Service Modularity: What Role Do Customers Play in Structuring Interactions at the Interface?

Abstract:

We argue that modularity theory needs to be extended to accommodate services where customers play a role in structuring interactions. We sought to find how a service provision can be dissolved into coarse-grained service modules that minimise interdependencies and make the whole system more comprehensible. We then focus on investigating the role of the customers in structuring interactions at the system interface. We find that, contrary to what has been proposed by literature on product modularity, transactions between participants such as customers and regulators were mainly conducted at thick crossing points. Finally, we propose an indicative proposition for service modularity that can be used to help better develop modular services.

Keywords: services, modularity, crossing points

1. Introduction

In recent years, the notion of modularity has gained attention in scholarly literature, particularly in the fields of production, organisational, and supply chain management (i.e. Baldwin, 2015; Baldwin and Clark, 2000; Schilling, 2000). Recently there is increasing attention to the organisational features of modularity (e.g. Brusoni and Prencipe, 2001; Campagnolo and Camuffo, 2009). A few scholars take steps toward adopting this approach in the context of services (e.g. Bask et al., 2010; Brax and Toivonen, 2007; Pekkarinen and Ulkuniemi, 2008; Tuunanen and Cassab, 2011; Voss and Hsuan, 2009). However, little is known about how service modularity should be understood and better developed. Furthermore, most literature describes modularity in services from the producer’s point of view. Thus, to what extent can modularity theory explain the dynamics of services sector innovation? What role do customers play in shaping innovation in services? The overall objective of this paper is to provide a critique of the empirical literature on service modularity, to apply the theory using the case of mobile payment systems,¹ and to discuss future research opportunities.

The remaining of this paper is organised as follows. Section 2 critically reviews the field of service modularity and the role of the customer. Section 3 unfolds the research methodology being utilised and describes the data and results from the field. This section also provides a comparative analysis by contrasting the empirical case and explaining how the theoretical gap would be bridged by providing insights from the case. Section 4 synthesises the research and formulates the result and analysis. Section 5 concludes by putting the case studies in context and addressing their limitations and contributions towards theory building.

2. Theoretical discussion

¹ This study extends the research agenda by focusing on the concept of modularity in new service development and assessing the empirical evidence from the case of mobile payment systems in developing nations. This study is both timely and relevant. The mobile payment industries are currently in a transition period where hundreds of tried-but-failed mobile payment solutions appeared, along with some future promising-but-uncertain innovative mobile payment solutions being introduced in the market (Dahlberg et al., 2008; Dahlberg et al., 2015; Dermish et al., 2012; Evans and Pirchio, 2015).
Modularity is about managing complexity in technology, production, and organisation (e.g. Baldwin and Clark, 2000; Langlois, 2002; Sanchez and Mahoney, 1996). It is also hinges on the system in what the system theorist Herbert Simon (1969) described as ‘nearly decomposable’ system—that is, a system that is comprised of a set of relationships such that module boundaries can be outlined and shown which elements of communications and coordination are close within the module and which are scattered among modules (Baldwin and Clark, 2000; Bask et al., 2010a; Garud et al., 2008). This would make information hiding possible (Parnas, 1972): decisions can be made in one particular module regardless of what is going on in the other modules.

Due to the rise of modern technology, modularity has shown its distinctive contribution in the literature of technological change and economic institutions (Baldwin and Clark, 2000). Modularity expresses that the division of labour in design has enabled organisations to exploit new opportunities through vertical integration by promoting specialisation in technical manufacturing as well as organisational innovation. Service providers can benefit from modular thinking and strategy (Spring and Araujo, 2009) by efficiently streamlined service processes and organise demand heterogeneity (Pekkarinen and Ulkuniemi, 2008). Indeed, service modularity is still in its infancy (Carlborg and Kindström, 2014) and research into specific area of services still relatively limited (Bask et al., 2010a).

Modularity is a principle that shows how a system can be designed with sub-systems while reducing complex interdependencies via standardised interfaces in order to retain efficiency. Modularity offers greater flexibility through recombination and efficiency through economies of scale from reusability. Modularity, as a core idea in the domain of general systems theory, can be defined as “a very general set of principles for managing complexity. By breaking up a complex system into discrete pieces—which can then communicate with one another only through standardized interfaces within a standardized architecture—one can eliminate what would otherwise be an unmanageable spaghetti tangle of systemic interconnections” (Langlois, 2002, p19).

