Reinstaller, Andreas

Working Paper
The Modularity of Technology and Organisations. Implications for the Theory of the Firm

WIFO Working Papers, No. 398

Provided in Cooperation with:
Austrian Institute of Economic Research (WIFO), Vienna

Suggested Citation: Reinstaller, Andreas (2011) : The Modularity of Technology and Organisations. Implications for the Theory of the Firm, WIFO Working Papers, No. 398, Austrian Institute of Economic Research (WIFO), Vienna

This Version is available at:
http://hdl.handle.net/10419/128933

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes. You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
The Modularity of Technology and Organisations
Implications for the Theory of the Firm
Andreas Reinstaller
The Modularity of Technology and Organisations
Implications for the Theory of the Firm
Andreas Reinstaller
WIFO Working Papers, No. 398
June 2011
The modularity of technology and organisations: Implications for the theory of the firm

Andreas Reinstaller

Abstract

This paper gives a selective overview on contributions studying issues of complexity, near-decomposability and modularity in relation to economic behaviour and the theory of the firm. In the first part the paper reviews contributions studying the relationship between human problem solving in the face of complex problems and the emergence of specific technological and organisational designs. The second part the paper reviews recent research that has studied the impact of modular designs in the organisation of production at the firm level on industrial organisation and dynamics. The paper draws some conclusions on future avenues of research.

Address: Austrian Institute for Economic Research (WIFO), Arsenal Obj. 20, 1030 Vienna, Austria. E-mail: andreas.reinstaller@wifo.ac.at.

JEL Codes: L2 D21 O31

Keywords: Complexity, technical and organisational change, industrial organisation, problem solving.

Introduction
Since the late 1980s an increasing number of researchers in the field of economics and strategic management have started to conceive technologies, firms and the economy in general as complex hierarchical systems (e.g., Anderson et al 1988, Levinthal 1997, Arthur 1999). Such complex technological, economic and social systems are difficult to design and control. The concept of modularity refers to a set of principles related to the design and management of complex systems. It is based on the seminal contributions by Herbert Simon (e.g. Simon 1996) on human learning in complex environments and near-decomposable systems.\(^1\) They have given rise to a large and still growing body of literature that has studied complexity, near-decomposability and modularity in relation to economic behaviour and the firm.

This paper gives a selective overview on important contributions to this field of research. It is organised as follows: In the first part it will define the concept of modularity and examine how it is related to complex systems and near-decomposability. In the second part it will discuss several aspects related to modularity as a strategy of human problem solving. The final part of the paper then reviews recent contributions that have applied the concept of modularity to the theory of the firm.

Complexity, near-decomposability and modularity: definitions
According to Simon (1996), p. 195, a system is complex if it is “made up of a large number of parts that interact in a nonsimple way”, and where the “whole is more than the sum of its parts” insofar as given the characteristics of the parts and the ways they interact it is “not trivial to infer the properties of the whole”.\(^2\) As a consequence it cannot be comprehended and controlled by a single person. Many publications rely on Kauffman’s NK Model (Kauffman 1993) to illustrate and study complex systems in economics, organizational studies or management sciences. The NK model studies the performance of a system \(\Pi\) subject to the two parameters, \(N\), the number of components of the system, and \(K\), the degree of interdependence of these components. The system is defined as a binary string of the \(N\) components corresponding to different system configurations and associated performance levels derived from them. The number of different configurations is then \(2^N\), and the performance \(n_i\) of each component \(T_i\) is a function of the component itself and the \(K\) other components impinging on it, \(n_i = n_i(T_i; \ T_1, \ldots, \ T_k)\). The performance \(\Pi\) of the system is then given by the average over all component performances, \(\Pi=1/N \Sigma n_i(T_i; \ T_1, \ldots, \ T_k)\).

Figure 1 illustrates one possible configuration for a system with twelve components, \(N=12\), that are on average influenced by six other components, \(K=6\). We may assume that the depicted complex system is a schematic representation of a production process.

---

\(^1\) Alexander (1964) developed the idea of modularity in architecture. However, his work has not influenced subsequent research on modularity and near-decomposability in the area of economics and management.

\(^2\) Near-decomposable or modular system architecture and problem solving through decomposition are concepts that have been applied to virtually all domains of economic life. So, the term “system” used in this section is a template for concepts such as “product”, “firm”, “organization” or “technology”. The term “problem” is instead a generalized way to say that every system has a purpose and that under some circumstances a system may not be able to fulfil this purpose adequately or that for a given purpose there is no adequate system yet.
that consists of twelve tasks, \( T_i \), that are interdependent and have to be executed to produce some output at some cost and with a specific quality. Interdependencies are marked by an “\(^x\)”. The performance of any task is therefore determined by the number of entries in the related row. The entries in each column in turn indicate which other elements are actively influenced by a specific task. Elements in the lower off-diagonal part of the matrix represent feed forward linkages and therefore capture some inherent sequence or hierarchy between the tasks and the elements in the upper off-diagonal part of the matrix represent feedback linkages. They indicate that a change in a downstream task affects tasks upstream. The performance of task \( T_4 \), for instance, is affected by upstream tasks \( T_1 \) and \( T_2 \) and downstream tasks \( T_9 \) and \( T_{11} \). A change in any of these tasks will affect its performance. On the other hand, task \( T_4 \) affects tasks \( T_1 \), \( T_2 \), \( T_3 \), \( T_5 \), \( T_7 \), \( T_8 \), \( T_9 \), and \( T_{11} \), such that any change in its performance will positively or negatively affect these other tasks. Any adjustment in the production process shown in Figure 1 is likely to be difficult especially if the effect of the linkages on the performance of any task is not additive.

