as intended, both the artefact and the environment in which it is placed must be relatively stable.
2.3 Design and development

The word ‘design’ is a noun as well as a verb. Design-as-verb denotes a purposeful intellectual activity that produces a design-as-noun, where this design-as-noun is a representation of the intended form and function of an artefact that provides sufficient guidance for the realisation of this artefact. The design activities and realisation activities are supposed to be separated:

\[
\text{design activities} \rightarrow \text{design} \rightarrow \text{realisation activities} \rightarrow \text{artefact.}
\]

The design activities include communication between the designer and client, acquisition of knowledge about the environment in which the artefact is to function, DPF, generation of alternative (partial) representations of the artefact, its context and the realisation activities, evaluation of alternative designs, and eventually selection of the design that is to be realised. The realisation activities include modification of the environment (to better accommodate the artefact), acquisition of resources, production/construction of components of the artefact, and eventually delivery to the client.

The reification of design-as-noun and the postulate that a design-as-noun, being a representation of the intended form and function of an artefact, precedes its realisation are essential to the idea of design-as-verb in the context of rational choice. A design-as-noun allows an \textit{ex ante} assessment of the flows the artefact will co-produce in the real world, and it allows the designer (and the client and the realiser as well) to judge the merits of alternative designs. Without the separation in time – by a moment of rational deliberation – of design-as-verb and realisation, the artefact would not be designed, but developed. As realisation activities proceed, the “time and space horizons for design” (Simon, 1981, p.178) move, revealing new means and ends that call for a change in the design.

The distinction between developing an artefact in a mesh of design activities and realisation activities, and designing an artefact and subsequently realising it is closely related to the notion of stability. Unlike development, design requires that the form and function of the artefact as well as the environment in which the artefact is to function are stable, at least for the period of time during which the artefact is to perform its function.

2.4 The design process

The design and realisation of an artefact in response to the CP can be viewed as a sequential transformation process:

\[
\text{Client Problem (CP)} \rightarrow \text{Design Problem Formulation (DPF)} \rightarrow \text{Design (D)} \rightarrow \text{Artefact (A)}.
\]

The DPF is produced by the designer, albeit in interaction with the client. It is the designer’s professional skill to transform the CP (= dissatisfaction with a (future) world state) into a DPF: a precise definition of goals and constraints (operationalised in terms of performance indicator/target value pairs) and available means (operationalised as a set of design variables/feasible option range pairs). A complete DPF also defines a \textit{test}: a procedure that, when executed, produces the answer to the question “how well will the artefact that is specified by this design solve the CP when it is realised?”. The designer then proceeds by generating and testing alternative designs (selecting a single option for
each design variable and estimating whether the chosen combination will affect the performance indicators in such a way that their respective target values are achieved, until he has found a design that suffices for the client (i.e., passes the test) and satisfies the designer. Eventually, the realisation activities result in the artefact A.

As stated in Section 2.1, a complete design represents the artefact, its environment and the realisation activities required. This implies that, using the terminology of Section 2.2, a design-as-noun is itself a transient structure that guides a flow of realisation activities. The same can be argued for the DPF as an intermediary ‘artefact’ between the CP and the design. Similar ‘linguistic artefacts’ (Bucciarelli, 2002, p.225) can be identified when looking even more closely at design processes. All of these artefacts – maps, diagrams, schedules, budgets – are structures that guide the interaction between different actors in a design process.

The suggestion that the transformation process CP → DPF → D → A is a simple sequential process is misleading, as several iteration loops can be identified in this sequence:

- the problem analysis loop CP ↔ DPF in which the client and designer interact
- the solution finding loop DPF ↔ D, which is the domain of the designer
- the implementation loop D ↔ A in which the designer and realiser interact
- the evaluation loop A ↔ CP in which the client discovers that he has a new problem.

Although these loops are often used to represent the dynamics of a design process as in Figure 1(a), they hide the underlying transformations that actually proceed sequentially in time as in Figure 1(b).