This concept of modularity can be both an organisational characteristic and a technical characteristic, and has implications for both the performance and organisational structure of firms and industries (see Langlois, 2002). The theory of modularity (Baldwin, 2007; Langlois, 2002; Schilling, 2000) describes how the boundaries of firms and the structure of vertical contracting are positioned in the industry. Modular theory of the firm proposed the notion of crossing points to explain where the task networks are located. The whole task network will be decomposed into more sets of specialised sub-networks in this crossing point with transactions presenting among them.

Baldwin (2007), building on transaction cost theory, proposes the notion of task networks as the systems of production whereby both “thick crossing points” and “thin crossing points” exist. Thin crossing points usually are linked with low transaction costs that require minimum interaction as well as limited information exchange between stakeholders and with ‘information

---

Friedrich Hayek’s (1945) postulate of markets as tools for managing information contributes to how we would define modularity. Every economic actor holds some sort of local knowledge about their own resources utilisation, but only have to consult prices to make economic decisions in acquiring those economic resources. The knowledge that every economic agent utilises can affect prices in the market where no other agent has the information and knowledge of it. Prices in the market are functioning as interfaces that interconnect every participant, whilst hide other irrelevant information at the same time. Contrastingly, a central planner is non-modular and is anticipated to acquire and take action based on this information. Any market information that has been ruled out in the first place does not have to interact with a central planner.
hiding’ among them. This usually happens for arm’s length transactions in which the cost of identifying, calculating, and compensating goods or services are easily transferred. She also posits that modular interfaces appear to emerge at “thin crossing points” inside the task network (Baldwin, 2007). She also explained that “regardless of its intended purpose, modularisation necessarily creates new module boundaries” (p. 179) with minimum regular transaction costs and related thin crossing points.

In contrast, thick crossing points, with many interdependencies to manage, require substantial exchanges of information as well as communication and cooperation through informal (relational) contracts and/or formal contracts. It is not difficult to imagine that complex system such as a mobile payment system comprised of network links that could be characterised as service-provision relationships that are mutually exclusive. Thus, all of the coordination activities start from the firm toward a concept of transactions within bigger institutional formations as bounded relationships in order to create mutual value creation.

Transactions conducted within a bigger institutional framework sometimes can be classified in terms of products that can be viewed as tangible bounded relationships. In this case, discreet goods or products that are standardised parts of a complex system represent modular structures regarded as comparatively “thin”. The transaction costs are low, but the interactions are quite busy, and the networks are quite complex in creating and developing mutual value together (Baldwin, 2007; Langlois, 2002). Spring and Araujo (2009) argue that the differences between thick crossing points and thin crossing points do not distinguish between services and goods, since what they define as “services” can be identified and transferred, and thus categorised and shaped by thin crossing points. Furthermore, technologies, particularly IT, tend to be difficult to differentiate.

Instead of debating between services and goods, Vargo and Lusch (2008) take a different approach by differentiating indirect service provision (i.e. through goods or products) and direct service provision. To a certain degree, it can be asserted that indirect service provision (in this case, a mobile payment system) is identical with thin crossing points with boundaries that are relatively easy to recognise and this leads to increasing efficiency for the service beneficiary and the service provider. Both parties benefit from decreasing transaction costs due to facilitated resource acquisition and the capability to arrive at a market price. However, this indirect activity will also reduce the effectiveness of service provision compared to direct service provision such as consulting or health care services (de Blok et al., 2010).

However, it can be asserted that direct service provision such as medical services can be more effective than indirect service provision (e.g. via medical equipment or third party involvement) due to higher transaction costs—albeit perhaps less efficiently. Yet, there remains several exceptions, particularly as technology becomes cheaper and better able to minimise transaction cost in such dynamic interactions and environments (e.g. health care software or health care systems that can interactively provide information and assistance to the patient daily).