\[
\begin{array}{cccccccccccc}
T_1 & T_2 & T_3 & T_4 & T_5 & T_6 & T_7 & T_8 & T_9 & T_{10} & T_{11} \\
T_1 & o & x & x & x & x & x & x \\
T_2 & x & x & o & x & x & x & x \\
T_3 & x & x & o & x & x & x & x \\
T_4 & x & x & o & x & x & x & x \\
T_5 & x & x & x & o & x & x & x \\
T_6 & x & x & x & x & o & x & x \\
T_7 & x & x & x & x & x & o & x \\
T_8 & x & x & x & x & x & x & o \\
T_9 & x & x & x & x & x & x & x \\
T_{10} & x & x & x & x & x & x & x \\
T_{11} & x & x & x & x & x & x & x \\
\end{array}
\]

**Figure 1**: NK Model with \( N=12 \) and \( K=6 \) in a task structure matrix (TSM). Illustration of feedbacks from Gomes and Joglekar 2008, p. 446.

Kauffman (1993) has shown that it becomes more difficult to improve the overall performance of a system like the one in Figure 1 if its elements are more tightly coupled. If \( K=0 \), i.e. no linkages exist, then each single component of the system can be improved and contributes independently to its performance. Hence, a unique configuration of the system exists that is associated with the global optimum which can be reached by gradually improving each single component. If on the other hand, complexity is at its maximum, \( K=N-1 \) which implies that all components are linked to each other, then small changes in one component trigger adjustments in all other components. Higher interdependence therefore leads to less correlated outcomes for small changes in the elements of the system. The number of local optima increases in \( K \), and this makes it very difficult to find the global optimum. Only a complete search of the state space will allow agents to identify a global optimum. However, exhaustive search time explodes as the number of components \( N \) increases. For agents with limited time, memory and
cognitive capacities the problem will quickly become intractable. Indeed, it is comparable to other NP complete problems such as the travelling salesman problem (see Kauffman 1993, p. 63ff.). Hence, with increasing system size and interdependency the danger for a system to remain locked-in in a local optimum increases, as does the difficulty to manage the system.

As will be discussed in detail in the next section of this paper, a strategy to deal with this complexity is to decompose the original complex problem into manageable subproblems, solve these and recompose the system. This process will transform the system into a composite one, "constructed through the superposition of: (1) terms representing interactions of the variables within each subsystem; and (2) terms representing interactions among the subsystems" (Simon and Ando 1961, p. 132). The subsystems are structurally independent from each other but operate together to fulfil the primary purpose of the system. Such composite systems are referred to as near-decomposable systems. Near-decomposability and modularity are architectural features shared by most complex systems.

Figure 2 gives an example of a near-decomposable system. As compared to Figure 1 most feed forward and feedback loops between the twelve tasks, T1..T12, are now concentrated in four subsystems or modules, M1..M4, made up of three tasks each. Between these modules only a few linkages exist. Hence, the tasks within each subsystem are tightly coupled whereas across subsystems they are much more loosely coupled if compared to the system depicted in Figure 1. Now task T1 is linked to both tasks T2 and T3 through feed forward and feedback loops, whereas T1 impinges only on task T1 in module M2 and task T7 in module M3 and is in turn influenced only by task T11 in module M4. In Figure 1 instead T1 affected T3, T4, T5 and T10, and was influenced by T3, T4, T5, T6, T8 and T11. This reduction of linkages to tasks outside the own subsystem implies that each subsystem hides information insofar as other subsystems cannot access and modify this information. It acts as a black box that transforms some inputs into specific visible outputs for other modules or the system user.

Figure 2: Near-decomposable system consisting of twelve tasks, T1..T12, and four modules or subsystems, M1..M4, depicted in a task structure matrix (TSM). Adapted from Baldwin and Clark 2000, p. 60. The blocks inside the matrix capture interdependencies of tasks inside a subsystem; “x” off the main diagonal indicate either feed forward or feedback loops between tasks.
If we compare Figures 1 and 2 now with the two system arrangements depicted in Figure 3 it is possible to illustrate the departure of analyses of technical and organizational change based on complex or near-decomposable systems from more traditional approaches using neo-classical production functions. In neoclassical production theory homothetic production functions are used to capture microeconomic and aggregate production processes. This implies that proportional changes in all inputs are equal to the identical proportional change in the aggregate input such that increasing the scale of operation of each input is equivalent to increasing the scale of operation of the aggregate input. In homothetic production functions output is also assumed to vary as a function of the marginal product of each input and the input levels, 
\[ y = \Sigma_i \frac{\partial y}{\partial x_i} x_i, \]
where \( y = f(x) \) is a production function transforming inputs \( x \) into an output \( y \).\(^3\) An important property of this representation of the production process is that it implies that the technology of a firm is separable. Separability means that the marginal rate of technical substitution between two inputs does not respond to changes in another input. This implies that an optimization by stages is implicitly assumed to be possible.