![Figure 1](image)

**Figure 1** Design process as a sequence of transformations

The outcomes of the two transformations CP → DPF and DPF → D are largely determined by the designer’s knowledge and skills. The DPF and D will be biased by what in the literature on social construction of technology is called the designer’s regime (Bijker et al., 1989). Poor communication skills may lead to a flawed CP → DPF transformation known as an ‘error of the third kind’ or ‘type III error’, because it leads the designer to look for solutions for the wrong problem (Mitroff, 1974). In the transformation D → A, the realiser may deviate from the design due to misinterpretation of D or in response to changing circumstances in the context of A.
In the examples that follow in the next three sections, I will create variety in designs by assuming what may seem to be caricatures of regimes. Interesting though they may be, I will not dwell on type III errors, iteration loops, or aberrations that may occur during the realisation of artefacts, because this would distract from the principal purpose of this essay, which is merely to illustrate three types of designs relevant to socio-technical systems.

3 System design

In the first fictitious design case, the client is some governmental agency that wishes to significantly reduce the CO\(_2\) emission by traffic. For this example, let us say the ambition level is a reduction of at least 30% on a national scale. The client does not know how to achieve this goal, and therefore has a problem. One can easily imagine that, when asked to deal with this problem, designers from different regimes would make different interpretations of the problem, which would result in different design trajectories:

• An electrical engineer might consider replacing the combustion engine in cars by an electromotor, powered by batteries that would be charged with ‘green’ electricity, or by fuel cells (either non-carbon or with a much higher fuel efficiency than current engines). The transport system would otherwise remain unchanged; only the petrol stations would need modification.

• A control engineer (with a taste for economics) might propose a road pricing system that would track the movements of vehicles and charge their owners for using the road network. Tariff differentiation in space and time would allow the National Government to discourage use by certain user groups (e.g., high-emission vehicles) or during certain periods of the day (e.g., to reduce congestion). Both measures would lead to a lower CO\(_2\) emission.

• A civil engineer might think of an even more radical modification of the transport infrastructure, changing from a system based on petrol-powered automobiles with free access and movement on an open network of roads to a public transport system based on electrically powered vehicles such as (metro) trains operating exclusively on a dedicated network, supplemented with a highly regulated taxi system to cover the ‘last mile’.

In all cases, the system engineers would not only consider technical feasibility and CO\(_2\) emission levels, but also a range of other system performance indicators, including transport service level, safety, and of course cost. Their designs would also include a ‘migration path’ describing how the transition from the present system to the new system could be made over time. The National Government could complement the system design with tax policies to balance costs and revenues.

Obviously, these sketches do no justice to the complexity of the design problems, but they do illustrate these characteristics of ‘system design’.
The artefact is ‘tangible’ in the sense that it involves physical objects such as roads or rail tracks, vehicles, batteries, engines, RFID transponders and computer networks.

The designs will typically consist of technical specifications, drawings and models. These ‘linguistic artefacts’ are highly formalised and the associated conventions provide strong guidance for design activities.

The designs are based on a large set of assumptions with respect to the artefact and the environment in which it is to function. These assumptions are predominantly technological, derived from the natural sciences; much less from the social sciences, with the exception of economics.

The environment is assumed to be stable, or at least predictable to such an extent that the robustness of an artefact – its ability to function within certain ranges of conditions that can occur in its environment – can be assessed. The range of conditions is typically specified in the DPF.

Computer models typically perform the test that is defined in the DPF to verify whether the designed artefact solves the CP when it is realised. These models are constructed on the basis of the same set of assumptions that are used as the basis for the design.

4 Decision process design

The second fictitious design case revolves around the same client: some national government agency wishing to reduce CO₂ emissions of traffic by 30%. But now the DPF is quite different. The artefact to be designed is not a traffic system that, when realised, will deliver the transport service capacity of the same level as the present situation while emitting 30% less CO₂, but a decision process that, when realised, will result in a transport policy that will achieve the goals and also have the support of all relevant stakeholders (industry, citizens, etc.).

To illustrate that the concepts from Section 2 also apply to this kind of design problem, I will sketch four different design trajectories that lead to very different decision process designs. These design trajectories are inspired on the four ‘calculated rationalities’ described by March (1978): bounded rationality, contextual rationality, game rationality and procedural rationality. Although these rationalities are not true ‘regimes’ as defined by Bijker et al. (1989), they do reflect particular schools of thought in the policy literature.