Our observation is that, in the literature on service modularity, argument-building on the design choices underpinning the decomposition of service offerings into modules is scarce. Most literature investigates whether they could recognise modularity in a given practical setting. We identified four decomposition logics: single-level process-oriented, single-level outcome-oriented, multi-level outcome-oriented, and multilevel combined orientation. We found a relationship between the decomposition logic and the modularity types. The aim of the modularisation did not seem to explain the decomposition logic fully; rather, we found that the
decomposition logic applied was related to the service routineness. Thus, it is important to be more explicit and give detailed attention to establishing clear boundaries for the service system being decomposed, the decomposition level(s) on which the functional parts of the service system are specified and how dependencies between modules are minimised.

A successful modular supply requires an intensive and often time-consuming design process (Baldwin and Clark, 2000). Thus, one should first identify why a service offering should be modularised: is it to provide variety, to lower costs or to balance variety and costs with an eye on efficient customisation? Second, the input and throughput uncertainties need to be considered, as the routineness of the services offered may have consequences for the appropriate orientation (outcome, process, combined) and decomposition level. Alongside cost reductions, modular architectures can offer greater transparency to clients on what can be delivered. Providing an overview of modules, and how they can be mixed and matched, could guide the service-specification process (de Blok et al., 2010). How the service modularity concept, including the modularity types specified and decomposition logics identified, contributes to balancing the variety and costs of non-routine service offerings is a relevant topic for our theoretical contribution.

We also realise that there is no unified definition of service modularity. We found that service modularity literature is mostly influenced by manufacturing modularity that follows the traditional systems view and approach. Thus, it does not capture the multi-layered or multi-faceted characteristics of services yet. There is also perhaps too much focus on generalisation and debates on prescription/prediction. Yet what we need, that is currently missing, is to understand what we can do to overcome forces that constrain the alignment of technical, organisational, and services modularity. With regard to practicality, it appears that there are no clear guidelines on how to utilise the concept in practice. Lastly, most studies we found focused on a single case study or single companies without comparative analysis and so they are difficult to generalise into broader populations.  

3. Research context and methodology

Even though mobile payment has attracted a lot of attention and criticism from academics and practitioners (e.g. Au and Kauffman, 2008; Chatain et al., 2008; Dahlberg et al., 2008; Dahlberg et al., 2015; Duncombe and Boateng, 2009; Jenkins, 2008; Ondrus and Pigneur, 2006; Pousttchi et al., 2009), most studies still focus on how customers adopt a particular payment system. Attempting to shed light in the literature gap, we were particularly looking at how modularity principles are applied, how they can be practically implemented, and what important strategic decisions need to be addressed in order to develop and provide modular service offerings. By doing that, we can provide empirical insights into how modular thinking and strategies can enhance more efficient and effective service provision.

The choice of the cases is derived from strategic sampling, considering their social, economic, and technological context (Dermish et al., 2012; Pousttchi et al., 2009). The social and economic conditions of each particular country will also help to grasp the development context and contemporary state of the enterprise surrounding mobile payment systems. It also appears that the technological context in which these projects take place determines the success or failure of project deployment (Duncombe and Boateng, 2009; Evans and Pirchio, 2015). The

rationale behind the case selection was not to choose representative cases of a given category but to choose cases which display a high degree of the phenomenon under study (Pettigrew, 1990) and are polar in the sense of covering the ‘known range and variation’ (Hakim, 2000).

We draw on both quantitative and qualitative data gathered during two-years of desk research and field study. Those data were obtained from primary as well as from secondary sources. Multiple data sources allowed us to triangulate several different sources of evidence (quantitative and qualitative) as well as various kinds of collection methods (such as document analysis and interviews) within the case and generate more solid and robust substantiation of constructs and underpinning arguments for its contributions to knowledge. We adopted a comparative case study method that is often utilised in the field of management of technology, innovation, and strategic management, and applied the similar framework to examine several cases (Eisenhardt, 1989a; Leonard-Barton, 1990).

The main sources of our qualitative data were 19 in-depth interviews conducted with key actors and stakeholders in the mobile payment industry. Interviewees included technical directors, company developers, project managers, industry experts, as well as technical experts from regulatory bodies. Interviews were conducted in a semi-structured way and lasted 175 minutes on average. Qualitative insights presented by our interviewees demonstrated to be significant and important for apprehending the dynamics of modularity in service development. This phase has also allowed us to recognise and analyse patterns or trends as they evolved from within cases being studied (Eisenhardt, 1989a).