Figure 3 now gives two examples of systems that capture essential characteristics of neo-classical production techniques. Panel a) illustrates an independent block task structure. Here within each module tasks are not separable as they are tightly coupled, however, the different modules are not linked to each other. Hence, each module can be changed or improved independently of other modules. A change in task \( T_1 \) in module \( M_1 \) will not affect the marginal rate of technical substitution between any two tasks executed in another module. The output of the system shown in Panel a) will be the aggregate of the outputs of each module. Panel b) in turn illustrates a sequential task structure that resembles a Neo-Austrian production function where the output is related to the roundaboutness of the production technique, i.e. it increases in the number of sequential production stages. In Panel b) these are captured by the hierarchically related tasks. Given the sequential structure of the system in Panel b), optimization by stages is clearly possible. The performance of a downstream task will only be affected by the task immediately preceding it. Both systems can be optimized as a high level of substitutability between modules or tasks is given. A simple hill-climbing optimization routine will find a global optimum. As we have argued before in complex systems this is typically not the case.

A comparison of Figures 1, 2 and 3 illustrates also that a principal objective of technical and organizational change is amongst other aspects the achievement of high levels of substitutability. However, no such feature of any production technique may be assumed to be given a-priori. The discussion also shows that interdependencies between elements in a complex system are also likely to be incompatible with the proportionality assumptions of homothetic production functions. Hence, the analysis of technical and organizational change based on complex or near-decomposable systems is more general than analyses based on standard production theory which implicitly assumes some form of decomposability without making this explicit (cf. Marengo et al 2003). This comes however at the cost of diminished analytically tractability.

\(^3\) If \( \frac{dy}{dx_i} \) is invariant with respect to changes in \( y \) then the production function has constant returns to scale.
Modularization is one heuristic through which a decomposition of a complex system into independent or quasi-independent modules can be achieved. In modular system designs the interaction between modules is typically organised according to a specific architecture and the interaction is mediated through standardised interfaces. The architecture establishes which modules are part of the system and which is their role. It also establishes a hierarchy between different modules which results from the recursive decomposition of the subsystems and their functionality. Interfaces establish how the different modules interact. They define which information or output from a module is visible or passed on to other modules. The architecture and the interfaces define the overall design of a modular complex system. The modularity of a system is therefore defined through

- a primary purpose of the system requiring specific functions or service characteristics,
- mutually balanced modules or subsystems transforming inputs into visible outputs,
- an architecture organising the modules in nested hierarchical relationships,
- interfaces connecting the modules and determining which information is hidden and which is visible across modules, and
- a complete description of all modules and interfaces such that the system is a bijective representation in the domain of possible system configurations.

Often the terms modular system and near-decomposable system or modularity and near-decomposability are used interchangeably. However, this is not correct. A modular system is one obtained through a process of modularization leading to specific design rules. These establish the architecture of the system as well as standards for the interaction of modules through interfaces. They also determine which system parameters are hidden within modules and which are visible to other modules. These design rules ensure that “plug and play” flexibility is achieved (see Baldwin and Clark 2000, p. 88ff) if the interface standards are kept constant for some time. This means that

---

4 A system can also have secondary purposes or purposes for which it was not initially developed. See Baldwin and Clark (2000), p. 28 ff.
in a fully modular system modules can be replaced by other modules operating on the same range of visible parameters and using the same interface standards without compromising the operation of the entire system. If such design rules exist and structure the interaction between sub-systems then a near-decomposable system is also modular, else it is just near-decomposable. Figure 4 gives an example of a complex system modularized through specific design rules. In this example the design rules specify for which module interacts with which, and which tasks in a module interact with other modules through standardized interfaces (implying standardized information formats and parameter ranges). Hence, as long as the design rules are observed “plug-and-play” flexibility is achieved.

![Figure 4: A modular system design. Interactions between modules are mediated standardized interfaces specified in design rules. The matrix contains now a column specifying which module interacts with other modules and through which tasks this interaction is managed on the basis of a standardised interface.](image)

Typically a system will not be fully modular, but there are special cases where this is true. The two extremes on the scale are “fully modular” and completely “integral” (cf. Ulrich and Eppinger 2008). Modularization increases with the independence of components which is determined by

- the coupling between components, which is tighter when the number of modules increases as discussed in relation to Figure 1.
- the substitutability of components in terms of whether they are easy to replace or whether they can be easily used in other systems, and
- the degree of standardization of modules and interfaces.

Full modularity (Fig. 4) implies a loose coupling between modules, perfect substitutability and complete standardization of modules and interfaces, whereas perfect integration (Fig. 1) captures tight coupling, a very limited substitutability and the absence of standardization of modules and interfaces.
The behavioural foundations of near-decomposability and modularity

Problem solving and the emergence of near-decomposable or modular designs

According to Simon (1976), p.293, “factoring the total system of decisions that need to be made into relatively independent subsystems, each one of which can be designed with only minimal concern for its interactions with the others” is the solution to manage complex systems such as firms. This behavioural strategy is typical for human problem solving in the face of substantive and procedural uncertainty (Simon 1978). It requires a division of labour as well as a division of information and knowledge and a reorganization of tasks and decisions related to the management of the whole system. Marengo et al (2000) show that through decomposition every complex problem can be transformed into one of minimal complexity. The difficulty of transforming a complex system into a near-decomposable one is however to identify and specify the subsystems or modules such that the need as well as the cost for problem-solving across sub-system boundaries is minimised.