When taking bounded rationality¹ as the basis, the designer would define a decision process that aims at a simplification of the problem formulation to such an extent that the solution space can be searched within reasonable time. The design would typically describe the decision process as consisting of the three phases defined by Simon (1977, p.41): an ‘intelligence’ phase in which goals and constraints would be defined, a ‘design’ phase in which various alternative system designs similar to those outlined in the previous section would be developed by system engineers, and a ‘choice’ phase where the alternative systems would be evaluated on the basis of a limited set of criteria for which impact assessment methods are available, such as for
example the transport service level (% of demand met by the system) and cost effectiveness (kg CO₂ year⁻¹ €⁻¹). Eventually, the system with the highest cost effectiveness that would meet a pre-defined standard for the transport service level of, say, 90% would be selected.

- The ‘regime’ of contextual rationality would lead the designer to define a decision process that would let the actual circumstances define the decision-making agenda. The National Government agency would respond to political pressures of the moment, and the decision would ‘happen’ as soon as a ‘window of opportunity’ – a propitious configuration of problems (not only the CO₂ problem, but also any other business at hand, such as high oil prices, badly congested highways, and a referendum on the EU constitution coming up), feasible solutions and political will – would open (Kingdon, 1995).

- A decision process design based on game rationality would define incentives in such a way that the maximum overall utility is achieved when all actors strive to maximise their individual interests. Such a design could, for example, involve a policy based on the market value (in € kg⁻¹) of international CO₂ emission rights. Using this figure, the economic value of 30% of the present total CO₂ emission by traffic would be established and the National Government agency would use this budget to put out a tender for CO₂ emission reduction schemes (positive incentive). To urge the industry to actively develop such schemes, the National Government agency would use the threat of a CO₂ tax on all carbon-based fuels (negative incentive), the revenues of which the Government would use to buy international emission rights.

- The ‘regime’ of procedural rationality would direct the decision process designer to develop decision-making procedures and conventions that are most compatible with tradition. This could of course be something very close to the ‘usual politics’ of the design based on contextual rationality, but one can also envision a decision process design that copies the approach taken to reduce waste from packaging (OECD, 1998; de Bruijn et al., 2002) to strike a delicate balance between content, progress, openness, and protection of stakeholder interests.

Like in the previous example, these sketches are very ‘broad brush’, but they do illustrate these characteristics of ‘decision process design’:

- The artefact is ‘intangible’ in the sense that it is a process of social interaction. A decision process does not involve the kind of physical components that together constitute a technical system.

- The designs typically consist of ‘rules’ that actors should observe in the course of the decision process. These rules are designed to guide the actors participating in this process towards consensus on the policy issue that is to be decided upon. The ‘linguistic artefacts’ in the design are informal (‘natural language’) and provide limited methodological guidance for designers.

- The designs are based on a large set of assumptions with respect to actor behaviour in response to these rules. These assumptions are predominantly behavioural, derived from the social sciences, notably those addressing aspects of decision making and rationality (Jaeger et al., 2001).
• The environment in which the artefact is realised is a socio-political system that usually is very dynamic. The designed set of ‘rules’ can only be realised as intended when they are sufficiently compatible with established decision-making practices.

• The test that should be part of the DPF is usually implicit: the design meets the test when the rules it defines are accepted by all the actors involved.

5 Institutional design

Institutions are defined by North as

“the rules of the game in a society or, more formally, the humanly devised constraints that shape human interaction. In consequence, they structure incentives in human exchange, whether political, social, or economic.”

(North, 1990, p.3)

By this definition, marriage is an institution. Although the ‘rules of the game’ associated with marriage may change from culture to culture, they evidently shape human interaction. Whole kingdoms have been joined through marriage, causing great shifts in wealth and power more easily than war, and making the relation between princes and princesses an object of strategic planning. Likewise, the rules for succession have affected (and still do affect) both social and political behaviour. Constitutional monarchies, people’s republics and all other forms of Government are institutions, as they define how states are ruled.

By looking for the ‘rules of the game’ in a society, one can identify a variety of institutions: legal courts, systems for education, healthcare, registration of property rights, taxes, trade, industry, etc. Although there may be no clear hierarchy for institutions, social and economic institutions will always be constrained to some extent by political and governmental institutions.

