Chapter 4

Structuration Theory and Actor-Network Theory as Conceptual Frameworks for Analysis

Table of contents

4.1 Introduction

In the previous chapter I discussed various functional and variance-based theories used in researching online electronic services and highlighted some of their conceptual flaws. Furthermore, I emphasised the use of process and context-based theories as substantive social theories to understand the implementation of an Internet-based self-service technology. The previous chapter also alluded to my belief that Internet-based self-service technologies should be viewed more broadly to include the understanding of social and behavioural dimensions. Therefore, neither technological determinism nor strong social constructivism may be suitable to explain the process of online self-service technology implementation (Orlikowski and Barley, 2001). A model located somewhere between the poles of technical determinism and social constructivism is deemed to be appropriate.

The purpose of this chapter is to discuss structuration theory and actor-network theory, two social theories which together form the basis of the initial conceptual frameworks created to analyse the case study. A conceptual framework can be defined as the structure, the scaffolding, or the frame of a study (Merriam, 1998). Some researchers refer to it as the lens through which we view the world (Walsham, 1995; Orlikowski and Baroudi, 1991) or the territory to be explored (Carroll and Swatman, 2000). I envisage that using structuration theory and actor-network theory to form these conceptual frameworks will account for the gaps in knowledge ascertained in the literature review and also address the research questions expressed in chapter 1. In addition, it will guide me through the analysis and the way I eventually interpret the findings, in future chapters.

Walsham (1995:76) cautions that:

> *… although a theory can provide a valuable initial guide…, there is a danger of the researcher only seeing what the theory suggests, and thus using the theory in a rigid way that stifles potential new ways and avenues of exploration .*

Following the advice of Walsham (1995), and Carroll and Swatman (2000), I will maintain an open attitude and willingness to change my initial assumptions and theories captured in this

framework. However, in designing this conceptual framework I will remain faithful to my philosophical orientation and stance discussed in chapter 2.

In the next section I draw upon the key concepts, ideas, terms, definitions, and models espoused by these theories. I start by describing structuration theory, a more dominant theory used for understanding the social context of information systems. Although analysis with this theory has been more common in information systems and has provided interesting insights in the past, I argue that a different understanding can be gained by using the actor-network theory (ANT). I will demonstrate this in chapter 7 by first analysing the case study using structuration theory, showing the insights such an analysis gives, and then pointing out some interesting answers it provides to the research questions. I will then analyse the same case study using ANT in chapter 8 and compare the manners in which ANT and structuration address the research questions.

4.2 Structuration theory

4.2.1 Introduction

The research questions in section 1.4 demonstrate the idea that SST implementation is implicated in broader social context and processes, even more so compared with conventional organisational information systems. This was confirmed in the literature review chapter. Since structuration theory is used to conceptualise the linkage between context and process in society, it is a good candidate for analysing these types of questions. A number of studies have used structuration theory to carefully analyse the manner in which context is involved in the production of action. This body of research has been broadly associated with content-process-context schema research (Walsham, 1993). I therefore envisage that structuration theory can play a substantial role in helping us understand the societal, organisational and personal contexts within which the self-service technology is embedded. Structuration theory suggests that this linkage is crucial for understanding Internet-based self-service technology, which is enabled or constrained by the social context in which it is implemented, and, in turn, is a medium for maintaining or altering that context. Since this is a logical starting point for analysis, I will discuss this theory further. In the next section, I discuss the main elements of structuration theory.

4.2.2 Key elements of structuration theory

Structure and agency

Structuration theory was developed by British sociologist Anthony Giddens (1984). Giddens' main aim for his theory was to integrate two opposing strands of social thinking: structuralism and functionalism, emphasising notions of social structure; and interpretivism, bestowing prominence to human agency and meaning. However, Giddens is concerned neither with the experience of the social actor nor with the existence of any form of social totality. Instead, in structuration theory, his main emphasis is on understanding how social practices are ordered across time and space. In structuration theory, Giddens (1984) attempts to recast structure and human agency as the duality of structure and action. In other words, social structure is drawn upon by agents in their day-to-day actions, and is therefore produced and reproduced by this action. At the same time, action is both constrained and enabled by structure. Accordingly, action is only possible then because of structure, and structure itself can only be constituted through action. At this point it is important to consider carefully what Giddens means by agency, and therefore how it operates in a structure.

The agent and agency

One of Giddens' main tenets in structuration theory is that human social activities are recursive. These activities are not brought into being by social actors. Instead, by expressing themselves as actors, activities are continually recreated by social actors. It is in and through activities that agents reproduce the conditions that make these activities possible. However, unlike knowledgeability displayed in the natural environment in the form of coded programmes, human agents exhibit cognitive skills. Human knowledgeability is therefore essential to human action. According to Giddens (1984), it is especially the reflexive form of knowledgeability that is involved in the recursive ordering of social practices. Giddens (1984) defines reflexivity as the monitored character of the ongoing flow of social life. In other words, human beings are purposive agents, having reasons for their activities and being able to elaborate discursively upon these reasons. Giddens (1984) is quick to caution that his use of the terms such as 'purpose', 'reason' and 'intent' should not be disentangled from the context. Instead, human action should

be seen to occur as a *durée,* that is, as a continuous flow of conduct, as does human cognition. As such, purposive action is not made up of a series of separate intentions, reasons or motives.

Reflexivity then should be articulated as the monitored character of the ongoing flow of social life, and not as an aggregate or series of separate intentions, reasons and motives. Giddens (1984) argues that this reflexive monitoring of action depends on rationalisation. Rationalisation is understood as a process inherently associated with the competence of the agent, and not as a state. Furthermore, action cannot be divorced from history. Neither should action be viewed as a combination of acts. Action also has to be observed in conjunction with the agent's surrounding world and what Giddens terms the 'acting self'. So, reflexive monitoring, rationalisation and motivation of action are treated as embedded sets of processes. These dimensions are viewed as routine in the process of human conduct. In fact, an actor is evaluated by other actors according to the level of competence exhibited in the rationalisation of his or her action.

However, not all processes are accessible to the discursive consciousness of the social actor. Giddens (1984) explains that the 'stocks of knowledge' or 'mutual knowledge' incorporated in encounters are not directly accessible to the consciousness of actors. Most knowledge in social encounters is of a practical nature, accessed within an agent's practical consciousness that enables social actors to 'go on' within the routines of life. Nevertheless, he does clarify that the lines between discursive and practical consciousness are fluctuating and permeable, depending on the experience of the individual agent and the evaluation between social actors engaging in different social contexts. Discursive and practical consciousnesses are viewed separately from the unconscious motives of the agent (more on these distinctions later). He distinguishes between discursive consciousness – where reflexive monitoring, the ability to explicitly describe actions and motivations, takes place; practical consciousness – the ability to act in a knowledgeable way and where there is rationalisation of action; and lastly unconscious motives. For now, it is sufficient to know that the unconscious does not assist us in advancing our knowledge of how social actors are able to control and sustain their conduct.
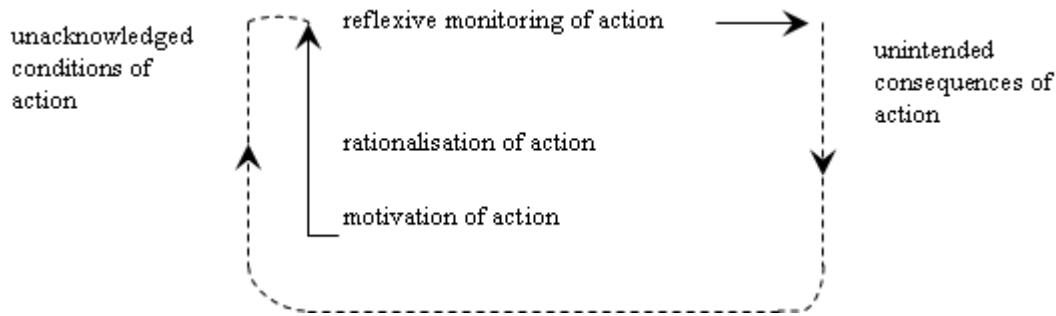
Figure 4.1 The stratification model of the agent

Source: Giddens (1984:5)

The stratification model of the agent in figure 4.1 illustrates how individual actors and other actors continuously monitor the flow of their activities. Giddens distinguishes between reflexive monitoring and rationalisation of action from its motivation. In doing so, Giddens contrasts 'reasons', as the grounds for action, with 'motives', which refer to the wants which prompt the action. Motivation is defined as a potential for action. Whereas reflexive monitoring and rationalisation are directly bound up in routine or continuity, motivation occurs in relatively unusual circumstances and serves in some way to break with routine. Motives tend to supply overall plans, projects or programmes. However, much of a social actor's day-to-day conduct is not directly motivated. Giddens (1984) points out the contrast between the ability of competent actors to report reasons for their actions and the motives for their actions. This difficulty may have something to do with the fact that motives are linked to the unconscious. Nevertheless, while unconscious motives are an important feature of human behaviour, practical consciousness takes centre stage in structuration theory. In explaining practical consciousness, Giddens emphasises the role of the agent in day-to-day life. Day-to-day life occurs as a flow of intentional action. However, acts have unintended consequences, and these unintended consequences may feed back to the unacknowledged conditions of further acts. Giddens also distinguishes between intentional and unintentional doings. An intentional act is characterised as one in which the instigator knows the outcome or quality of an act and how to go about achieving this quality or outcome. Unintentional doings refer to those doings which would not have happened if the individual had behaved differently, but are not within the scope of the agent's power to have brought about.

The next concept we review is agency. Giddens does not view agency as the intentions people have in doing things, but as their capability of doing things. As such, for Giddens, agency implies power. Agency concerns events in which an individual is the instigator, in the sense that the individual could have chosen at any point in his or her conduct to have acted differently. Agency therefore refers to doing. For Giddens (1984), human agency is the 'capacity to make a difference', also known as 'transformative capacity' (1984:14). Agency is intimately connected with power. In fact one of the defining characteristics of agency is power, since the loss of the capacity to make a difference is also powerlessness. As Giddens (1984:283) himself explains, 'there is no more elemental concept than that of power'. In practice, human agents almost always retain some transformational capacity, albeit small. Power plays a central role in the exploitation of resources. Thus, power is inherent in social action, as it relates to domination involving resource authorisation and resource allocation. Authoritative resources derive from the coordination of the activity of human agents, while allocative resources stem from control of material products or aspects of the natural world.

For Giddens, action addresses not only individual moments of doing, but rather the *durée* of human life, the 'continuous flow of conduct'. Giddens is particularly interested in the problem of order (Jones, 1999). The continuity of social life, that is, the recursive ordering of social practices, is of particular importance. According to Giddens (1984), recursive ordering is only possible because of the continuity of practices that makes them distinctly similar across space and time. Thus recursive ordering presumes reflexivity. Reflexive monitoring of action is grounded in some sort of rationalisation, itself based on a motive, an overall strategy, plan or programme for the range of conduct.

It follows then that both discursive and practical consciousness are 'directly bound up with the continuity of action' (Giddens,1984:6). Individuals, in particular, acquire ontological security through their engagement in predictable routines and encounters. Although Giddens (1984) claims that routine is the predominant form of social activity, this does not imply that action is programmed. On the contrary, routinisation in most social conduct has to be continually 'worked out' by those who sustain it in their day-to-day conduct. Thus local practice is always indeterminate. Interactions are always shaped by actors as they interpret their local context, even when such interactions are routine.