Contributions to the literature on modularity in the realm of economics, management sciences and strategic management generally rely on two distinct theoretical approaches to explain the micro-economic incentives to partition complex systems in order to minimise the cost of problem-solving across sub-system boundaries (cf. Marengo and Dosi 2005, Gomes and Joglekar 2008): (i) the information processing view, and (ii) transaction cost economics and imperfect contracts theory. The information processing view maintains that human information processing capabilities are limited. Therefore, the number of alternatives that can be considered and the capability to identify clear causal chains on the operation of a complex system and its effects are severely restricted. The objective of breaking down a complex problem into higher-level parts is therefore to minimise the complexity of the system such that it becomes easier to understand and manipulate. This in turn will increase the efficiency of problem solving and reduce the time needed to find solutions. The decomposition schemes will therefore reflect the efforts of system designers to achieve information processing efficiency. The transaction cost view instead maintains that if problem-solving activities necessary to manage a complex system are uncertain or non-standard and require specific assets these are best organised inside specific modules. This will minimise the management cost and economise on transaction costs across modules. The decomposition schemes will therefore be determined by the relationship-specific features of the assets of the agents involved in the problem solving process.

If a system design has to be modular and not just near-decomposable the search for suitable partitions of the system will also involve the specification of interfaces and

---

5 In cognitive psychology the phenomenon is referred to as “chunking” (cf. Gobet et al 2001).
6 Using terminology from computer science the literature refers sometimes to the boundaries between different modules in a system as “encapsulation boundaries” (see for instance Langlois 2002).
7 Baldwin (2008) mentions also knowledge-based theories of the firm as these establish that the boundaries of a system (typically a firm) shift with changes in the knowledge-base. As Baldwin points out these contributions are generally not able to specify precisely the system architecture and the system boundaries. For this reason they are not considered here.
modules prior to any other design and problem solving activity. A designer called
“architect” (Baldwin and Clark 2000, p. 68) establishes the central design parameters of
the system a-priori at four levels (see Ethiraj and Levinthal 2004, p. 162): (1) the
appropriate number of modules, (2) the mapping of service characteristics and technical
characteristics to the modules,8 (3) the interaction of service and technical
characteristics within each module, and (4) the interfaces defining the interaction
between the modules and specifying what information remains hidden and which is
available to other modules. In other words, an architect defines what a system will do,
which parts of the system will do what, and how they will fit together, connect and
communicate. She establishes design rules that are binding for any other agent working
on the system. By defining crucial aspects of the system before starting any other
problem solving activity the architect restricts the search to specific parts of the space of
possible designs.

It is much more difficult to design modular systems than comparable interconnected
systems as the designers must possess considerable knowledge of the inner workings of
the system and its elements (Baldwin and Clark 1997, p. 86). As architects and system
designers will usually construct design rules on the basis of previous knowledge their
search for solutions is at first localised. To extend the breadth of their search process
they can rely on a number of actions to change the given structure (or first guesses about
the actual structure of a complex system) into a new structure. According to Baldwin and
Clark (2000), chapter 5, these actions are

- splitting a system into two or more modules;
- substituting one module design for another;
- augmenting the system by adding a new module;
- excluding a module from the system;
- inverting the (hierarchical) relationship between modules to create new design
  rules;
- porting a module to another system.

Splitting is the typical decomposition operator that subdivides modules in order to
disentangle their inner complexity. Through substitution existing modules are replaced
by better ones. Substitution is therefore an action leading to a modular innovation in a
system (cf. Langlois and Robertson 1992, Sanchez and Mahoney 1996). This type of
innovation does not alter the structure or architecture of a given system. The next three
actions, augmenting, excluding, and inverting are instead operators leading to an
architectural innovation (cf. Henderson and Clark 1990) insofar as the structure and the
design rules of a given system will change as a consequence of their application. Through
augmentation new modules are added to a given system whereas through exclusion a
module is left out. Through inversion information previously hidden to other modules is
moved up the design hierarchy so that it becomes visible to other modules (Baldwin and
Clark 2000, p. 138). In this way redundant information and related problem solving
tasks can be consolidated into one single module.

8 In other words, system analyst and designers establish a link between service characteristics providing value to the users of a system and
the internal structure of the system, i.e. its technical characteristics (cf. Savioiti and Metcalfe 1984). Baldwin and Clark (2000) talk in this
context about the definition of „design parameters“.
Finally, porting occurs when a module breaks loose of a particular system and is able to function with a number of different systems. For this to work it needs some functionalities translating information from different sources in such a way that it can be processed internally and the outcome transmitted back to other modules. Portable modules are essentially interface techniques that support combinatorial innovation in the sense that functionalities of different systems can be linked together to obtain a larger range of service characteristics to final users of the system, or by making specific services of one system available to another system. However, portability is not equivalent to another important aspect of learning and adaptation in complex environments which is integration (or crossing-over). This is a process where parts of two modules are combined into a new module to produce a new set of service characteristics (cf. Holland 1992, p. 97 ff). This operator is at the core of recombinant search. Only this operator will allow agents to explore the design space extensively and generate any design (cf. Fixson and Park 2008, p. 1310). However, its application potentially reduces the degree of modularity in a system.

Baldwin and Clark (2000), p. 144, argue the listed operators are complete insofar as every conceivable modular design or system structure can be generated from them. This implies that by scanning the design space broadly agents may be able to find a decomposition that effectively maximises a given objective function. Schaefer (1999) however has proven rigorously that finding an optimal modular design partition is an NP-complete problem. Hence, it is unlikely that agents will be able to find the decomposition of a system maximising a given objective function in finite time. Designing a modular system that captures the “real” underlying structure of a given design problem is therefore equally hard in terms of the computational burden as finding a global optimum in a complex system. Furthermore, the restriction of search induced by design rules may imply that those parts of the design space where an optimal decomposition can be found are excluded a-priori from the search.