Institutions are ‘formal’ insofar as the ‘rules of the game’ have been codified in laws and regulations. Laws and regulations can be changed, which suggests that institutions can be designed (Weimer, 1995). To illustrate the concept of institutional design, I will again use the fictitious case of a National Government that looks for a way to regulate traffic in such a way that CO2 emission will be reduced. As in the two preceding sections, I will sketch different design trajectories that correspond to particular problem conceptions. This time, the variety is not the result of differences in technological regime (the car with combustion engine remains the principal mode of transport) or decision rationality (I disregard the policy process), but of different political philosophies that I have associated rather casually with three forms of social structures: markets, hierarchies and networks (Thompson et al., 1991).

• A designer inclined to liberal capitalism would choose to design a market. To solve the CO2 emissions, the Government could issue tradable emission rights up to the maximum acceptable level, and the ‘invisible hand’ of the mechanism of supply and demand would ensure maximisation of the overall utility of the traffic system. The market would automatically provide strong incentives for the development of low-emission vehicles, but unfortunately also for finding ways to circumvent the system.
A designer inclined to centralistic communism would prefer to design a hierarchy to regulate the use of automobiles. Both cars and roads would be public property. The demand for transport would be investigated by region, and, based on this demand, resources would be allocated by authorities on different levels down the hierarchy. The plans for production and allocation of resources would be optimised within the constraints of available technology, cost-benefit assessments, and of course the CO2 emission limits.

A designer inclined to idealistic socialism would design a network for voluntary resource sharing amongst individuals. The location, itinerary and seat availability of privately owned and operated cars would be made public to facilitate car pooling, there would be guidelines for ‘proper conduct’ when signalling a vehicle to catch a ride, and there would be incentives for sharing, such as dedicated highway lanes for cars with more than one passenger.

Obviously, the outline examples are caricatures, but they serve to illustrate these characteristics of ‘institutional design’:

Here, too, the artefact is ‘intangible’ and the design consists of rules, but there is a marked difference with a decision process. A decision process is a ‘one of a kind’ sequence of interactions (e.g., one that leads to the decision to subsidise bio-fuel), whereas an institution moulds a multitude of such processes (e.g., anyone who wants to drive a car will apply for a drivers license). The design of an institution is realised when the ‘rules of the game’ that constitute the institution have been internalised by society to such an extent that they indeed give shape to human interaction. The realisation of the artefact is this ‘internalisation’: the institution becomes a structure that subsequently changes the flow of social interaction.

Like decision process designs, institutional designs are based on assumptions with respect to actor behaviour in response to these rules, but where decision process designs are highly pragmatic and aim for closure on a particular decision within a relatively short time span (months or years), institutional designs are strongly influenced by ideology and aim for changes in social interaction that last for generations.

The assumptions that underlie institutional designs are based on theories from economics and social science that address behavioural and cultural aspects of collective action (Schotter, 1981; Ostrom, 1990; Thompson et al., 1991).

Newly designed institutions must be closely aligned to existing institutions. The more a newly designed institution is incompatible with the institutions that are in place, the less likely it is that the design will be accepted or, when realised as artefact, will constrain actor behaviour as intended. For example, when emission rights are imposed, a black market for ‘bootlegged’ carbon fuel is likely to emerge. The model of tradable emission rights fits nicely with the present liberalisation trend in western economies. The central resource allocation would require a large and powerful bureaucracy to be already in place. The car-sharing model would only fit in a society in which social networks are already strong and ‘helping your neighbour’ is a matter of course.
Although certain design principles for the design of institutions have been formulated (Ostrom, 1990), the test that should be part of the DPF is mostly conceptual and argumentative. Compatibility with the existing institutions is difficult to test, but social simulation and gaming are developed for this purpose (Duke and Geurts, 2004; Hedström, 2005; Mayer and Veeneman, 2003).

6 Discussion

The obvious differences (tangible/intangible artefact, ‘hard’/‘soft’ scientific disciplines as knowledge base for designers, many/few ‘linguistic artefacts’ that bring rigor to the design process) separate system design on the one hand from decision process design and institutional design on the other hand. These differences also seem to offer strong arguments for the idea that decision processes and institutions develop, whereas systems can be designed. Technology implies the capacity to predict the flows that will be co-produced by a structure and the environment in which it is placed. This capacity is largely provided by scientific research. When compared with the physical sciences, the social sciences to date have a very limited prediction capacity. By consequence, the “time and space horizons for design” are so near that design as a process of rational choice is virtually impossible. This may explain why the literature on policy- and decision-making rarely mentions design, and why decision technology thrives only in contexts in which a single rationality dominates (and then automata usually drive out human actors from the decision process, which to my eyes turns decision process design into system design).