Routines are important because most daily practices are not directly motivated. In addition, the knowledgeability of human actors is to some degree bounded by the unconscious on the one hand, and by unintended consequences of their actions on the other. However, this does not imply that actors do not skilfully engage in their daily practices. At least at the practical consciousness level, it is assumed that all human beings are knowledgeable agents, understanding what it is that they do, in a manner that is not incidental to their activities. In other words, people can monitor and reflect on their own practice and that of others. They can reflect on the effect of planned and unplanned effects of their intentional actions. Subsequent actions are based on reflexivity, suggesting that these actions are not mere repetitions of what was done before. Therefore, this observation that agents engage in routine should not be construed as predetermined or even simplified behaviour. Instead, Giddens (1984) maintains that knowledgeability embedded in practical consciousness exhibits an extraordinary complexity. The essence of the duality of structure, then, is the complexity of human action within the ongoing flow of life stretching over lifetimes, and institutions giving continuity or 'fixity' to social practice. Action, therefore, is crucial to the stability and overall existence of social life. However, the ability of human beings to monitor their own conduct in a reflexive way and the unintended consequences of human conduct imply that all action carries within it the seeds of change (Walsham, 2001).

Structure and structuration

Giddens (1984) offers a unique conceptualisation of structure. According to Giddens (1984:28), structure is better understood as 'normative elements and codes of signification' that exist as memory traces in the mind of the agent. Therefore structure is not to be seen as a thing, an object or some kind of skeleton or morphology (Orlikowski, 1992). Instead, structure is defined as rules and resources recursively implicated in social reproduction. As such, institutionalised features of social systems have structural properties only in the sense that relationships are stabilised across time and space (Giddens 1984:xxxi). In contrast to conventional views of sociology, then, that view structure as being 'external' to action, in structuration theory structure is viewed as 'a source of constraint on the free initiative of the independently constituted subject' (1984:16). Following this line of thought, social practices are not composed of 'structures', but rather exhibit 'structural properties' (1984:17). This structuring property allows for the 'binding' of

time and space in social systems, which make it possible for discernibly similar social practices to exist across varying spans of time and space and which lends them a 'systemic' form.

Structural properties that manifest themselves in social practices (action) are often deeply embedded in these recursively reproduced practices. As such, they act as rules and resources for action, constraining, but also enabling the form of action that can take place. Rules and resources, that is, 'structuring properties', like the idea of structure itself, should not be viewed as 'things' or 'objects', because such perspectives perpetuate a mechanistic view and deterministic outcomes. It is therefore useful to consider rules and structure, briefly.

What sets structuration apart from other social theories is its dismissal of the traditional notions of rules as formalised prescriptions external to the human agent. Instead, from a structuration perspective, rules do not occur in isolation, but in sustained practice in the reproduced relations between individuals that we recognise as social systems. The human agent's capacity to deal with an unknown variety of circumstances is enabled and constrained by rules implicit in the agent's mastery of social practice. As such, rules can serve to constitute meaning in a context and to sanction the conduct of the human agent. Apart from engaging these rules to understand the circumstance or to determine how to act, the agent's ability to act depends on resources. This ability to act can be allocative or authoritative. As pointed out earlier, 'allocative' refers to the human agent's command over objects, goods or material phenomena, while 'authoritative' refers to the human agent's command over persons or actors (Giddens,1984:33). Rules and resources, whether or not we can identify for them a 'real existence' (procedure manuals, laws, raw materials, people), become rules and resources 'only when incorporated within processes of structuration' (Giddens, 1984). In simple terms, 'the rules and resources constituting structure exist only in the agent's heads' (Jones, 1999:99). There are numerous structures that agents enact when engaging in a social context. This implies that in any structurational analysis, one must foreground some structures and background others (Giddens,1984.)

The duality of structure

Giddens (1984) argues that structure consists of the rules and resources that are created through the actions of individuals through practices and routines. A duality emerges as structure constrains action, but, simultaneously, action serves to maintain and modify structure. According to Giddens (1984), as opposed to viewing structure and agency as two independent sets of phenomena (dualism), these two concepts are better portrayed as a 'duality'. As a result, structure and agency are conveyed as being dependent upon each other and recursively related. In other words, human action is enabled and constrained by structure, but structure is also the result of human action. Thus, the focal point of duality in structuration theory is the manner in which action and structure presuppose each other. Thus, the structural properties of social systems are both medium and outcome of the practices they recursively organise (Giddens 1984). The dimensions of the duality of structure are given in the following well-known diagram in figure 4.2. The separation of this duality into vertical dimensions is simply for analytical convenience. As the discussion in the next section clarifies, the dimensions of structure are inextricably interwoven in and with the production of action.
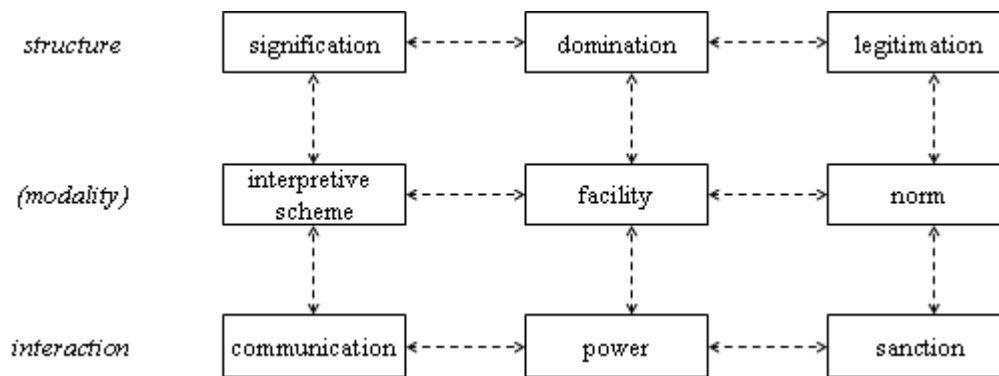
Figure 4.2 The dimensions of the duality of structure
Source: Giddens (1984:29)

Giddens (1984) draws upon the earlier work of Durkheim, Marx and Weber (see Giddens, 1984) to break down social structure and human interaction into three dimensions: signification, domination, and legitimation are interlinked by the three modalities, as illustrated. 'Signification' refers to the rules that constitute meaning, while 'legitimation' refers to the norms and the resources that determine relations of domination. Given that these structural properties are

evident in action, the latter are also analytically deconstructed into communication, power, and sanction. Giddens (1984) also introduces modalities of structuration to further clarify these analytical dimensions. From this, we can 'relate the knowledgeable capacities of agents to structural features' (Giddens,1984:*28*). For this reason, the three modalities relating to signification, domination and legitimation are interpretive schemes, norms and facilities, respectively.

According to Giddens (1984), interpretative schemes are the stocks of knowledge that enable actors to understand things as things, whether they are physical (eg table, chairs), or more abstract and conceptual (eg events, words). This implicit (background) knowledge and explicit (foreground) knowledge are acquired through experience. Actors map their experience of the world into cognitive schemes. These are drawn upon in making sense of their own and others' actions (Walsham, 1993). Communicative action in particular is sustained by, and sustaining of, structures of signification through interpretive schemes. While interpretive schemes are the rules for understanding what to know, norms are understood as the rules for understanding how to act. Actors use rules such as standards of morality to sanction or legitimate their actions as appropriate conduct. Therefore norms constitute structures of signification. Facilities are the material and non-material resources which actors bring to bear on their actions, enabling them to exert power over social action. Material resources can be allocated by those who control them. Non-material resources include 'status, special skills, charisma etc, that an actor may bring to an action situation' (Lyytinen and Ngwenyama, 1992:23). These facilities enable actors to draw on and reproduce structures of domination, or the asymmetry of allocative and authoritative resources. Such structures are fluid rather than concrete, because they demonstrate what Giddens (1984) terms a 'dialectic of control'. This means that even those who are subordinate can influence the activities of their superiors. The playing out of this dialectic can lead to asymmetries that can be dramatically or imperceptibly shifted over time.

Giddens (1984) elaborates on two further concepts, namely social integration and system integration. According to Giddens, social integration refers to the reciprocity, that is, autonomy and dependency, between actors in contexts of co-presence. On the other hand, system integration refers to reciprocity between actors or collectivities across extended time-space. Giddens distinguishes between the cohesive effects of social interactions which take place when actors are physically present and wider systemic effects of interactions across distance. Whereas

social integration preserves a concern for praxis in situ, system integration refers to reciprocities between absent agents, that is, agents who are physically and/or temporally situated in different settings (Cohen, 1990). Social integration is particularly useful for social practice involving self-service technologies where wider-scale practices as opposed to when face-to-face interactions are observed.

Another important concept in structuration theory is time-space distanciation. This involves the 'stretching of social systems across time-space, on the basis of mechanisms of social and system integration' (Giddens,1984:377). As the recursive and reflexive structuration of social interaction extends between people over geographical distance and over time, so does the embeddedness or 'bite' of those practices increase. For example, the user, equipped with access to Internet-based self-service technology, may expect to use the Internet successfully in most parts of the developed world, given that these practices have been more widely accepted there for some time. However, should the user wish to conduct an online transaction in a developing country context, these practices are less widely observed, but may, in the future, become standard. If social practice becomes reasonably stable over time and space, then routines – practices in which actors habitually engage – develop. Routines constitute 'the habitual, taken-for-granted character of the vast bulk of the activities of day-to-day social life' (1984:376).

The use of Internet-based self-service technology, once a social practice to be painfully acquired, may, with the years, become commonplace, a routine part of a person's life. 'All social interaction is situated interaction – situated in time and space. It can be understood as the fitful yet routinised occurrence of encounters, fading away in time and space, yet constantly reconstituted within different areas of time-space. The regular or routine features of encounters, in time as well as space, represent institutionalised features of social systems' (1984:86).

4.2.3 Structuration theory in information systems research

Structuration theory is increasingly becoming a well-received theory in the IS literature (Walsham, 2002, 1993; Rose and Scheepers, 2001; Orlikowski, 2000, 1996, 1992; Walsham, 1993; Barley, 1986). For example, there have been a number of attempts to incorporate information systems within the theoretical framework of structuration theory (eg DeSanctis and Poole 1994; Orlikowski 1992). Following Giddens (1985), Orlikowski and Robey (1991) explore the 'duality of technology' and argue that an individual's actions are neither determined by technology, nor are they capable of constructing technology as they see fit. There is a duality of structure, so that technology constrains and enables individual action, while being a product of individual action. In other words, while we develop technology, technology affects our activities. This recursive relationship recognises that while individuals design technologies to enable new actions, these technologies constrain our action. This is often summarised as technologies constrain/enable action.

In Orlikowski and Robey (1991), the tenets of structuration theory are applied to help understand the relationship between IT and organisations. In their work, the 'duality' of technology is explored. IT is seen as the social product of subjective human action within specific structural and cultural contexts, and simultaneously an objective set of rules and resources involved in mediating human action, hence contributing to the creation, recreation and transformation of those contexts. The concept of the duality of technology is explored further in Orlikowski (1992). Orlikowski's approach to rethinking information technology is well in keeping with the tradition of Giddens' work. DeSanctis and Poole (1994) proposed the adaptive structuration theory (AST) framework to provide insight into group decision support systems (GDSS). However, the AST approach comes in for sustained attack from Jones (1999), who points out that Giddens' concept of structure is incompatible with the more traditional view adopted in AST, and that no substantive theoretical justification is offered, to produce a contingency-type model of technology 'impacts' which Giddens has specifically criticised.

As 'meta-theory' from a social contructivist stable, structuration theory does not provide 'middle range theory about specific phenomena that can be explored or tested directly and empirically' (Orlikowski and Robey, 1991). Neither is it 'specific about the technology' (Monteiro and

Hanseth, 1996). This research style is largely incompatible with Giddens' own, which is bound to lead to tensions. The inherent weakness of some of this theorising is that it tends to reinforce the equation of technology with structure and structural constraint. In IS this tends to take the form that technology is built by human agency; thereafter it constrains what we do – characterised as the 'discontinuous separation of design and use' by Orlikowski (1992). This equation of technology with structural constraint is not consistent with structuration theory. However, the deployment of further structuration theory concepts, such as time space distanciation, routinisation, and system integration helps to explain IS practice while avoiding this problem (Rose 1999).