Generally the optimal partition of a complex system into manageable modules with respect to some objective function is not known. The decomposition therefore happens in a gradual recursive problem solving process (cf. von Hippel 1990). The search for new solutions is based on practices experience has shown to be successful, and on forward-looking mental models that are developed to abstract self-contained functions of the system from the whole and identify related parts. The interaction of these two types of problem solving ensures that the complexity of a problem can be successfully reduced. It allows agents to scan large parts of the space of possible designs (cf. Gavetti and Levinthal 2000). The balance between these problem solving strategies will depend on the novelty of the problem. If the problem solving process involves little novelty agents learn much from prior experiences, if instead novelty is considerable they need to develop forward-looking mental models on the relationship between the functionalities of a system, the value they provide to its user and the structure and components of the system. These are essentially beliefs about the true state of the world. As the designers are proven wrong during the development process they will frame the relationship between these concepts through local, experience based learning (cf. von Hippel 1990).
Benefits and dangers of near-decomposability and modularity

A crucial advantage of near-decomposable or modular system architectures is that they allow different parts of the system to be worked on concurrently. They also cut down on the coordination burden between the people involved in the search for solutions to sub-problems. This shortens the time needed to improve or adapt the system vis-à-vis systems with a comparable degree of complexity that is not near-decomposable or modular. Furthermore, near-decomposability and modularity accommodate uncertainty related to both the operation of the system and its capability to adapt to changes in its environment that affect its capability to fulfil its purpose and hence its value in use. In an innovation context these two types of uncertainty may be viewed as technological uncertainty and market uncertainty. One the one hand, as it is possible to single out working functions separate improvement, modification and testing lead to a better understanding of the causal relationships underlying the operation of the system. This narrows down the search space for solutions to problems and reduces uncertainty related to the system itself. On the other hand, near-decomposable systems are easier to reconfigure due to the low degree of interdependence between modules which ensures that sub-systems are easier to replace. This increases the domain of potential applications as a system can be more easily adapted to suit changing purposes. This reduces uncertainty related to the environment.

Complex systems that are near-decomposable or modular are also adaptive. Simon (1962) therefore conjectured that near-decomposable or modular systems will, on average, be better able to compete in environments subject to constant change and unanticipated external shocks than comparable complex systems that are not near-decomposable or modular due to their better rate of adaptation. This conjecture has been validated in simulation studies carried out by Frenken, Marengo and Valente (1999). They have demonstrated that the decomposition of the system into sub-systems, even if it is only an approximate decomposition, allows agents to improve their performance in very short time. Therefore, in evolutionary environments where constant and speedy adaptation is crucial for survival, strategies aiming at a high rate of improvement of a complex system outperform optimizing strategies which aim at the maximum end result. However, near-decomposable or modular systems generally will not reach, or even approach, the global optimum in the state space. Their crucial advantage is their ability to adapt more quickly and therefore to improve their performance more rapidly vis-à-vis changes in the environment in which they operate. Indeed, a recent paper by Gomes and Joglekar (2008) shows that modularity increases the transactional efficiency and reduces task completion times.

There are, however, limitations to modularity that under some circumstances outweigh the benefits. Some authors point out that the development of modular system architectures may not pay off (cf. Langlois 2002, Arthur 2009). They involve high fixed costs that do not arise in the development of comparable integrated systems.\(^9\) As a

---

\(^9\) The IBM /360 mainframe computer introduced in 1964 was one of the first strongly modularised products that have been introduced on the market. For this reason it is often used as a case in contributions on modular product development (cf. Baldwin and Clark 2000, chapter 7). The sunk cost investment for the 360 series is estimated at $5 billion in 1966 USD of which between $500 million and $1 billion was development costs. IBM (successfully) bet the company on the development of this product. At present day value this amount is close to what the US government has spent in the Manhattan project.
consequence it pays to partition a system only if there is a sufficient volume of use for all modules such that the fixed costs of modular design decrease in the intensity of use of the system.

Another strand of literature points out that for modular systems there is a trade-off between the speed of search and the breadth of search (cf. Fleming and Sorenson 2001, Ethiraj and Levinthal 2004, Brusoni et al 2004). Modular systems clearly offer the advantage of fast adaptation because they enable a greater number of possible system configurations to be tested in parallel and improve the system through the substitution of modules. However, as has already been mentioned previously, modular system architectures also imply that the search for an adequate system configuration is more localised as design rules restrict the search in the space of possible designs. Integral systems instead enable a broader even though slower search.

On the one hand, due to the interdependence of their parts changes in the system reveal more readily inconsistencies or problems. Hence, integrated systems command a joint improvement of parts and therefore get locked-in to specific system designs. On the other hand, some authors argue that modularity is negatively associated with technical problem solving efforts (cf. Gomes and Joglekar 2008). One important argument in this context is that only the interplay of modularisation and integration (i.e. recombinant search) leads to break-through innovations (cf. Fleming and Sorenson 2001, Fleming 2001, Sorenson 2003, Arthur 2009). Modularisation without integration will quickly exhaust the potential to create new combinations and therefore get locked-in to specific system designs. Exclusive reliance on modularization also blinds the designers for potentially important interactions at the level of modules. Indeed, modularisation is an instrument for the incremental improvement of systems. Several studies therefore have suggested that designers should strive for the development of near-decomposable systems with an intermediate degree of interdependence across modules (cf. Fleming and Sorenson 2001, Ethiraj and Levinthal 2004, Rivkin 2001, Rivkin and Siggelkow 2006).