Secondary sources of qualitative data gathered from about 540 documents, consists of company annual reports, industry reports, trade journals, technical journals, newspapers and magazines articles, as well as websites articles. We also analysed 126 patents on mobile payment systems worldwide. Primary sources of quantitative data came from a sector specific database of the GSM Association (GSMA) and the International Telecommunication Union (ITU). We also extracted data from WorldBank Global Findex to complement the study. We then utilised the pattern-matching strategy advised by Yin (1994, 2003) by comparing an empirically-based pattern with an alternative or estimated one. The use of that aforementioned strategy presumes the establishment of rival or contested explanations and justifications that include independent variables that are mutually exclusive.

We use an explorative case study of Oi Paggo (Brazil), TCASH (Indonesia), and M-PESA (Kenya) where service modularity to be found, in order to develop propositions about how such practices are being managed and organised. The cases were selected taking three specific criteria into account: (1) adoption level, (2) geographic location, and (3) data availability. As the most profound case on mobile payment adoption, M-PESA will serve as a benchmark to contrast with the other two cases. M-PESA also represents the dynamics of a country with a high adoption rate. TCASH represents a country with a moderate adoption rate, while Oi Paggo represents a country with a low adoption rate in mobile payment services. Indeed, it is also expected that these cases will provide adequate pictures of emerging economies from different continents.

4 When a particular finding can only be gathered from interviews, we need at least two people having the same conclusion to corroborate this finding. We also value interviews from a later time rather than interviews of the same people at earlier time to take into account that opinions might change due to new insights or experience during the development phase.
Modularity in services has been viewed as a pivotal factor in developing service orientation within the firm (Bask et al., 2011; Böhmann et al., 2003; Voss and Hsuan, 2009). Pertaining to this view, services should be developed as modular units and assembled by mixing and matching those units accordingly to customer demand (Bask et al., 2011; Rahikka et al., 2011; Tuunanen and Cassab, 2011; Ulrich, 1994; Voss and Hsuan, 2009). On one hand, modular services must have standardised baseline services, customised services, as well as combinations between the two. On the other hand, modular services must reuse processes possible during the implementation in order to accommodate flexibility and customisation (Bask et al., 2010a). These attributes are evident in our cases of mobile payment systems.

Tuunanen and Cassab (2011) state that service process modularisation allows firms to achieve market impact efficiency from the extension of services via reusability and variation of existing processes. However, taking into account what has been shown from the case studies, servitisation also poses challenges, not only internally (such as design strategy, organisational transformation, etc.) but also externally (such as changes in customer needs and preferences, changes in regulation and power structure, etc.). Thus, service providers must deliver stellar service packages to the customer in the front end, but at the same time, they have to operate efficiently in the back end while managing changes and complexity in their environment.

In the previous section, we discussed the implementation of modular principles in service development. We argue that modular service systems represent some sort of functionality that can be implemented using components (either software or hardware) to any degree of variations as long as the implementation is strictly compatible with the conceptual model (through a design rule). The modular principle provides a guiding design template that includes the structural dimension as well as dynamic aspects of abstract components that can be implemented using components determined by the system designers. However, echoing Simon (1962), we did not find a system that was totally built up of independent modules—there are always some unrecognised inter-modular interdependencies.

The cases confirm our proposition in the way that the application of service modularity was affected by several key service attributes, such as: customers engaged in the production of services (co-producer), both service process and service product are combined in the final service offerings, and that service packages will incorporate both technical as well as human factors. Yet, even though modularity theory in general has been around for the last two of decades, it still lacks a unified definition.