Yet system design for technical artefacts in the context of large-scale socio-technical systems remains difficult. First, because on the long time horizon of the design problems of the type I have discussed in this essay, the environment in which the artefact has to function is unlikely to be stable. Second, because these design problems only rarely have a single client, a single designer and a single realiser. They are ‘messy’ multi-actor problems of a complexity that exceeds that of the ‘tower of Babel’ negotiations referred to by Bucciarelli (2002, p.220) because in the context of large-scale socio-technical systems, the design of tangible artefacts (technical systems) and the design of intangible artefacts (‘rules of the game’ that guide social interaction) are interlaced in a larger process of design. Even for my fictitious cases, a more elaborate analysis of this process as a succession of design transformations would not produce a single, sequential chain like the one in Figure 1(b), but more likely a directed graph in which many paths split and join later on as different design sub-problems – some dealing with technical artefacts, some with social artefacts – are concurrently formulated and re-formulated by different (although possibly overlapping) multi-actor design teams over time in a process as depicted in Figure 2.

Figure 2  Design process as a system of concurrent transformations
The graph in Figure 2 illustrates how a client problem CP1 evolves (say from CO2 emission reduction to more efficient use of transport infrastructure) as client and designer interact and settle on two design problem formulations: DPF2 concerns a social artefact (say the institutional framework of rules and regulations that provide incentives for transport efficiency) and DPF3 a technical artefact (say the systems needed to implement the new regulations). For the social artefact, two alternative designs are considered (say electronic road pricing and a new tax scheme for cars and petrol), and as the public policy process proceeds, a private consortium – anticipating on the policy outcome – works on the design D1 of a technical artefact (road pricing infrastructure). Once the institutional design D1 has been chosen and has been realised (laws ratified by parliament, bureaucratic support in place), the technical design problem is reformulated (say to deal with privacy issues) as DPF4. Of two alternative designs (say GPS tracking and toll booths), the first is chosen and realised. Now that both artefacts A1 and A2 are functioning in their environment (road pricing becomes a reality on highways), a new CP emerges due to unforeseen effects (say toll evasion) or changes in the environment (say tax compensation schemes that neutralise the incentives for efficiency).

Obviously, the real process would consist of many more interacting transformations, but even in this highly simplified example, it can be seen that for different transformations in the process, different actors play the role of client, designer and realiser. The original clients are actors who are concerned about CO2 emissions. As the focus shifts to traffic, the process of DPF moved to the transport policy arena with a new diversity of stakeholders. As the road pricing concept appears to become dominant in policy discourse, certain industrial actors anticipate by making an innovative design and win over the more conventional bids by their competitors.

The point I want to make is that even when each of the transformations in Figure 2 is the result of rational choices of actors, the process itself and the resulting system are emergent. It is tempting to suggest that process like the one in Figure 2 can be designed in the sense that ‘rules of the game’ for such processes can be designed. However, to ensure that actors will behave in accordance with the structure provided by these rules, another decision process would be required in which the actors agree to these rules. This line of reasoning leads to an infinite regression.

The following example – taken from my personal experience – has a simpler design context: one organisation is both designer and realiser, and in a way also the client (the organisation wishes to make money on the system). The example illustrates the interdependence between system design and institutional design, and also a form of adaptation by which a system develops that is different from the implementation loop D ↔ A.

Greenwheels (Rosenthal, 2005) is the name of a Dutch car-sharing organisation that in ten years time has expanded from a three cars in Rotterdam to some 500 cars in 50 cities in The Netherlands. Five years ago, I obtained an electronic pass with a personal ID number and an access code in exchange for a deposit of 250 Euro and a monthly fee of 5 Euro. I use the ID number and access code when making a reservation for a car by telephone or via the internet, sometimes months in advance, sometimes literally ‘last minute’, provided that a car is available nearby. After checking ‘my’ car for any damage, I hold the electronic pass close to the windshield. The on-board computer then contacts the central server via the mobile telephone network (using SMS), and seconds later the doors will unlock, provided that I indeed have made a reservation for that moment. If the car is damaged or dirty, I check the paper log to see whether this has been
noted already. If not, I make a note myself and report the damage to the Greenwheels operator to avoid being held responsible myself. A small display reminds me that I must enter my access code before I can take the ignition key from the glove department. Seconds later, I am on my way. The tank should be at least one quarter full, and I’m supposed to return it that way. At the petrol station, I pay electronically using the Greenwheels debit card. At all times, the on-board computer displays my travel time and distance, as well as the time my reservation ends. At the end of my trip, I return the car to its own Greenwheels parking space, place the ignition key back into its magnetic slot in the glove department, and lock the car by holding my pass close to the windshield again. The trip data are then sent to the central server. Once a month, I receive an invoice by e-mail that specifies my trips in detail. The amount due, calculated on the basis of distance and time using the formula that corresponds to my contract (Greenwheels offers different formulas), is debited automatically from my bank account.