Competitive advantage typically not only depends on the speed with which systems can adapt to changes in the environment, but also on the sustainability of this advantage over time. In the presence of spillovers and imitation competitive advantages arising from the modularisation of a system may erode rather quickly. Components and the system architecture are easier to imitate in modular systems as modularisation implies also a high level of codification of knowledge. Hence, there is a trade-off. Modularisation enables a fast improvement of a system and performance gains, but at the same time it lays also the basis for this competitive advantage to be eroded through imitation. A number of studies have shown that intermediate levels of interdependence in a system offer important benefits when the aim is to sustain a competitive advantage over time. Rivkin (2001), for instance, shows that the largest performance gaps between innovation leaders and imitators arise when innovation leaders design their systems with moderate interdependence. Ethiraj et al (2008) instead provide evidence that performance differences between innovation leaders and imitators are persistent for integrated systems and systems with intermediate interconnection.
Modularity and industrial organisation

As technologies, firms or more generally organisations and the economy in general are best conceived as complex hierarchical systems fulfilling some purpose (e.g., Anderson et al 1988, Levinthal 1997, Arthur 1999) the concepts of complexity, near-decomposability and modularity have been applied to different level of aggregation in economic analysis. The most prominent strand of literature has applied them to firms and their strategies in the areas of organisation, innovation and product design. Others instead have explored the implications of the emergence of modularity at the firm level on industrial organisation. This section will present some of the aspects this rich body of literature has discussed.

The mirroring hypothesis of problem-solving and system structure

Most of the recent work on modularity in the area of strategic management and the theory of the firm can be traced back to the contributions by Henderson and Clark (1990), von Hippel (1990), and Langlois and Robertson (1992). These papers explored the relationship between product architecture and the knowledge and information processing structure of a firm, task partitioning in the innovation process, and the innovative potential of industries based on modular products respectively. These seminal contributions rely on the concepts of problem decomposition and near-decomposability as the basic tool kit in their analysis. A remarkable aspect of these papers is that they share what some authors call the “mirroring” hypothesis (cf. Baldwin 2008). This hypothesis links the knowledge and information processing structure that emerges from the process of problem decomposition to the internal structure of the system that is being designed (cf. Henderson and Clark 1990, p. 27). More specifically, all these early studies link an organisation’s task structure to the actions undertaken to develop, make and sell products. This has implications for the theory of the firm.

Figure 5: The hypothesis on fundamental isomorphism between task structure and design structure.

Figure 5 gives a summary overview on the bottom line of the “mirroring” hypothesis and on the implications it has had in subsequent studies on industrial organisation, the theory of the firm or product development. In order to solve a complex problem tasks will be divided between different people. The way these people communicate, share information and knowledge reflects the hierarchical relationships and interdependencies among the components of the system and their technical characteristics (or design
parameters). Hence, task structure and design structure are assumed to be isomorphic (cf. Baldwin and Clark 2000, p. 47).

The “mirroring” hypothesis can be extended to organisations and products as particular task structures create specific patterns of knowledge in an organisation. People exchange information and knowledge within workgroups that deal with closely related tasks. Exchanges across workgroups in contrast become less frequent. The organisation therefore reflects problem solving efforts related to the design and production of an artefact. If the task structure mirrors a rather integrated internal structure of the product it has designed an organisation may become quite inflexible (Henderson and Clark, 1990). Sanchez and Mahoney (1996) suggest that modularity offers ways to preserve flexibility: if firms design modular products they will create modular organisations which in turn will increase their strategic flexibility and as a consequence raise their economic performance. At the firm level the “mirroring” hypothesis therefore links the architecture of the products of a firm to its performance through the transmission channel of strategic flexibility.10

Moving across the boundaries of the firm the “mirroring” hypothesis implies that modular product-and organisational architectures favour vertical and horizontal disintegration (cf. Langlois and Robertson 1992). By limiting the interconnection between modules, by standardising the information exchange, and by hiding information (and knowledge) inside specific components modularisation creates breakpoints where firms are able split up and outsource activities, or where new industries emerge as the modular design of a product with “plug-and-play” offers opportunities to entrants with ideas on how to improve or extend its functionalities. Hence, the existence of complex modular products will lead to an increase of the specialisation in an industry and intensify the division of labour in the economy as a whole (see also Baldwin and Clark 2000).

Combined together the two main implications of the “mirroring” hypothesis suggest that as modularity conveys important competitive benefits products are generally becoming more modular over time and this development is associated with changes in the organisation of firms and in the structure of an industry. This will be discussed on the basis of a few important contributions in the remainder of the paper.

**Modularity and the boundaries of firms**

**The rise of modularity to a new industrial paradigm...**
Langlois and Robertson (1992) have examined industries producing modular systems such as high-fidelity and stereo sound reproduction systems or the microcomputer industry. These products are complex artefacts that offer a wide range of service characteristics to their users through different components that are produced by different firms. Common industry standards of compatibility link the firm in this industry producing the different components into a modular production network.

---

10 Strategic flexibility implies that firms are able to pursue parallel development efforts and adaptations across modules, by multiplying design options, and by achieving “economies of substitution” (Garud and Kumaraswamy 1995). Recent studies have found some mixed evidence for this hypothesis (cf. Worren et al 2002, Todorova and Durisin 2008) indicating that modularity increases are positively associated with firm performance but that firms often fail in their quest for flexibility.
According to Langlois and Robertson the division of labour between firms in these industries is essentially determined by the interplay of economies of scale of assembling the components and transaction costs to users of finding out about the different components. As both types of costs are low firms have little incentive to supply more complex products integrating all components and users have the benefit of being able to maximise the utility from using these products by choosing and assembling the different components themselves. They can also be better tuned to user needs more rapidly because innovation takes place concurrently for each component. As modular production networks seem to offer superior benefits, Langlois and Robertson conclude that their importance will increase.