A study by Barley (1986, above) described the introduction of computer tomography scanners into American hospitals, exploring how the actions of the stakeholders and the institutionalised traditions within the organisation influenced each other as 'occasions for structuring'. By far the most common starting point is the 'dimensions of the duality of structure' model (figure 4.2), using Giddens' concepts as a checklist for guiding social analysis. A fairly straightforward use of these concepts occurs in Karsten (1995), where Lotus Notes implementations in three organisations are analysed. Jones and Nandhakumar (1993) go further in their analysis of the development of an executive information system by reflecting upon the theory – thus completing the circle. Walsham (1993) provides sustained longitudinal case study analysis covering issues of IS strategy, development, implementation and evaluation in three contrasting organisations. Walsham and Sahay (1996) used structuration theory with actor-network theory to investigate problems in developing geographical information systems (GIS) in an Indian government department. They analyse the social context and process of implementing GIS in India, and the inter-linkages between them. They emphasise two aspects of social context relating to government organisational structures and the scientific tradition, and relate these to the initiation, operationalisation and continuation phases of the GIS implementation process. In addition, Lyytinen and Ngwenyama (1992) found that social activity, including work processes, is enabled and constrained by social structures that are produced and reproduced through human action.

More recently, Orlikwoski (2000) proposed an extension to the structurational perspective by developing a practice lens which examines emergence, improvisation, and change over time as people reconfigure their technologies or alter their habits of use, and thereby enact different technologies-in-practice. Despite the growing popularity of Internet-based SSTs in contemporary

social settings, there is little evidence of any reference to the structuration approach to understand this contemporary phenomenon. The next section adapts Orlikwoski's (2000) latest notable advancement in applying structurational theory in IS into a conceptual framework for analysis.

4.2.4 A structurational framework for analysis

Orlikwoski (2000) proposes an extension to the structurational perspective by developing a practice lens which focuses on how users' recurrent interaction with technologies enacts distinctive structures of technology use. This model obviously has a particular attraction for the study of IS innovations like Internet-based self-service technologies, compared with previous IS models that attempted to incorporate structuration theory. One of the criticisms of past structurational models is their tendency to focus on stable technology. What sets this model apart is its proclivity to examine emergence, improvisation, and changes over time as designers reconfigure their technologies or as users alter their habits of use, and thereby enact different technologies-in-practice. Recently, Schultze and Orlikowski (2004) applied this model to understand the use of Internet-based self-service technologies in a business to business (B2B) setting. However, this model has not been applied in a business to consumer (B2C) setting and in the context of competing service channels. Furthermore, they tend to focus on 'use' with scant attention being devoted to 'design' facilities, norms and interpretive schemes that are so central to the study of implementation processes.

Table 4.1 lists the key conceptual elements for the initial framework for analysis. The relationships between these key concepts are illustrated in figure 4.4. Figure 4.4 takes a significant departure from the original model by incorporating design as an important analytical element.

Table 4.1

The social context of SST implementation: an initial synthesised analytical framework using structuration theory

Conceptual components	Associated conceptual elements
Practice lens	- Facilities, norms and interpretive schemes - Systems of signification, domination and legitimation, - Routinisation - Enabling and constraining features - Intended and unintended consequences

Assumption(s)

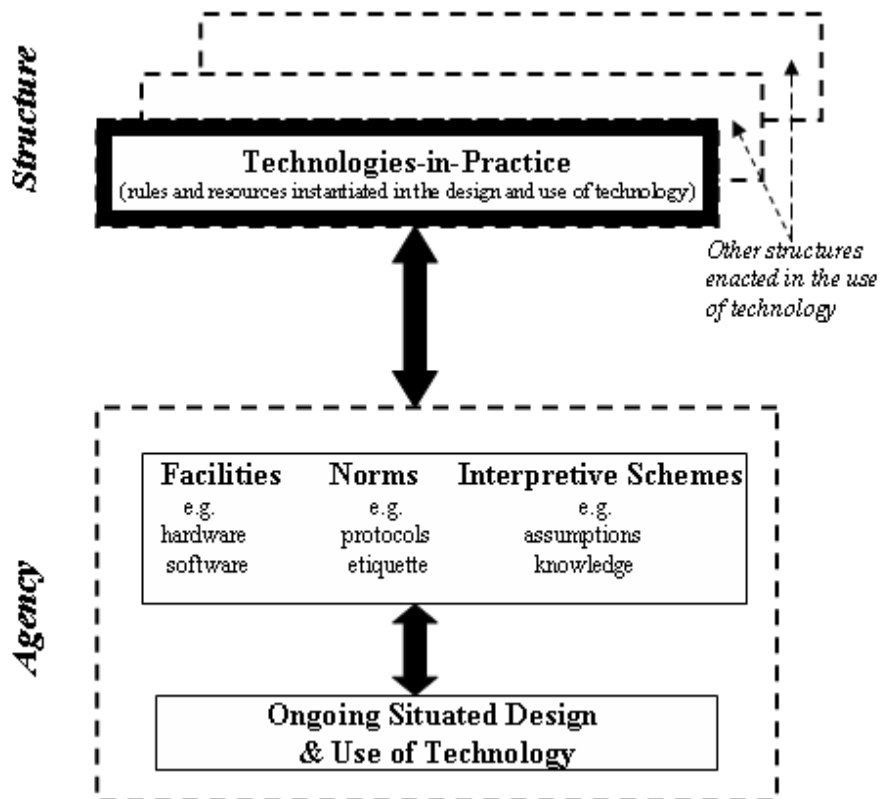The stability of technology and its application is always provisional.

Figure 4.3 Extended enactment of technologies-in-practice model

Source: Orlikowski (2000:410)

Principle

Users always have the potential to change the habits of use, and in this way change the structures they enact in their recurrent practices. Similarly, designers have the potential to change the structures they enact in their recurrent practices.

Agency

Agency refers to the capacity of human actors to do things and as such implies power.

Facilities

Actors draw on the facilities available to them such as hardware and software.

Norms

Actors draw on the norms such as protocols and etiquette that inform their ongoing practices.

Interpretive schemes

Actors draw on their tacit and explicit knowledge of their prior action and the situation at hand.

Ongoing situated design and use of technology

Actors recursively apply their knowledge, facilities, and habits of the mind and body to 'structure' their current action that inform their ongoing processes.

Structure

Structure refers to rules and resources recursively implicated in the reproduction of the technology-in-practice.

Technology-in-practice

Ongoing enactment of technology reinforces it, so it becomes regularised and routinised, an expedient and habitual response to repeated use of a technology within the daily demands of social life.

Rules

Rules refer to techniques or generalisable procedures applied in the enactment or reproduction of social practices (Giddens, 1984).

Resources

Actors depend on two types of resources. Allocative resources (such as land, raw materials, technology) refer to capabilities or to forms of transformative capacity.

Authoritative resources are non-material resources involved in the generation of power, derived from the capability of harnessing the activities of human beings.

Design and use

Design and use of technology using the practice lens involves a repeatedly experienced, personally ordered and edited version of the technological artifact, being experienced differently by different individuals and differently by the same individuals depending on the time or circumstance.

Other structures

The enactment of a technology-in-practice is situated within a number of nested and overlapping social systems. This means that people's interaction with technology will always enact other social structures along with the technology-in-practice, for example English as the primary language over the Internet.

In summary, these structuration concepts provide a rich framework to analyse empirical SST situations.

4.2.5 Critiques and limitations of structuration theory

Despite its growing popularity in IS, a number of social theorists have expressed common concerns about structuration theory's treatment of structure and agency. Archer (1982) argues that conflation, the problem of reducing structure to action or vice versa, weakens the analytical power of structuration theory. Archer (1982) proposes that to allow for their analytical separation, human action should be viewed over the short term, while structures should be seen as more enduring. In addition, Giddens' conceptualisation of structure as 'rules and resources' existing only in memory traces has led to criticisms of subjectivism. Critics argue that Giddens does not so much resolve the dualism of action and structure, as offer victory to the knowledgeable human actor (Clegg, 1989).

Another criticism is directed at structuration theory's inability to explain historical change. Giddens' view of structuration offers a conceptual mechanism for explaining the reproduction of social structure. However, some researchers argue that why some forms of social reproduction succeed and become institutionalised, and others do not are questions of more relevance to contemporary researchers (Stinchcombe, 1990). It appears that for these types of questions, structuration theory offers no direct answers.

Nor is it clear whether structuration allows for improvisation as a kind of purposeful behaviour. Ciborra and Lanzara (1994) propose a distinction between formative and structuring properties. The formative context suggests a background of meaning that can account for shift and drift phenomena. However, upon closer examination there is little distinction between structuration and formative properties. In fact, the treatment of the formative context is not unlike Giddens' (1984) suggestion that structure is paradigmatic.

Other criticisms are aimed at the lack of concrete empirical example in Giddens' own work. Critics contend that Giddens' abstract conceptual focus offers few clues as to how to gather useful understanding of the world of practice.

In IS studies, the limitations of structuration are that it does not allow us to examine the relationship between people and technology beyond the recognition that technology both enables and constrains us. After all, structuration is a theory of social organisation that explains change and stability in a social system over time. For example, structuration does not directly address how power and values are embodied or found in the use of technology. As a result, critics argue that structuration is not capable of unpacking exactly how technology regulates us, and how we react to technology. This leads Monteiro and Hanseth (1996) to argue that structuration simply does not provide a fine-grained analysis of the interaction between individuals and technology. Proponents of the use of structuration theory in IS also conclude that scholars need to better theorise the information technology artefact and move beyond the simple constrain-enable distinction (Orlikowski and Iacono, 1991).

Structuration theory in information systems research has also been criticised (Archer 1982; Barley, 1986; Orlikowski, 1993) for its relative neglect of technology. For Giddens, structure does not exist in material artefacts, such as technology, but in human memory traces and is seen

to be enacted through social practices (Jones and Karsten, 2003). Giddens and Pierson (1998:82) argue that 'technology does nothing, except as implicated in the actions of human beings'. Monteiro and Hanseth (1996:330) present the following critique of structuration theory's relative neglect of technology:

> *Our principal objection to conceptualizations like [Orlikowski and Robey 1991; Orlikowski 1991; Orlikowski 1992; Walsham 1993] is that they are not fine-grained enough with respect to the technology to form an appropriate basis for understanding or to really inform design.*

Monteiro and Hanseth (1996) were not claiming that structuration theory cannot deliver a satisfactory level of precision. It may be in its application that researchers have failed to derive a more fine-grained analysis. In applying structuration theory to the case analysis, I will assess the extent to which structuration theory naturally lends itself to how more specific technological elements and functions of an IS relate to organisational implementation issues.

Nevertheless, what is missing from structuration theory are concepts that allow the interrogation of the relationship between individuals and technology. It appears that such concepts can be found within actor network theory (ANT). ANT is largely concerned with the interactions between technology and individuals. According to some of the most prominent interpretive researchers in IS, ANT contains a wealth of concepts for understanding the relationship between technology and individuals, such as actors, networks, the process of inscription, and reconfiguration (Hanseth, 2004; Monteiro, 2000; Monteiro and Hanseth, 1995). They maintain that addition of these concepts will allow for the further theoretical development for the interplay between technology and the social. These concepts will be discussed in the next section.