It is the case that modularity practices in services was influenced by modularity in product manufacturing. It appears that the current literature mostly approaches modularity from a traditional systems view that tends to be static in insight and oversimplifies the complexity of the ‘modular world.’ However, due to the distinctive attributes of services, modularity in services setting is usually more heterogeneous and complicated than in products. Business services can exist in different dynamic relationships (i.e. B2B, B2C, B2G, etc.), that might involve low or high human involvement, can be knowledge-intensive or rely on no knowledge at all, and can also be physically or virtually based. Thus, modularity in services must be viewed as multi-layer or multi-level. Evidently, there is a need for a re-conceptualisation of service modularity that is not purely simplistic, but also incorporates heterogeneity and multi-layer/multi-faceted characteristics of services.
We propose that a modular service development should really start with the identification of service requirements, in line with Edvardsson and Olsson (1996). A modularity approach will provide better adaptability and greater flexibility to a firm in producing service provisions. However, a systematic modularisation mechanism must be followed to facilitate the process of segregation and decomposition. As suggested by Silvestro and Silvestro (2003), service systems, service processes, and service concepts must go together in harmony to create and provide optimal value to the customer.

By decomposing services into a series of service modules, then service process activities that consist of identical service content can be mapped and grouped according to their similarities and reusability. The focal point of modularity in services is that it is built upon the concept of independency in function in which every service function ought to be independent of other functions (Geum et al., 2012; Schilling, 2000). Thus, service modules should be defined in a mechanism where interactions between modules are minimised but might be high within a module (Ulrich, 1995).

In decomposing services, it is important to distinguish functional elements as well as physical elements. Physical elements relate to the physical systems which are installed to hold up the services or to support functional location in which those service activities carry out. Functional elements are single service transformations and operations that give rise to the entire functioning of services system. Thus, to accurately plot the activities to the service modules, mapping those activities becomes crucial. It is also particularly important to separate common elements and distinctive elements within a service module.

‘Traditional’ modularity theory emphasises technological discontinuities and architectural changes (Langlois, 2002). It is thus assumed that, implicitly, technological change occurs inside the boundaries of certain modules and does not have an impact upon the interdependencies between the modules and their accompanying inter-organisational relationships. Our cases of mobile payment systems suggest that this is probably not the case. We found that when the technological dynamic is changed, firms need to coordinate with their suppliers, vendors, and other related parties in order to share information according to the situation and thus adapt and change their behaviours and expectations reciprocally.

Mobile payment systems are currently being developed in such modular ways, but we found that not every modular service system is the same. We did not find any case that is fully modular or fully integral. They are somewhere in between or close to one end of the spectrum. We also can show that modularisation in the sector follows different trajectories and patterns. The changing path can be seen in the mobile payment system, perhaps due to its technology-intensive nature. However, patterns in other industries might be different.

We contend that the benefit of modularity does not come without cost: pre-existing interdependencies can stimulate unanticipated problems, conditions might change and require specialised modularisation. Thus, the decision to apply the modular principle ought to be deliberate. Rather than questioning whether we should become modular or not, a more fruitful pursuit would perhaps be to ask how to maximise the benefit of modularisation.

Research on modularity has also provided a significant addition to the literature of economic organisations and technological change. This shows that, especially during the design stage, technical modularity and advancement in the division of labour has opened new and alternative
pathways for the organisations other than vertical integration, through promoting vertical specialisation in production manufacturing as well as in innovation (Brusoni, 2005).

What seems to be counter-intuitive is that some revisionist studies argue that the interest in modularity has gone too far. Rather than investigating difficulties and challenges that service firms are facing in developing and operationalising modularity, there is a propensity to generalise empirical findings and examinations that are too narrow and context-specific. We are not advocating an alternative option or explanation, more modestly, we were working toward shifting the arguments away from frivolous debates to an empirical discussion that enquires about what factors could limit the alignment of technical modularity, organisational modularity, as well as market modularity. Another related aim is to examine what firms can do to solve these issues. We explore new evidence from the mobile payment sector—which is a large scale, innovative and cutting-edge industry—as a benchmark of wider services industry patterns and trends. Also, the thesis suggests that, even in these industries, there are strong counter-forces pushing the structure of the organisation to transform into more integrated, rather than arm’s length, structures. Our cases on mobile payment systems show how cognitive complexity and competitive dynamics cause limits to modularity. In the industry where technological change is very high and unpredictable, codification cannot reduce complexity. We thus argue that modular collaboration needs more and better coordination through hierarchical arrangement and corporate management.

5. What can we learn from service modularity?

Modularity theory proves to be useful in explaining the dynamics of mobile payment ecosystems that were not captured using other theories. It describes the boundary of the systems, interdependencies among the modules, as well as power relations among different economic actors. Modular thinking also provides an explanation of why some particular aspects of property have been more compliant than others. Modular strategy benefits from both specialisation and economies of scale. When viable interfaces are attainable cheaply, complementary components will be created by the competitive force that is released in a value chain.