Arora et al (1999) have explored the trade-off between economies of scale and transaction costs and its implications for inventive activities in modular production networks further. They argue that the production of standardised components or modules benefits from economies of scale. For this reason it is carried out by specialised firm. On the other hand, following Langlois and Robertson (1992), they argue that the combination of modules is more efficiently carried out by the users of the assembled product, or firms that are close to final demand. If now for technical reasons the production and combination of modules is not separable, producers will carry out both activities. The precondition is that the market is large enough. In this case the economies of scale will offset the transaction costs of acquiring and processing user-specific information. If on the other hand, production and combination of modules are separable specialisation between producers of standard modules and firms combining modules into complex artefacts will take place if the cost of transportation of the general modules is small if compared to the cost of interacting with local users. The implications for the innovation process are that in the former case the entire innovation process will be concentrated in a few larger markets, whereas in the latter case innovation activities will be more prominent in smaller markets where the combination of modules creates quasi-rents.

Sturgeon (2002) finally analyses the emergence of modular production networks and contract manufacturing in the US electronic industry. He goes so far as to argue that this may be viewed as the paradigmatic case of a “New American Model” of industrial organisation. Sturgeon characterizes the firms in this new industry structure to consist of two types: i) deverticalised lead firms, and ii) “turn-key” suppliers. The lead firms focus on the design and the marketing of complex artefacts. The “turn-key” suppliers, i.e. contract manufacturers, instead produce modules on the specifications provided by lead firms. Industry standards and the transmission of highly codified information lower the transaction costs in this interaction. Hence, lead firms provide new combinations, whereas “turn-key” suppliers exploit economies of scale as suggested by Arora et al (1998). In such a network barriers to entry and exit are lower as the mutual interdependence between single firms is lower as well. Hence, firms are more flexible to access and exploit location specific factors and markets. The services of “turn-key” suppliers can also be shared by a variety of lead firms. An important competitive advantage of this “New American Model” is that it builds-up external economies of scale (Sturgeon 2002, p. 489).
Langlois (2002, 2003) has explored the location and form of transactions and hence the boundaries of the firm in such modular networks further. On this basis he advances a modularity theory of the firm. In essence he argues that the modularisation of economic activities in an economy solves rights assignment and control problems leading to holdup by repartitioning and reintegrating property rights and critical knowledge across firms (Jensen and Meckling 1992). Dynamic efficiency requires that firms seek to place all relevant knowledge and property rights related to specific technological processes under their control. The aim thereby is to internalize externalities or tightly coupled activities subject to the cost of setting up and maintaining the control as well as other aspects such as the presence of economies of scale. In the extreme case each module can become the business of a single specialist firm which has complete control over all aspects of the module. Technical standards on the other hand permit to externalize mechanisms of coordination that were previously integrated into one large firm to the market. Hierarchical coordination becomes increasingly unnecessary also because modular product and process architectures reduce transaction costs and decrease the minimum efficient scale of production as external economies of scale are built up. As a consequence, large vertically integrated corporations of the “Chandlerian” type (Chandler 1977) will disappear, and we should observe a modularization of economic organizations. The invisible hand of the market will become more important again.

Schilling and Steensma (2001) provide empirical evidence that integrated hierarchical organisations have been replaced by non-hierarchical, modular forms of organisation in many industries. Using a large dataset for a number of US industries they show that heterogeneity of industries’ production processes in terms of inputs and demand drive this process. They are positively associated with greater use of modularity in the organisation of production. Modularity seems to convey a higher level of flexibility to meet uncertainty in upstream and downstream markets. Rapid technological change in an industry increases the pressure for flexibility. Therefore it favours the adoption of modular forms. The use of industry standards on the other hand reduces the need for integration. They facilitate the use of modular forms when pressure for flexibility is high.

Other empirical work also documents that in several industries loosely coupled production systems have emerged where system integrators or innovation platforms take over the role to coordinate and exploit complementary and dispersed capabilities and skills among specialised organisational units (for an in-depth review see Patrucco 2011). These are specific organisational arrangements that have the aim to share a number of core components and interfaces in complex products and production systems. They are set up in order to effectively adapt or create new complex products. Innovations are therefore created by keeping the core parts of an existing design largely unchanged and by modifying or adding non-core components. This implies that novelty is generated by reconfiguring products without redefining their entire architecture. In this way firms are able to cut innovation costs. The presence of system integrators or innovation platforms

---

11 Langlois refers to this process as “demodularisation”.
12 Reinstaller (2007) shows that such a development depends also on the institutional arrangements surrounding the firm. The repartition and reintegration of property rights and knowledge may be different subject to the characteristics of the arrangements. Hence, ceteris paribus we might observe differences in the degree of decomposition of the division of labour.
significantly affects the competitive dynamics in an industry as the repartitioning and reintegrating property rights and critical knowledge across firms may change the balance of power between suppliers and assemblers.

Independently on whether the “mirroring hypothesis” is valid, the research reviewed in this section provides evidence in support for the rise of a new organisational paradigm in some industries related to the modularisation of products and technologies. The evidence on innovation platforms seems also to indicate that even in the face of increasing modularity in production, some form of hierarchical coordination mechanism continues to exist.