4.2.6 Conclusion on structuration theory

Looking beyond its common criticisms discussed above, a number of the key tenets of structuration theory still have a particular appeal for the study of contemporary phenomenon like Internet-based self-service technology. First, structuration theory implies that social enquiry should not give prominence to either the individual experience of actors or the social totality, but rather focus on social practice that lies at the root of the constitution of both individuals and

society. Second, social practices are created by knowledgeable agents with causal powers, that is, powers to make a difference. Instead of viewing human agents as cultural dopes or the product of class forces, focus should be directed to their capacity for self-reflection in day-to-day interaction, a practice of 'tacit' consciousness of what they are doing and the ability under certain circumstances to do it. In the third place, structuration theory proposes that social practices are not random or voluntary, but ordered and stable across space and time. Simply put, they are routinised and recursive. In producing social practices which constitute society, actors draw upon structural properties (rules and resources) which themselves institutionalise features of society. Finally, structuration theory proposes that structure is activity-dependent. Structure is viewed as both medium and outcome of practices across time and space. Giddens (1984) refers to this process as the 'double hermeneutic', the double involvement of individuals and institutions. In the words of Giddens (1984:14), 'we create society at the same time as we are created by it'. These tenets will serve to broaden our understanding of the SST implementation phenomenon.

4.3 Actor-network theory

4.3.1 Introduction

Although the use of structuration theory to understand the implementation of SSTs offers a natural starting point for addressing my research questions, I explore here the actor-network theory (ANT) as a possible addition to help with a further understanding of the case. One of the main motivations for including actor network in my analysis is that it may offer additional insights with regard to the social dynamics of SST implementation. Structuration theory, including its enhancements, has a strong tendency to neglect the role of technology (the SST) in the social context. However, I would argue that IS implementation has both technical and social merits at the same time, and that therefore it might be appropriate to try to overcome the distinction between technical and social for a better understanding. What sets ANT apart from conventional theories is that neither the inherent properties of the technology, nor some properties of the social context – such as user traits and other actors – drive the success or failure of an implementation. Instead, it is the associations that exist and are created between the technology and its surrounding actors, that is, actors that are both technical and social. ANT presents a view of translation, which by focusing on associations rather than properties is radically different from some of the ideas proposed by structuration theory. In this section I will

describe ANT so that I can use it in my analysis of the case study. I will also highlight some of its uses in IS research and finally discuss its critiques.

4.3.2 Key elements in actor-network theory

In this section, I provide an understanding of the key elements in actor-network theory. ANT employs an idiosyncratic vocabulary, terms which at first glance may appear similar in other theories, such as systems theory, but have quite a different meaning in ANT. For example, ANT provides its own particular 'interpretation' to notions such as 'network', 'black-box' and 'translation'. These peculiarities will be discussed briefly in the subsequent section.

Overview of actor-network theory

Actor-network theory was pioneered by Michel Callon and Bruno Latour (Callon and Latour, 1981; Callon, 1986), and later extended and further developed by the original authors and other researchers (Latour, 1999; Law, 1991; Law and Hassard, 1999). Drawing on key concepts and assumptions from the social shaping of technology, Callon and Law (1989) make use of the actor-network approach to understand the heterogeneous and interrelated character of social and technological components. ANT contends that both social and technical determinism are flawed and advocates a socio-technical account (Callon and Latour, 1981) in which neither the social nor the technical is privileged. According to ANT, what appears to be social is partly technical and what appears to be technical is partly social (Law, 1991). ANT differs sharply from views that posit that purely technical and purely social relations are possible.

Based on these perspectives, actor-network theory, or the 'sociology of translations' is concerned with studying the construction and transformation of the heterogeneous networks (Law, 1992). These are made up of people, organisations, agents, machines and many other objects that constitute the world, existing of both humans and non-humans. As Latour (2005:10) comments on ANT's treatment of non-humans, 'They have to be actors and not simply the hapless bearers of symbolic projection' – a significant distinction from structuration theory, a distinction that will be tackled later (see section 4.6). ANT also delves into the ways in which networks of relations are arranged, how they emerge and come into being, how they are constructed and maintained, how they compete with other networks and how they are made more durable over

time (Tatnall and Gilding, 1999). In other words, when humans interact with other humans, these interactions are mediated through non-human artefacts of various kinds, and such interactions are mediated through additional networks of non-human artefacts, objects and humans. Hence, if material artefacts in these networks disappear, so too would 'social order'. ANT also investigates how actors enlist other actors into their world and how they bestow qualities, desires, visions and motivations on these actors (Latour, 1996). ANT thus offers a unique approach to theorising innovations such as self-service technologies and their implementation; an approach that resists the essentialist notions inherent in the conventional treatment of self-service technologies.

Understanding the essence of ANT

A key feature of actor-network theory is a perspective of the social world that shows it to comprise heterogeneous networks that form actors. All phenomena are the effect or the product of heterogeneous networks (Law, 1992). Even persons are made up out of a heterogeneous network, as Law describes (1992:379-380):

> *However, I will press the argument in another way by saying that, analytically, what counts as a person is an effect generated by a network of heterogeneous, interacting, materials. This is much the same argument as the one that I have already made about both scientific knowledge and the social world as a whole. But converted into a claim about humans it says that people are who they are because they are a patterned network of heterogeneous materials. If you took away my computer, my colleagues, my office, my books, my desk, my telephone I wouldn't be a sociologist writing papers, delivering lectures, and producing 'knowledge'. I'd be something quite other – and the same is true for all of us. So the analytical question is this. Is an agent an agent primarily because he or she inhabits a body that carries knowledge, skills, values, and all the rest? Or is an agent an agent because he or she inhabits a set of elements (including, of course, a body) that stretches out into the network of materials, somatic and otherwise, that surrounds each body?*

In attempting to understand social phenomena, researchers need to study actors without imposing on them *a priori* definitions or expectations. More specifically, ANT researchers must avoid *a priori* distinctions between the technical and the social. Or, put in another way, actor-network theory seeks to avoid both technological determinism and social reductionism (Monteiro, 2000).

ANT does not accept any form of reductionism, either technological or social, that splits up the technical from the social and supposes that the one drives the other. It states that there is no reason to assume, *a priori*, that either objects or people in general determine the character of social change or stability (Law, 1992). As such, ANT supports analytically treating objects and people in the same way; non-humans and humans together form the heterogeneous networks. This is best elucidated by means of an example. In our day-to-day lives, we are influenced by a wide range of factors, including social and technical, but also political and historical factors. For example, when driving a car we are influenced by traffic regulations, previous driving experience, road conditions and the car's manoeuvrability (Monteiro, 2002). To understand the phenomenon of driving a car, we should consider all these influencing factors together.

ANT offers three methodological principles to address the need to treat both human and non-human actors fairly and in the same way (Callon 1986): agnosticism, generalised symmetry and free association. First, in agnosticism, the researcher abstains from censoring or judging the actors, whether they are human or non-human. ANT therefore prescribes analytical impartiality towards both human and non-human actors. In the second place, generalised symmetry requires that researchers explain conflicting viewpoints of different actors in the same terms by use of an abstract and neutral vocabulary. This rule prohibits change registers when moving from the technical to the social aspects of the problem to be studied, and dismisses the affordance of any privileged explanatory status to technical or social actors. Finally, in the principle of free association, the researcher is instructed to abandon all previous distinctions between the natural and the social. There can be no boundary between the two, although they might be separated later, as the result of analysis and understood as outcomes or effects, but cannot be divided *a priori*, assuming it is the given order of things (Law, 1999). Similarly, Callon (1999:183) states:

> *ANT was developed to analyse situations in which it is difficult to separate human and non humans, and in which actors have variable forms and competencies.*

This treatment of human and non-human has a particular appeal for understanding contemporary forms of IS innovations such as Internet-based self-service technology.

Understanding the actor

According to Latour (2005:217) an actor-network 'is made to exist by its many ties: attachments are first, actors are second'. When we speak of actors, Latour (2005) proposes that we always add the large network of attachments that make the actors act. As alluded to earlier, actors are those elements in a context that shape action while pursuing their interests. An actor is something that acts or to which activity is granted by others. It implies no special motivation of human individual actors or of humans in general. An actor can literally be anything, provided it is granted to be the source of an action (Latour, 1996). But as Law (1992) argues, actors are also an effect generated by a network of heterogeneous, interacting, materials. In other words, social agents are never located in bodies alone, but rather are patterned networks of heterogeneous relations. By punctualisation or black-boxing, actor-networks themselves 'make up an actor'. Thus each actor is made up of actors and at the same time is part of an actor. Or, in the vocabulary of actor-network, each actor is itself a simplified actor-network and is at the same time part of other actor networks.

Law (1992) states that all attributes we normally associate with human beings, such as thinking, acting, writing, loving and earning, are generated in networks that exist beyond the body. Thus an actor is also always a network. Since actors are actor-networks in infinity, the researcher must choose how the network under research is 'zoomed in and out' and which actors are included. 'It entails that the "actor" of an analysis is of the "size" that the researcher chooses as most convenient relative to the direction of the analysis (Monteiro, 2000:82). However, being seen as an actor and thereby producing a simplification of complexity, either by researchers or, more importantly, by other actors, shows the result of a mobilisation process with black-boxing effects. The ordering these simplifications produce is neither natural nor 'obvious'. They are made obvious or natural in order to achieve an effect, namely to curb the opposition or alternatives. When choosing the size or shape of an actor, the researcher is not completely free, but bound by other actors' practice and what is obvious and natural to them (Monteiro, 2000). Thus, actors should have some obviousness and naturalness to them. In summary, an actor in ANT parlance is an effect of heterogeneous relations between humans and objects, and an actor is also, always, a network (Law, 1992).

Understanding the actor-network

The actor-network is realised through the common 'enrolling' of human and non-human participants into a network through processes of negotiation and translation (McLean and Hassard, 2004). It is through their alignment with each other that actors form an actor-network. This alignment is achieved through the translation of interests and the enrolment of actors into the network. Translating involves showing how an actor's non-aligned interests may become aligned. Alignment is established in inscriptions that give a particular precedence in terms of a viewpoint. Latour uses the term 'immutable mobile' to describe such network elements, meaning that when they are moved around in time and space, they remain stable and unchanged (Tatnall and Gilding 1999; Latour 1999b). For example, a car can be considered an immutable mobile when it displays a relational pattern of certain properties (such as infrastructure, oil industry, driving licences, traffic signs, garages, etc). Such a network can move through time and space without changing these properties. It displays properties of irreversibility (Walsham, 1997).

Law (1992) asserts that the actor-network approach is curious about how actors and organisations mobilise, juxtapose and hold together the bits and pieces of their composition. Law (1992) is also interested in how they are sometimes able to prevent those bits and pieces from following their own inclinations and 'making off'. Furthermore, of particular interest is how they manage to conceal for a time the process of translation itself and so turn a network from a heterogeneous set of bits and pieces, each with its own inclinations, into something that passes as a punctualised actor. Law (1992) emphasises 'time' because once a network is formed, it is not formed once and for all. It can always become unstable, since new actors, the desertion of existing actors or changes in alliances can cause the 'black-boxes' of networked actors to be opened and their contents reconsidered (Callon, 1986b). A black-box, which is also a network in its own right, is a way of talking about the simplified points that are linked together in an actor-network (Callon, 1986b). Latour (1987:108–121) describes five alternative strategies for enrolling others in the punctualisation or creation of a black-box:

- To appeal to the other's explicit interests ('I want what you want')
- To get the others to follow our interests ('You want what I want')
- To suggest a short detour ('I will take care of your interests, if you follow me')

- To reshuffle interests and goals by tactics such as inventing new goals and inventing new groups ('We all want this')

- By becoming indispensable to others ('You need me to get what you want')

A network recursively generates and reproduces itself, and relies on the active maintenance of its simplifications or 'punctualisation' for its continued existence. The term 'network', contrary to conventional uses of the word, does not imply some fixed thing, but describes a dynamic, actively shifting alliance of actors. A network becomes durable partly owing to a structure where each point is at the intersection of two networks: 'one that it simplifies and another that simplifies it' (Callon 1987:97). As already pointed out, actor networks are relentlessly produced and reproduced. The point here is not whether the actants of a network are social or technical, but, as Latour (1987:140) points out, 'which associations are stronger and which are weaker'.