Modular principles enables entry by innovators to provide particular inputs that the innovators can manifest comparative advantage, even in a situation where such companies have no competence or little experience as integrated system providers of a bigger part of industry outputs. This, however, does not reduce the capacity and function of organisations in developing interfaces as well as communicating and managing production processes. A balance between the interests of the system developer and those of external partners becomes possible.

We anticipate that the application of service modularity will be influenced by three key attributes that differentiate services from products. First, in service production, customers are engaged in the process and become co-producers along with a service provider. Second, both the service process and service product are inseparable elements in the final service offering. Finally, every modular service package will incorporate not only technological factors but also human factors. Modularity within the service domain has been closely associated with productisation of services and our consideration of modularity in services has been influenced by literature on modularity in production and manufacturing. We have shown, however, both how earlier theoretical treatments can be made congenial to services and in our case of mobile payment systems we have shown how the application works.
One of the most important concerns in modularity studies is that there is no current ‘best’ or ‘ideal’ modular form or ‘best’ or ‘ideal’ architecture with which we could contrast and set side by side the results and balance between strategic needs and functional needs. Since every company (or service system) has a distinctive strategy, the architecture will not become similar or identical. Design Structure Matrix (DSM) emerged from an engineering paradigm, but as mentioned by several researchers, it is difficult to describe a single ‘correct’ decomposition method of a product or service. The series of related components might eventually become completely distinctive, resulting in different clusters of module and matrices in the end. Mapping the example of mobile payment systems also revealed a gap in the literature. It lacked the means to represent how the environment in which a service is deployed affects the value of its design.

These benefits do not, however, come without cost. Service modularisation might overlook interdependencies that exist and cause unexpected problems. However, it might rule out interdependencies in some values. Secondly, the dynamic nature of the environment might change and, thus, require different kinds of modularisation. Even though a modular system can evolve more easily and faster than a non-modular system, a modular system might end up with ‘non-ideal’ solutions, subject to the variation of the modules and subject to whether the variation of configuration is randomised or selected rationally.

In the case of service organisations, the attributes of services are easier to overlook. Modular organisations can be viewed as a panacea for overcoming complexity and attaining flexibility, specifically in addressing coordination problems of associating outputs to inputs. However, as long as the modular service firm can overcome the coordination challenge and the input is valuable enough, our experience with modularity theory and human artefacts would prompt us to anticipate a major role for modular configurations.

In the case of mobile payment systems, contrary to what has been proposed by literature on product modularity, we find evidence that transactions between participants such as customers and regulators were mainly conducted at thick crossing points with many transfers and participants involved. It implies that the boundary of what changes hands in providing mobile payment services are not easy to identify. Thus, in order to achieve efficiency, service providers must acquire competencies so that they can arrive at a market price.

Modularity can be used to describe and explain structures and relationships in the services sector. It also uncovers the dialectical process between innovator and regulator. On one hand, innovators aim to dominate the market and launch innovation more quickly in order to achieve critical mass. On the other hand, regulation stabilises systems in a way that promotes innovation. An additional element of this contribution to the services literature is that modularity seems to work well in knowledge-intensive industries that enjoy network effects.

We also propose a redefinition of modularity in services. The current literature does not include the peculiar characteristics of services: (1) the customer is engaged during the service production stages and, along with the service provider, becomes a co-producer of the service, (2) the final service delivery is an inseparable mixture of service process and service product elements, and (3) both human factors and technical factors will shape and incorporate the modular service package. We believe that this study shows the potential for new research to further advance the debate around modular architectures and service production.\(^\text{5}\)

\(^\text{5}\) Our study has several limitations. First, we characterise innovation in services within the context of modularity only. Analysing innovation in terms of not only modularity but also architecture (Henderson and Clark, 1990),

---

9
References


will perhaps provide interesting and complementary findings. Secondly, we did not consider the possible impact of our subject area allowing for more complex behaviour than that derived from modular recombinations alone. Third, we relied on three in-depth case studies that can be regarded as limiting generalisability.