... and the limits to specialisation

More recent contributions have come to view the “mirroring” hypothesis and other aspects of the “modularity theory of the firm” as presented in the previous section with increasing scepticism. Hoetker (2006), for instance, has tested the assumption that increased product modularity is associated with increases in organisational modularity empirically. His analysis shows that product modularity is positively correlated with supplier turnover. Modular products therefore lead to more reconfigurable organisations and hence higher organisational flexibility. Nevertheless, his analysis also shows that product modularity is not associated with the decision to outsource. He concludes that “product modularity contributes less or not at all to [...] firms shifting activity out of hierarchy” (Hoetker 2006, p. 514). Hence, there is no or little evidence that loosely coupled networks of firms supplant integrated hierarchies.

Hoetker’s study suggests that the development of loosely coupled networks of firms and deverticalisation of firms are separate phenomena that can exist one without the other. Several other contributions reach similar conclusions. Hoetker (2005) himself points to the importance of internal supply relations. In environments that are characterised by high uncertainty the value of internal supply relations is highest to downstream firms. Past relationships with other suppliers as well as differences in technical capabilities between internal and external long-term suppliers are of little importance under these conditions. However, environments of high uncertainty are those for which most proponents of the “mirroring” hypothesis postulate the highest advantage of modular production networks. Hence, firms may not be willing to outsource activities to specialised suppliers even if their products are modular.

Another important strand of criticism (cf. Brusoni and Prencipe 2001, Brusoni et al 2001, Brusoni 2005, Hobay et al 2005) highlights problems that derive from the assumption that knowledge and property rights in modular production networks are perfectly partitioned. These authors use case studies of aeronautical engineering and chemical industries to argue that the knowledge boundaries of the firm differ from the boundaries of the firm as defined by outsourcing decisions. The diffusion of modularity as a design strategy leads to an increasing division of labour across firms at the product level, “but only once someone has made it so. And this ‘someone’ is a firm that maintains a ‘higher-level understanding’ necessary to be able to frame problems, and the division of labour

---

13 Benassi (2009) draws similar conclusions.
14 Hoetker (2006), p. 514, discusses an example in case for the notebook computer industry.
around them, i.e. the system integrator” (Brusoni 2005, p. 1900). System integrators are firms that “guarantee the overall consistency of the product and [...] orchestrate the network of companies involved” (Brusoni and Prencipe 2001, p. 185). This however, requires that these firms maintain knowledge that is wider than their productive activities would suggest. They know more than they do. These firms are necessary as modular product architectures in themselves do not provide all the information such that the actions of different actors can be coordinated through the market mechanism. Hence, hierarchical organisation will take over this role and act as visible hand in such modular production networks. In this sense system integrators are different from lead firms (Sturgeon 2002) or firms that carry out combinations of modules (Arora et al 1998).

In line with our discussion on the limits of modularity other authors instead argue that in order to develop new and better performing designs firms will not only rely on modularizing products and organisations. They will change product architectures also by integrating and consolidating tasks (cf. Fixson and Park 2008). This is an iterative process of co-design in which firms will explore which and how activities can be decomposed and what types of interfaces are required to reintegrate the modules into complex artefacts (Sabel and Zeitlin 2004). Such a process can change the nature of competition in an industry as it can drive out firm that are specialised on a few modules and are not capable of competing at the level of a more complex, integrated artefact. From this larger more hierarchical firms emerge. While this does not necessarily contradict the assumption that a modularisation of industry is driven by the search for an efficient repartition and reintegration of property rights and knowledge, it shows that a modular division of labour may be just a transitory state in an industry. Its structure may well fluctuate between stages in which its structure is more modular and stages where it is more integrated.

**Summary and conclusions**

This paper has reviewed the growing body of literature that has studied the implications of modular system architectures for firm strategy and the division of labour across firms. Modularity is a concept used to characterise specific designs and design heuristics for complex systems. Modular systems have been shown to convey superior capability for adaptation and flexibility in uncertain environments. Many scholars have therefore explored the value of modular products or organisations for firms and concluded that firms will benefit from the use of this strategy. However, other research shows that the fixed structure as well as technical standards and interfaces predetermined in the design rules of a modular system restrict the search for solution to well defined areas in the domain of possible designs. It is therefore not possible to develop very novel system designs. Sustained competitive advantage is unlikely to be obtained from modularity when developing products or shaping organisations.

Other research has analysed the impact of the diffusion of modular design heuristics for the development of products and organisations on the division of labour across firms. The principal finding is that product modularity increases the division of labour across firms at the product level. Under well defined conditions it will give rise to modular production networks which are conceived as loosely coupled networks of firms in an
industry. Some authors argue that modular product designs and related industry standards lower different types of transaction costs and create external economies of scale thereby favouring indirect coordination of economic activities through the market. The hypothesis therefore is that modularity in production and organisations will lead to the demise of vertically integrated firms.

This hypothesis is rejected by a number of contributions. Chief reason for the rejection is the argument that there is a divergence between the division of labour between firms and the distribution of knowledge across firms in an industry. Standards and semi-fixed module interfaces do not provide enough information for firms in a market to perfectly coordinate their actions. Some firms must have the capability to frame the problems related to the production of a complex artefact. For this reason they have to rely on broader knowledge bases and carry out coordination tasks. Hence, indirect coordination in modular networks is unlikely to replace direct coordination through hierarchies.

References