One of the central concerns for ANT is the issue of power. It is argued, in particular, that power is effected through the production and reproduction of a network of heterogeneous 'actants' (McLean and Hassard, 2004). The notion of power is better explicated using the sociology of translation.

Understanding the sociology of translation

One of the central concepts in ANT is the sociology of translation. For Latour (2005:64)

> *ANT is the name of a movement, a displacement, a transformation, a translation, and enrolment. It is an association between entities which are in no way recognisable as being social in the ordinary manner, except in the brief moment when they are reshuffled together.*

More recently Latour (2005:108) defined translation as a relation that 'induces two mediators into existing'. However, 'translation' is a term used in many different ways. Literally the term 'translation' denotes two meanings, both relevant to ANT. In the first place, it is a change of position and a new interpretation. In other words, translation operates between actors: an actor gives definition to another actor by imputing these actors with interests, projects, desires,

strategies, reflexes, afterthoughts. And second, according to Callon (1986b), a translation is 'the methods by which an actor enrols others'. These methods involve:

- The definition of roles, their distribution, and the delineation of a scenario
- The strategies in which a future state actor-network renders itself indispensable to others by creating a geography of obligatory passage points
- The displacement imposed upon others as they are forced to follow the itinerary that has been imposed

In the creation of the actor-network or the process of translation, Callon (1986) discerns four 'moments': problematisation, interessement, enrolment and mobilisation. During the first moment of *problematisation* one actor, the initiator, makes an effort to make other actors subscribe to its own conceptions by demonstrating that it has the right solutions to, or definitions of the problem. Initiators try to demonstrate their quality of being indispensable to the solution of the problem during the initial stage. The problem is redefined or translated in terms of solutions offered by the initiator (Bloomfield and Best, 1992), who then attempts to establish themselves as an 'obligatory passage point' which must be negotiated as part of its solution. To pass through the obligatory passage point, the other actors must accept a set of specific conventions, rules, assumptions and ways of operating laid down by the first actor (Tatnall, 2000). Introna (1997) defines an obligatory passage point as a rhetorical device that presents the solution to the problem in terms of the resources of the agency proposing it. During the second moment of translation, *interessement*, an attempt is made to impose and stabilise the identities and roles defined in the problematisation on the other actors, thereby locking other actors into the roles proposed for them (Callon, 1986). Gradually existing networks are replaced by the new network (Grint and Woolgar, 1995). The third moment is *enrolment*. Enrolment occurs when a stable network of alliances is formed, and the actors yield to their defined roles and definitions. This involves a multilateral political process where the initiators seek to convince other actors. It is for this reason that Callon (1986: 211) states: 'To describe enrolment is thus to describe the group of multilateral negotiations, trials of strength and tricks that accompany the interessements and enable them to succeed.' The final moment is *mobilisation*. Mobilisation is a set of methods that initiators employ to ensure that allies do not betray the initiators' interests. During mobilisation the proposed solution gains wider acceptance and achieves stability. Stability implies that the

technology's content is institutionalised, and is no longer controversial, that is, it becomes taken-for-granted and is 'black-boxed'.

From an ANT perspective, material artefacts are significant in the structuring of these relations. As Lowe (2001:82) so lucidly puts it: 'Objects provide receptacles for human knowledge and vastly enable the process by which facts become accepted.' According to Latour (1987), the spread of ideas and their conversion to accepted facts is a rhetorical process. A fact is only established as such when a following is built up which uses such ideas increasingly and without modification. The fate of ideas depends on how those who come later use these ideas. In other words, ideas become more accepted, or taken for granted, as they are translated into inscriptions and incorporated into instruments and technology.

For example, the equivalent is the spread of self-service technology through the consumer-driven healthcare industry as healthcare insurance service providers have come to accept the technology as a fact, as a 'black box', a fact which can be accepted without modification. The networks constituted by SSTs are spread in this manner. Challenges are 'muted' as a consequence of the strength which the network achieves as least in part from many of the service providers who deploy these systems. Translation is necessary for stability in these networks, since actors from the outset have a diverse set of interests (Monteiro, 2000). Aligning these interests causes a network to become stable and durable. However, according to Mahring, Holmstrom, Keil and Montealegre (2004), the translation process does not necessarily pass through all the stages described above. It is plausible, then, that translation processes may fail and halt at any stage, depending on the strength of the network's inscriptions. In contrast to diffusion models, which assume technologies to be immutable, actors in ANT not only reshape technologies, but are themselves changed as the changing artefact spreads through the social network.

Besides the four stages of translation, the process of inscription is critical to building and stabilising actor-networks, as most artefacts within a social system embody inscriptions of some interests. Inscriptions refer to the way technical artefacts embody patterns of use (Monteiro 2000), or how certain viewpoints, values, opinions and rhetoric are converted into devices or materials such as reports, documents and scientific papers or frozen into codes or computer applications (Bowker and Star, 1994). For example, technological artefacts can also embody a worldview (inscription) that reflects the socio-economic context and rationality in which it was

created (Avgerou, 2001; Heeks, 2002). Chilundo and Aanestad (2005) also found that in the implementation of technologies in developing countries, the potential for a clash of rationalities is greater where the values inscribed in Western technologies conflict with values of developing countries.

During the development of a technology and in its placement in an actor-network, inscription takes place. This means that the technology does not have to be physically constructed for it to exist. It has to be conceived, but once it is conceived, it is a force to be reckoned with (Latour, 1996). Inscriptions also prescribe a program of action for other actors, which the latter may or may not follow, depending on the strength of the inscription (eg Latour, 1991). In relation to translation, inscription to a large extent takes place simultaneously and interrelatedly, that is, it starts as soon as a technology enters the picture and is beginning to be formed by its creators (Latour, 1991).

So, inscription implies that a material or technological artifact never begins as a blank slate. In other words, artefacts always embody the designer's beliefs, social and economic relations, previous patterns of use, and assumptions as to what the artefact is about (Akrich, 1992). For example, inscription is used when designers formulate and shape self-service technology in such a way so as to lead and control users. Inscription can also refer to the way technical artefacts embody patterns of use, including user programs of action. The term also encompasses the roles users and the system play (Monteiro, 2000). Since inscription can guide users to join or behave in a way that forces a definition of the form and function of the technology, many actors actively seek to inscribe their vision and interests into the artefact (Faraj, Kwon, and Watts, 2004). Inscriptions may also lead to irreversibility.

Another important phenomenon and concept of ANT, then, is irreversibility. Irreversibility refers to the degree to which in a certain situation it is impossible to go back to a point where alternative possibilities exist (Callon, 1991). Irreversibility is often the result of the inscription of interests into technological artifacts, whereby those interests become increasingly difficult to change (Hanseth and Monteiro, 1998; Mahring *et al*, 2004). As ideas are inscribed in technology artefacts and as they diffuse in their relevant contexts, they help achieve socio-technical stability (Latour, 1987). While technology artefacts are, in part, open to interpretation, there are some features that are in practice 'beyond' (re)interpretation and that increase stability in the networks

in which technologies are encompassed (Latour, 1991). According to Bloomfield, Coombs and Knights (1997), information technologies in particular 'stand in for' or 'speak for' human actors. In this way they can hide decision processes and become 'frozen organisational discourse' (Bowker and Star, 1994:187). Walsham (1997) notes that inscriptions developed in software, as frozen discourse, may resist change and display signs of irreversibility.

Hanseth and Monteiro (1997) suggest four notions of inscription and translation that should be emphasised in a study:

- The identification of explicit anticipations or scenarios of use held by the various actors during design
- How these anticipations are translated and inscribed into materials
- Who inscribes them
- The strength of these inscriptions, that is, the effort it takes to oppose or work around them

Another important concept related to translation and inscription is framing. According to Faraj *et al* (2004), inscribed patterns of use do not succeed when users do not conform to their assigned program of action. In many new technologies, users modify and adapt the artefact into new forms of use. Based on actual practice, unexpected uses are developed and new functionality is envisioned, leading to a new perspective on what the artefact does and is expected to do. Thus, when studying the user of technical artefacts, one necessarily shifts back and forth 'between the designer's projected user and the real user' in order to describe this dynamic process of negotiating design (Akrich, 1992).

According to Faraj *et al* (2004), it is necessary to measure which of these superimposed inscriptions actually succeeds in shaping the pattern of use in order to measure the strength of an inscription (Monteiro, 2000). Faraj *et al* define framing as the emergent outcome of the process of an artefact instantiation-meeting practice. Once a set of accepted facts is recognised as core to the technology, the process is hard to reverse: these facts are embedded in the expectations of users, evaluation routines (standards), and the beliefs of developers themselves. New versions of the technology must then incorporate or improve on this functionality. On a macro level, framing is defined by how key actors engage in actions in support of a certain vision or pattern of usage.

As with the micro level, this results in irreversibility (Hughes, 1994; Hanseth and Monteiro, 1997). Faraj *et al* (2004) claim that the examination of how users and actors together frame criteria for selecting and stabilising features is essential to understanding the irreversible and path dependent aspects of technology evolution.

4.3.3 ANT in information systems research

Researchers have long recognised the potential of employing ANT to investigate the successes and failures of technological innovations. Some seminal examples include Callon's (1986a) study on the failure of the domestication of the scallops of St Brieuc Bay and the development of the electric vehicle by the Electricité de France (Callon 1986b). Latour also used actor-network theory to analyse the development of a revolutionary public transportation system known as Aramis (Latour 1996) and to discuss the achievements of Louis Pasteur (Latour 1999b). More closely related to IS are organisation studies which have applied ANT in organisational analysis (Hassard, Law and Lee, 1999) and studies in accounting systems (Lowe, 2000, 2001).

More recently, many IS researchers have gone to great lengths to demonstrate the potency of ANT in IS studies. The use of ANT in IS has been popularised by the attempts of Scandinavian academics to apply the sociology of translation and inscription to the study of information infrastructures (Monteiro, 2000; Monteiro and Hanseth, 1996). Monteiro and Hanseth (1996) also studied the role of standards in electronic data interchange (EDI) systems and information infrastructure. Wagner (2003) recently drew on ANT to study the design and implementation of an enterprise resource planning (ERP) system in an academic environment. Mantovani and Spagnolli (2004) also used ANT to describe the struggle in a firm to negotiate and legitimate a new network technology.

ANT has been used to interpret the political processes of IT implementation (eg Walsham and Sahay, 1999). More specifically, ANT's treatment of information systems implementation as a complex socio-technical and heterogeneous network comprising actors, institutional arrangements, textual descriptions, work practices and technical artifacts has a particular appeal for this study. For example, Walsham and Sahay (1999) used ANT to demonstrate the mutual

dependency between technological properties and the social context in the implementation of GIS systems by Western developers in a local Indian setting.

Research drawing upon ANT seeks to:

> *examine more than just the technological system, or just the social system, or even the two side systems side by side; … but the phenomena that emerge when the two interact* (Lee 2001:iii).