Greenwheels car sharing works so well because it combines good system design and good institutional design. The chosen technologies solve the problem that characterises common pool resources: the electronic system with ID number and access code achieves such excludability that car use becomes a private good, whereas distributing keys would lead to club goods at best (Ostrom and Ostrom, 1977). The original technical artefacts (the on-board computer and the central server) are imbedded in a stable environment (the Internet and the GSM mobile phone network). The behavioural component of Greenwheels car sharing builds on firm institutions: many of the rules that clients are to observe have already been internalised by them – you pay for your petrol, and now even with a smile because it is not your own debit card you’re using! – or they are enforced by other institutions, such as banks and the police. The rules defined by Greenwheels themselves – return your car on time and to its original place, report damage immediately, keep the tank at least one quarter full, keep the car clean – are internalised easily not only because they are close to ‘civil behaviour’, but also because there are good incentives (a fine when returning late, but only when this forced someone else to take a taxi) and effective means for social control, as rule breakers can be traced easily.

The point I want to make using the Greenwheels example is that a (smaller scale) socio-technical artefact will perform successfully only when its design is highly adapted to the (larger scale) systems, decision processes and institutions in its environment. In the case of Greenwheels, this adaptation did not occur by iterating through the implementation loop D ↔ A. Neither Greenwheels nor any other of the commercially successful car-sharing companies evolved from small, ideologically inspired grassroots care-sharing organisations (Shaheen et al., 1998, p.41); they were designed. Informed by the failure to realise earlier designs, the designers adapted the DPF, adding economic constraints (avoid bankruptcy) and, in particular, new means, not only new technologies, but also alliances with public transport companies, taxi firms, and car-leasing firms.

7 Conclusion

The objective of my essay was to clarify three types of designs that can be observed in the context of large-scale socio-technical system development. By presenting alternative system designs, decision process designs, and institutional designs in a common framework, I have suggested that actors in the designer role could in fact have considered these alternatives and worked towards a rational choice of one particular
alternative. In the discussion section, I have argued that when it concerns large-scale socio-technical systems, there is no point in taking this view of design as a process of rational choice between discrete alternatives. The fictitious road pricing example showed that such systems develop through emergence because many processes of system design, decision process design and institutional design proceed concurrently and interdependently. The Greenwheels case illustrated that, on a smaller scale, socio-technical systems develop as the DPF develops (informed by the failures of other actors to realise earlier designs) until it is so well adapted to the environment (which is also developing) that the artefact can be realised successfully.

To my eyes, however, the idea that socio-technical systems develop through emergence and adaptation does not disqualify the RAP as a reference frame for good design. On the contrary, I see in both phenomena an argument to engage in research that may improve our understanding of socio-technical system development processes, the constraints they impose, and in particular the degrees of freedom they still have. Studying such development processes as a set of interdependent activities of system design, decision process design, and institutional design, performed by different rational actors who pursue their goals in different arenas, should increase our knowledge of their interaction and, eventually, hopefully, our capability to predict their course and outcome.

References


Notes

1 I like to thank my anonymous reviewer for pointing out to me that my interpretation of the characterisation by March (1978) of bounded rationality as optimisation under constraints may not do justice to the concept as it is elaborated later by Simon (1990) – see Todd and Gigerenzer (2003) for a convincing argument on this point. For the illustration’s purpose, however, this interpretation yields a better contrast with the other rationalities he describes.

2 The fact that car pooling presently accounts for some 15% of all work-related trips in the USA (Benkler, 2004) suggests that these conditions are met to some extent by the American society.