In addition, Hanseth and Aanestad (2004:117), in their introduction to a special issue of ANT in the *Information Technology and People* journal recently emphasised the superiority of ANT over structuration theory in its analysis of technology:

> *'The Structuration theory approach has been picked up by a vast number of scholars and a wide range of studies have been carried out. These have given us many valuable insights into the social processes related to adoption and use of information systems. There is one aspect of these studies that is of crucial importance here. That relates to the role of technology in these studies as well as the theories they are based on. These go equally well (or more precisely, badly) for both Structuration theory and institutionalism. The studies of information systems based on these theories do not address the role of technology in a proper way. This fact is largely a consequence of the fact these theories totally ignore technology. This makes ANT – and the technology studies part of the STS field – different. And in this respect ANT offers some unique and very important contributions to information systems.'*

In the special issue mentioned above, five exemplar papers were presented on ANT applications to different technologies and settings. Marres (2004) used ANT as a conceptual and methodological tool for the development of research practice. Adams and Berg (2004) used ANT to show how the notion of 'reliability' of health information is subject to negotiation. Allen (2004) studied the concept of inclusion in multiple technological frames using a socio-technical approach and Mahring *et al* (2004) tackled a comparative use of ANT and escalation theory to analyse dysfunctional IT projects. A study which has a particular appeal for this thesis is Faraj *et*

al's (2004) study, which used ANT as a basis for studying the evolution and complex processes of World Wide Web technology. Nevertheless, like structuration theory, ANT is not free from criticisms either.

4.3.4 An actor-network framework for analysis

Actor-network theory will be used to analyse the same case and investigate how explanations provided by structuration theory to the research questions can be understood differently using ANT. Although ANT and the sociology of translation have become prominent streams of research on IS innovations, few of these models were able to distinguish between the designer's beliefs, the user and the evolving artefact. Recently, Faraj *et al* (2004) built on actor-network theory as a basis for studying the complex evolutionary processes of modern information technologies. A particular appeal of this approach is the authors' attempt to reject the subject-object distinction. Furthermore, their theoretical lens was able to reveal the interdependencies between actors and how processes of inscription, enrolment, and framing dynamically enabled and constrained information technology development. However, their original model overemphasises the role of designers and neglects the user in these processes.

Table 4.2 lists the key conceptual elements for the initial framework for analysis. The relationships between these concepts are illustrated in an adapted model in figure 4.4. Figure 4.4 presents the core ANT processes utilised in this study for understanding Internet-based self-service technology implementation model.

Table 4.2

The social context of SST implementation: an initial synthesised analytical framework using actor-network theory

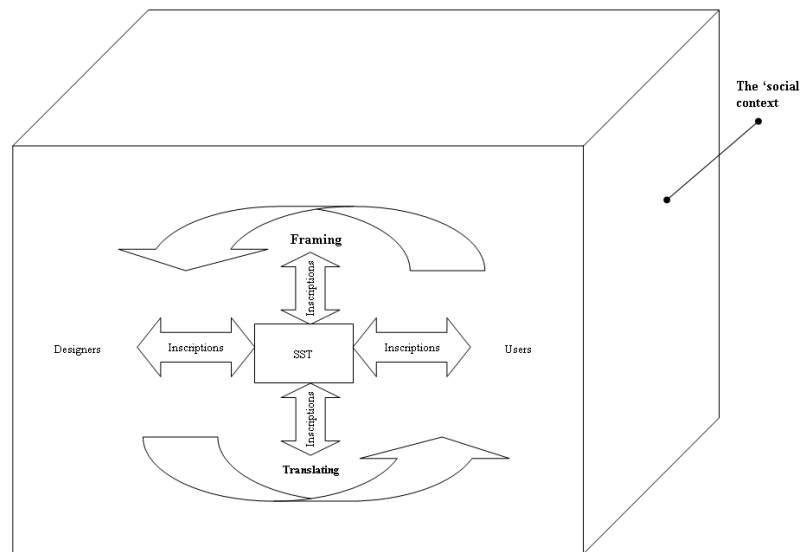| Conceptual component	Associated conceptual elements
Sociology of translation	• Translation • Framing • Inscription

Figure 4.4 ANT process model for SST implementation

Source(s): Adapted from Faraj, Kwon and Watts (2004: 190) and du Plooy (1998)

Principle

Implementation processes show recurring operations in which networks of actors continuously react and interact, creating a spiral of technology implementation. These processes operate recursively within a social context.

Assumptions

Technology artefacts are not merely physical because they include and embody the active projection of actors, such as the motives, intentions, interests and prejudices of the designers and users.

Processes

Implementation of technologies should emphasise broader processes of implementation encompassing efforts across organisations and communities.

> **Inscribing**
>
> Entities that make up a network are often converted into inscriptions of devices such as documents, reports, models and software. This process is concerned with how ideas, values and intentions of social actors become inscribed in technology. Inscriptions

prescribe a program of action for other actors, which the latter may or may not follow, depending on the strength of the inscription. An inscription that remain stable and unchanged, that is, exhibits strong properties of irreversibility, is termed immutable mobile (Latour, 1997, Walsham, 2001).

Translating

Translation describes a variety of ways in which actors actively seek to interest others in supporting the construction of a claim, enrolling them directly or indirectly in a coalition dedicated to building a fact or a machine (Latour, 1987). The process of creating these actor-networks consists of four major stages: problematisation, interessement, enrolment, and mobilisation. Translation processes do not pass through all these stages and may fail and halt at any stage.

Framing

The framing process describes the emergent outcome of the technology meeting practice. Key actors engage in actions in support of a certain vision or pattern of usage. However, unexpected uses are developed, leading to a new perspective on what the technology does and is expected to do.

It is envisaged that analysing these processes that are bound in a particular social context will draw greater attention to the dynamics of IS implementation and provide rich insights into their emergent consequences. To complement the above ANT process model for SST implementation, figure 4.6 will be used to describe the translation sub-processes in more detail.

Principle

Actor-networks are configured by enrolment of human and non-human allies via a series of negotiations in a process of redefinition where one group of actors imposes definitions and roles on other actors (Callon, 1986b).

Assumption(s)

The model in figure 4.5 assumes that the enrolment of human and non-human allies is developed through four distinct subprocesses.

Stages

Problematisation

In the problematisation stage, a group of one or more key actors defines the problem and suggests solutions that make them indispensable to the solution.
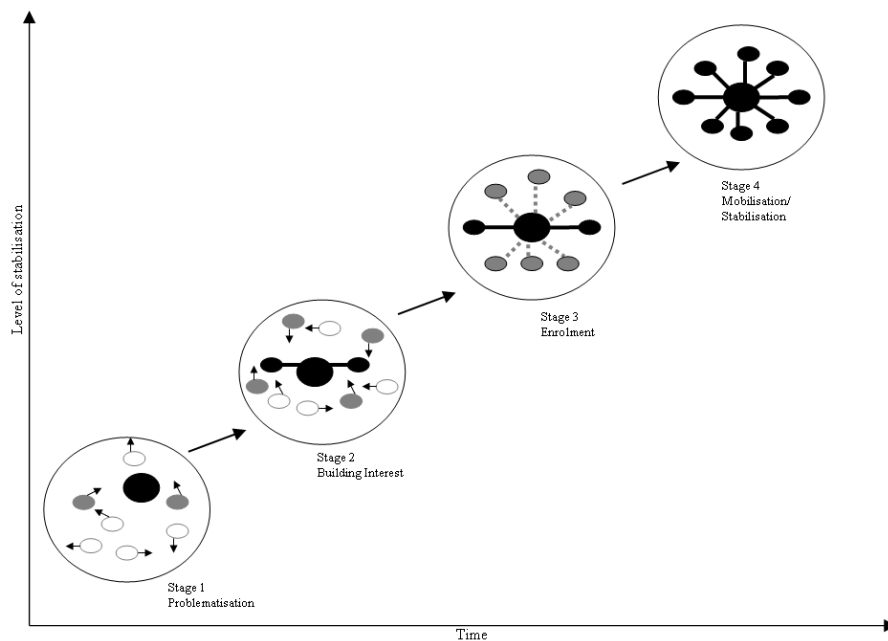
Figure 4.5 The four moments of translation

Source: Adapted from Callon in Law (1986:196–223)

Interessement

In this stage, the key actors build interest and lock key allies in, by finding ways to (re)formulate the problem or solution in such a way that key allies will associate their own interest with the formulation

Enrolment

In this stage, the problem or solution is established as an accepted fact, that is, the problem or the solution is legitimised, by controlling or influencing the production of facts, by using allies and spokespersons, and by inscribing problem or solutions in the organisational memory (eg documented agenda, minutes).

Mobilisation / stabilisation

Finally, during mobilisation, the solution gains wider acceptance. Compliance is ensured by monitoring the network and addressing descent as and when it arises. The key actors use the stability in the network to enact solutions.

The model above uses the notion of weak ties (depicted by broken, thinner and disconnected lines) and strong ties (depicted by darker and connected lines). Whereas actors during the earlier stages such as problematisation are characterised by fragmented alliances and instability (weak ties), through the process of translation, actors are progressively locked into alliances, whereupon they come together and the network stabilises (strong ties). The depiction of the orderly sequence of the translation sub-processes is merely an analytical convenience. It allows us the ability to construct an understanding around the sequence of events that lead to the outcome of the SST implementation.

To summarise, ANT views implementation as an emergent process initiated and guided by actors (agenda setters) with specific interests. Their agendas are enacted through processes of inscription, translation, and framing. Inscription and the sub-processes of translation are used to enrol dissidents who oppose the new agenda. By inscribing the agenda in material artefacts, actors enable material artefacts such as information systems to assume the role of actors in the network; that is, they stand in for the agenda setters. However, unexpected uses are developed, leading to a new perspective on what the technology does and what it is expected to do. The framing process describes the emergent outcome of the technology meeting practice.

4.3.5 Critiques on and limitations of ANT

Walsham (1997) recognised four main areas of critique directed at ANT in the literature. These include ANT's disregard for social structures; its disregard for moral and political analysis; its symmetrical treatment of humans and non-humans; and its descriptive power as opposed to power to explain, together with the problem of managing vast amounts of detail.

ANT has been criticised for neglecting macro social structures and focusing only on local contingencies, that is, how to relate the local and global. However, ANT proponents argue that

macro levels can be investigated with the same methodological tools as the micro-level, since the macro-structure of society is made of the same stuff as the micro-structure (Latour, 1991). ANT allows movement between levels of analysis. Latour (1999a: 19) denies a difference between macro structures and micro interactions. To Latour, differences between network and actor 'are two faces of the same phenomenon'. Walsham (1997) suggests drawing upon the structuration theory by Giddens (1984), which links levels of analysis from the individual to the global and offers models of social action and structure at multiple levels to overcome this problem. However, Latour (1999a:17) argues that the social possesses the bizarre property of not being made up of agency and structure at all, but is a 'circulating entity'. Following Latour's (1999) perspective then, the social is always circulating between actor and network. This treatment of society is similar in many ways to Giddens' treatment of structure and agency – as a duality. In addition, Callon and Latour (1981) argue for the use of the same framework of analysis for tackling both a 'macro-actor' and a 'micro-actor' and for making the notion of an actor-network scalable: one element of an actor-network may be expanded into a new complete actor network, and vice versa, a whole actor-network may be collapsed into one element of another actor-network (Monteiro, 2000). Impacts on the micro or respectively the macro level can thereby be analysed to show their effect on either of these levels. In this way, structuration theory is very similar to ANT. Giddens also (1984:139) dismisses distinctions between 'micro' and 'macro' social studies as being nothing more than a 'phoney war' and as 'misleading'. 'I do not think that there can be any question of either having priority over the other' (Giddens, 1984:139). Thus, for reasons outlined above, some researchers have dismissed Walsham's (1997) position on synthesising ANT with the work of Giddens.

A key criticism of ANT stems from its assumption of according symmetry between the social and the technological in the actor-network (Walsham 1997). The main critique is that people have been reduced to the same level as things and machines. A number of critics argue that treating all actors as equal is problematic: not all actors are equal; some exert a stronger influence than others. However, human qualities such as emotions, which play a vital role in human activity, seem to be lost (Mutch, 2002). Moreover, with the notion of inscriptions, technology seems to have been granted some deterministic property. But ANT does not claim humans and machines are the same – it merely states that one should first attempt to discover the influential elements that actually determine action, be it technical or non-technical (Monteiro, 2000). Hanseth *et al* (2004) counter this criticism by arguing that this is an unfounded claim.

They argue that while it is true that ANT assumes everything to be an actor-network, and accordingly so are both human and technologies, all networks are nevertheless inevitably different. In other words, different technological artefacts and different humans play different roles in social life. For example, to discover the influential factors that affect the way we drive our car, we need to know the engine's capacity (technical) as well as the driver's training (non-technical). Rather than distinguishing technical and non-technical a priori, ANT argues that they might have more in common than not. To perceive the term 'inscription' as being an action that is inscribed and hard-wired into an artefact is a misconception (Monteiro, 2000). It is merely used to describe how concrete anticipations and restrictions of patterns of use are involved in the development and use of a technology. It is neither the case that the object determines its use, nor that an object is infinitely flexible in its interpretation and appropriation by its user; it is an interplay between both extremes. Nevertheless, Collins and Yearley (1992) insist that ANT concedes too much to realist and technical accounts. Similarly, Grint and Woolgar (1997) accuse ANT of technicism in its need to refer to actual technical capacities of technology. Brey (1997) argues that ANT is somewhat similar to social-shaping approaches, and suggests that the notion of inscriptions can be viewed as a metaphor for the 'politics of artefacts'.

Walsham (1997) also criticises the amoral stance of ANT and its associated lack of insight concerning political viewpoints. He proposes that additional political and ethical theories might be needed to understand case findings. For example, the reason for the African continent almost totally being excluded from the Internet cannot be understood by simply investigating the network. He suggests that the empirical results from an ANT study should also be debated in terms of the moral and political issues.

Knights and Murray (1994) also criticise ANT for the way in which it gives little or no attention to the broader powers and inequalities that are both the condition and consequence of network formations. Latour's (1999:197) counter-argument is that critical theorists rely too much on inequalities of the social.

> *Critical theory is unable to explain why artefacts enter the stream of our*
> *relations, why we so incessantly recruit and socialize non-humans. It is not*
> *to mirror, congeal, crystallize, or hide social relations, but to remake these*
> *very relations through fresh and unexpected sources of action. Society is not*

stable enough to inscribe itself in anything. On the contrary, most of the features of what we mean by social order – scale, asymmetry, durability, power, hierarchy, and the distribution of roles – are impossible even to define without recruiting non-humans. Yes, society is constructed, but not socially constructed. Humans, for millions of years, have extended their social relations to other actants with which, with whom, they have swapped many properties, and with which, with whom, they form collectives.

In other words, while it is true that there are inequalities, inequalities are not *a priori*, dividing the social and technological. ANT does not accept any reductionism; neither machines nor human relations are determinate. It is argued that there is no reason to assume, *a priori*, that either objects or people in general determine the character of social change or stability. Indeed, in particular cases, social relations may shape machines, or machine relations may shape their social counterparts. The social and the technical might be considered separate when understood as effects or outcomes, but not as given in the order of things (Law, 1999). Both, through inscriptions, can have an impact on the resulting inequalities between actors. The same conclusion as before can be drawn: 'Moral and political issues should be debated from a solid empirical base, and actor-network theory offers a contribution to the latter if not directly to the former' (Walsham 1997:475).

ANT is also posited to be much more a method for describing than explaining (Bloomfield and Vurdubakis, 1999). However, Latour (1999) argues that ANT does not claim to explain the actor's behaviours and reasons, but only to find the procedures which render actors able to negotiate their ways through one another's 'world-building' activity. In other words, ANT was never intended to explain the behaviour of social actors, but in a much more ethnographic sense a way for researchers to study what, how and why actors behave the way they do, and not claim to explain this behaviour by all kinds of exterior forces unknown to the actors themselves (McLean and Hassard, 2004).

Another point of critique on ANT is the position of the researcher. The role of the researcher in labelling actors, defining passage points, scoping the actor-network, telling the story and so on is very influential in the results that an ANT study delivers (McLean and Hassard, 2004). The researcher enters the study with his or her own theoretical backgrounds, ideas and

preconceptions (Clarke, 2001). A way to deal with this critique is to adopt a more reflexive approach towards the researcher within the study. Moreover, Monteiro (2000:76) argues that 'employing ANT still requires a researcher to make critical judgements about how to delineate the context of study from the backdrop', that is, the researcher should be critical in his or her labelling of actors and in the analysis in general, thereby being guided by the actors themselves.

Some critics have argued that ANT might be too 'flattening'. In other words, by perceiving actors equally, important social constructions and discourses may be lost. Furthermore, critics continue to assert that reliance on the configuration of actor networks is not enough to explain why and how some actors are more empowered, while others are disempowered (McLean and Hassard, 2004). In applying ANT, the role of 'exogenous contingencies' such as economic crisis, deregulation and management principles may be undervalued. Some critics argue that conceptualisations of the market, economics, organisations, management or culture should be explored further (Bloomfield and Vurdubakis, 1999). Other researchers are suggesting that it is not sufficient to understand that actors hold particular beliefs or interests. Instead, an analysis is needed to understand how and why a certain actor has taken these beliefs for granted and how they have shaped the actor's interests .

Monteiro (2000) highlights a few issues regarding ANT as a methodology. These include that unpacking a network will cause an explosion in terms of complexity as each actant can potentially be expanded into another whole actor-network. Furthermore, ANT does not specify how to delineate one actor-network from the next (Monteiro 2000). Despite these criticims, I am still convinced that ANT points to a better possibility of understanding technology in the social context.

4.3.6 ANT and interpretive research

Looking beyond the criticisms of ANT, a number of researchers have argued that using ANT as a lens in interpretive research contradicts ANT's fundamental ontological stances (Cordella and Shaikh, 2003). I do not claim to have used 'the actor-network theory' as an ANT purist would have, even though there is currently not any single definition of what that might be. In fact, the approach that I have used is likely to offend an ANT purist. Notwithstanding this, it is my intention to show how some of the ANT concepts are extremely powerful and appropriate

analytical devices for understanding the SST implementation phenomenon. For interpretive case study research, ANT provides a lens through which to review the research setting and a language for discussing the dynamic events in which the research is located. While I would reiterate that I will not explicitly use ANT in a purist sense, it has indeed tacitly influenced my research owing to my prolonged exposure to and immersion in its concepts and applications.

ANT will guide my research primarily epistemologically related to my choice to study the process of SST implementation as based on a complex actor-network. I will emphasise how technology and related components, as well as other actors, influence the actions of the variety of the actors. Implementation of an SST should thus be seen as based on a negotiation process involving a heterogeneous and socio-technical actor-network. My interpretation of the implementation process will not be based only on my interpretations of the interviews, but also on my interpretations of the roles of the various social and technical actors involved. In this way I will try to make sense of how the various social actor interests and intensions are reflected in non-human components and further, how this is reflected in their use. ANT will also sensitise me to important aspects of SST implementation by focusing on how technology influences human behaviour (Monteiro, 2000). It will help me to transcend technology and social determinism by focusing on the mutual influences of humans and non-humans. Furthermore, the use of ANT for sense-making and as a device for interpreting and describing so as to improve understanding has long been supported by stalwarts in IS interpretive studies such as Walsham (2001). On a closing note, my use of ANT is not restricted to certain analytical levels, but will rather encourage me to open those black boxes which I found interesting, and to close others. By using ANT in this way, I have also tried to avoid the risk of not being open to field data which a more rigid use of theory may produce (Walsham 1995).

4.3.7 Conclusion on ANT

> *We contend that actor-network theory can be useful in studies of information systems where interactions of the social, technological and political are regarded as particularly important … We suggest that actor-network theory, and the theory on innovation translation, can be particularly useful for studies in areas such as the business use of the World Wide Web …'* (Tatnall and Gilding, 1999:963)

ANT provides an approach to analyse the case study that is promising. It offers a language of analysis that sensitises us to new ways of understanding. The dichotomy between the social and the technical is solved by the perception that both are intertwined. Moreover, ANT does not reduce *a priori* IS implementation to simplistic factors, but it is able to analyse it in all its complexity. It cuts across economic, political, strategic, social and technical issues related to IS implementation and allows for making sense of the unfolding implementation process (Monteiro, 2000). Though still a nascent theory in IS research compared with structuration theory, ANT has already demonstrated huge potential in IS research.

4.4 Comparing structuration theory and ANT

I have discussed two theories to help to understand the implementation of IS in organisations. In this section I will zoom in on some of the key differences between the two.

The main difference between the structuration theory and ANT is in their social-technical stance. In structuration theory, technology's capability to make a difference is unacknowledged. It is relegated to the status of a facility or tool employed by human agents. In actor network theory, on the other hand, technology becomes an independent actor in its own right, yet no distinction is made between the agency of technology and humans. In ANT, the social system and technology are considered inseparable.

In comparing these two theories, some researchers have argued that the different and incompatible treatment of agency is irreconcilable (Rose, Jones and Truex, 2005). They argue that neither structuration theory nor actor-network theory offers a particularly convincing account of the interaction of humans and machines, and that their different accounts of agency make them hard to integrate in any meaningful way. Some researchers are suggesting that structuration theory exaggerates the role of human agency in creating and producing its context. Proponents of ANT perceive the context to be both social and material, which is a hybrid of both human and non-human actors. Advocates of structuration theory argue that human and non-human agency cannot be labelled as equivalent.

However, despite the issue of agency, Walsham (2001) made a valuable contribution by combining these two theories in the same case, using structuration theory to guide broader social analysis, and ANT to describe the detailed socio-technical processes that took place. While these concerns over agency are valid, they do not hold for the purposes of this research since ANT is not applied here as a rigorous scientific theory, but rather as a flexible methodology.

> *… (ANT) not only provides theoretical concepts as ways of viewing elements in the real world, it also suggests that it is exactly these elements which need to be traced in empirical work.* (Walsham 1997:6)

So while some researchers have made some valid calls to review our understanding of agency, this was not deemed to be a particular objective of this study. Despite the theoretical differences between ANT and structuration theory, I use the two theories separately as sensitising devices (Walsham, 2001; Rose, Jones and Truex, 2005). More specifically by using these two theories, I will be able to conduct two separate analyses and assess how these two theories compare, and furthermore establish what different insights they bring to the case. I envisaged that insights from the analysis of these theories in chapter 7 and chapter 8 and the development of a conceptual framework in chapter 9 may advance our understanding of the 'agency problem'.

4.5 Conclusion

In this chapter I discussed structuration theory as a starting point from which to analyse the case study. Structuration theory was chosen because of its wide acceptance by researchers interested in technology and the social context, and hence for its compatibility with this study's research questions. However, after having considered the potential limitations of structuration theory, actor-network theory was added to the case analysis. In chapter 7 we will start with a brief analysis based on Giddens' structuration theory to show where this theory can be helpful in understanding the social context of self-service technologies. In chapter 8, ANT will be used to analyse the same case even further and investigate how potential problematic explanations from structuration theory can be understood differently. It is envisaged that these two separate lenses will help to provide a greater understanding of the social context of self-service technology implementation. The next two chapters focus on describing the case study to be analysed.

Chapter 5

Background to the Healthcare Insurance Context

Table of contents

5.1 Introduction

In chapter 1, I expressed the problem of self-service technology implementation, and how this was shaped and influenced by broader processes such as global, organisational, inter-organisational, group, and micro processes of change experienced by individuals. In the literature review a number of international- and national-level contextual influences could be seen to be playing out within the implementation of Internet-based self-service technology in the healthcare insurance context. The goal of this chapter is to provide a description of the case within this wider social context, before providing an organisational background in chapter 6. By sketching this broader context, it is envisaged that richer interpretations will be made possible in chapter 7 and chapter 8. The separation of broader and local contexts is largely an analytical device. In chapter 7 and chapter 8, the inextricable link between the broader and organisational contexts will become self-evident in the way they are implicated in the SST implementation. Thus, to understand the environment of the case study, this chapter presents background information on healthcare and healthcare financing models in particular, and situates the phenomenon of self-service technology implementation within a developing country and first world healthcare insurance industry dynamic.

The first section of this chapter provides a synopsis of global and local issues in healthcare generally and the implications for the healthcare insurance market in particular. The second section of this chapter provides a brief background to the South African healthcare insurance services industry and contextualises the role of other social actors, including regulators, insured members, service providers, brokers, and employers. The third section draws attention to the historical and current status of ICT and e-commerce in the South African context and the financial services industry specifically, with emphasis on the Internet and the enablement of online self-service technologies.

As Walsham (1993) advises, the analysis of information systems should not be restricted to the boundaries of that system. Accordingly, this chapter sets the scene for the following three chapters: chapter 6 being a description of the case study organisation; and the subsequent analysis of the case in chapter 7 and chapter 8, using the initial conceptual framework conceived in chapter 4.

5.2   Broader social issues in healthcare

5.2.1 Introduction

Healthcare worldwide in the new millennium can be characterised as facing enormous challenges. In some circles, global healthcare is being viewed as undergoing a severe crisis. Some of the key reasons for the crisis include the startling inequities between the haves and have-nots, the proliferation in healthcare technology and its costly application to medicine and healthcare in the form of healthcare innovations, and the sedentary lifestyle of the wealthier, coupled with rapid population growth and changing disease profiles during this century (World Health Report, 2002). Hence, setting the right financial incentives for providers and ensuring that all individuals have access to effective healthcare is becoming an even more daunting challenge. These factors have led to intense debates on the design and structure of healthcare financing systems. The search for the perfect solution has eluded many healthcare experts. It is not surprising then that healthcare financing systems in many countries continue to be both of an eclectic nature and in a state of enormous flux (World Health Report, 2000). To understand the environment of the case study, I explore some of the broader trends influencing the healthcare insurance industry.

5.2.2 A synopsis of global healthcare

Dr Gro Harlem Brundtland (2002:4), director general of the World Health Organisation (WHO), sums up lucidly the statistics presented in table 5.1 below:

> *The world is living dangerously either because it has little choice*
> *or because it is making the wrong choices about consumption and activity.*

The first striking observation in table 5.1 is the number of deaths or diseases related to causes that are viewed to be within the control of individuals. For example, chronic non-communicable diseases which are linked to factors such as smoking, obesity and a sedentary lifestyle cause 20% of the society's disease burden. While an enormous part of the crisis in healthcare can be attributed to individual behaviours related to risk, such as food intake, smoking and sexual behaviour, a number of broader factors are influencing the spread of non-communicable diseases. According to the World Health Report (2002), changes in food processing, production, distribution, and agricultural and trade policies are affecting the daily diet of hundreds of

millions of people. At the same time, changes in living and working patterns have led to less physical activity and less physical labour. Even in the United States, Merrill, Shields, and White and Druce (2005) report that many adults live a sedentary lifestyle, although a variety of health benefits have been associated with regular physical activity. Similarly, a 2002 report from the US Department of Health and Human Services indicated that 7 in 10 adults are not involved in regular physical activity, including 4 in 10 who are not active at all, contributing to an estimated 300 000 preventable deaths.

In addition, infectious diseases that are perceived to be beyond the control of individuals are responsible for 30% of the disease burden. People located in less-developed countries such as sub-Saharan Africa and Asia are among the most vulnerable to infectious diseases. Factors that are beyond the direct control of individuals include environmental factors related to water, sanitation, and a rising drug resistance to major pathogens. In developing nations (especially in Africa) that are battling to conquer infectious diseases that have always afflicted them, factors that were commonly associated with the industrialised nations are creating a double burden. For instance, more than 40% of all deaths in the poorest 20% of the world's population are already due to non-communicable diseases (World Health Report, 2002). In addition, the consequences of HIV/Aids, the fourth biggest cause of mortality in the world, are extending beyond mortality as children are orphaned and entire economies are affected (Sen and Bonita, 2000).

At the same time there have been impressive and unrivalled gains in health status worldwide in the 20th century, with increasing life expectancy being observed in both developing and developed countries (Sen and Bonita, 2000). In fact, life expectancy at birth in most developed countries at over 80 years is double that of the most disadvantaged countries. However, already more than half (57%) of the world's population over the age of 65 years live in less-developed countries.

Table 5.1

Summary of key global health statistics

Subject	Measures
Underweight children (developing countries)	170 million
Overweight (worldwide)	1 billion
Deaths from obesity-related diseases	0.5 million per annum
Lung cancer from smoking	
Men	90%
Women	70%
Global deaths from tobacco-related causes	8.8% (4.9 million per annum)
Global deaths related to alcohol	4% (1.8 million per annum)
Physical inactivity	1.9 million deaths per annum
(causes about 15% of some cancers, diabetes and heart disease)	
HIV/Aids infections	40 million people
Global burden of infectious diseases	30%
Chronic non-communicable disease burden	20%
(Five risk factors: unsafe sexual practices, alcohol use, pollution, occupational exposures, and tobacco use)	
Life expectancy (global average)	
1950	46 years
1998	66 years
Europe	73 years
Low- and middle-income countries	68 years

Source: WHO (2002)

The greatest proportion of deaths caused by alcohol was in the Americas and Europe.

HIV/Aids is the fourth biggest cause of mortality in the world: 28 million (70%) of the

population with HIV infection are concentrated in Africa.

These rapid demographic changes, characterising an ageing population, are adding to the heavy burden of chronic non-communicable diseases, infectious disease and HIV. These challenges will potentially worsen already marginalised regional and national health inequalities, and place added pressure on healthcare financing systems. Trends in healthcare expenditure which show a rise from 3% of world GDP in 1948 to 7.9% in 1997 are already indicative of such pressures (WHO, 2002). This dramatic increase in spending worldwide has prompted societies everywhere to look for health-financing arrangements which ensure that people are not denied access to care

because of affordability issues. In the next section I focus on these particular trends in South Africa and their implications for healthcare funding.

5.2.3 An overview of the South African healthcare status
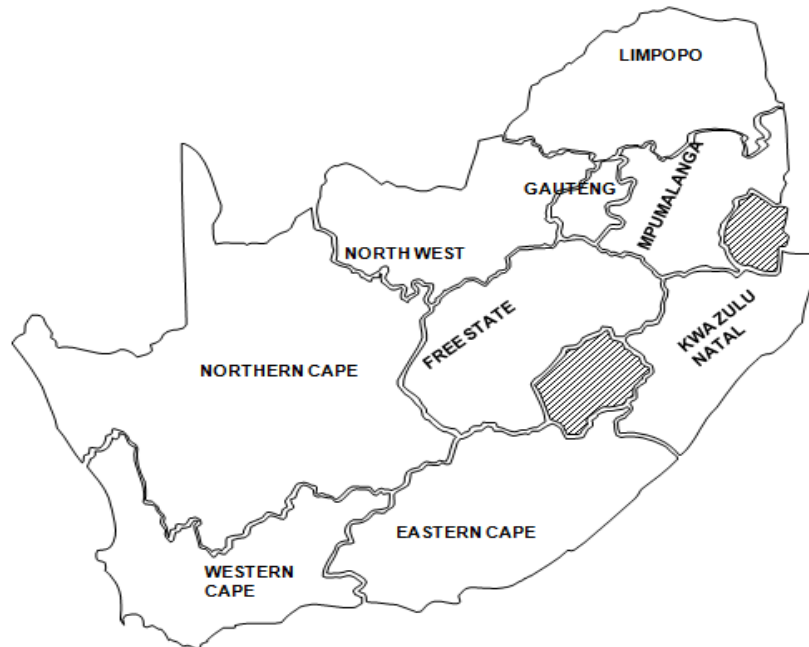
Map 5.1 Map of South Africa[1]

Source: Africa Institute

In some ways South Africa's healthcare position represents a microcosm of the globe. In other ways, the challenge is quite unique. During the era from 1948 to 1994, the apartheid system[2] affected people's health in many ways. Various social conditions contributed to ill health (Hassim, Heywood and Berger, 2007). For instance, the migrant labour system forced millions of black men to leave their homes and families, and work as cheap labour for industries and mines.

---

[1] South Africa is situated in the southern tip of Africa and has a land mass of more than 1.2 million square kilometres, which makes it bigger than France, Germany and Italy combined.

[2] 'Apartheid' was a term coined by the Afrikaaner National Party (NP) during their 1948 election campaign to promote a policy of 'separate development' in South Africa, along racial lines. Many experts suggest that the Nazi philosophy and propaganda at the time influenced the ideas of Afrikaner leaders in their political thinking. In effect it granted the white government state control over the African, Indian and coloured labour market. According to Terreblanche (2002), the 'practical' conception of apartheid was propogated by emerging Afrikaner entrepreneurs whose profits at the time depended on an abundant supply of cheap non-white labour. Over time the NP's segregationist legislation was not limited to discriminatory labour practices, but also to discriminatory social engineering programmes and draconian security systems used to perpetuate white social, political and economic domination. However, the concept of non-white inferiority in South Africa has a history of over 350 years, from periods of more blatant slavery during Dutch colonialism to the British imperialism of the 19th century, propaganda which was based on ideological strands from Social Darwinism. South Africa had its first democratic elections for all races in 1994.

These men lived in single-sex hostels in urban areas. These overcrowded living conditions in the hostels resulted in a number of illness epidemics as well as mental health concerns and problems of alcoholism. In addition, the apartheid government passed special laws and policies to enforce racial inequality in accessing healthcare services (Hassim *et al*, 2007). For instance, health departments were structured according to race. Furthermore, black students were prevented from training as doctors or dentists at white universities. Moreover, black doctors and nurses were not allowed to supervise white nurses, even if they were more qualified. Another issue was the unequal spending on health services. On average, spending according to race in 1987 was R137 for blacks compared with the R597 for whites (Hassim *et al.*, 2007). There was also a lack of significant challenge by civil society, including health professionals and workers, to the apartheid health system. Many white health workers either collaborated with the system or did not oppose it. For example, doctors even helped police cover up the murder of political activists like Steve Biko (Hassim *et al*, 2007). In 1998, the Truth and Reconciliation Commission (TRC) established that:

> *The health sector, through apathy, acceptance of the status quo and*
> *acts of omission, allowed the creation of an environment in which the*
> *health of millions of South Africans was neglected, even at times actively*
> *compromised, and in which the violations of moral and ethical codes*
> *of practice were frequent, facilitating violations of human rights.*

More than 10 years after the first democratic elections, the South African healthcare system, plagued by the profound legacy of apartheid, continues to be characterised by a number of distressing contradictions. On the one hand, it espouses an undisputedly high standard of modern medical education and a private sector that aspires to high standards of patient care (Benatar and Van Rensburg, 1995). According to a study by the United Nations, South Africa's private system ranked 39th out of 162 nations for technological innovation and achievement. On the other hand, the public sector is beleaguered by fragmented and duplicated services, wide disparities in health and access to healthcare, and the lack of preventive and rehabilitative services (Ntuli and Day, 2004).

In addition, South Africa's burden of disease is not shared equitably among the population. The pervasive social inequities that have their roots in the apartheid era are evident in morbidity and

mortality figures among racial groups. For example, the current life expectancy is 68 years for a white male and 46 years for a black male (Bradshaw and Nannan, 2004). Child mortality among black children has risen in the last decade, with HIV/Aids and low birth weight among the highest causes of death. Recent infant and under-five mortality rates have been climbing closer to 100 deaths per thousand live births, compared with the national target of 30 deaths per thousand live births (Ntuli and Day, 2004).

Furthermore, the devastating impact of the Aids epidemic has profoundly affected the health system (Carton, 2003). HIV prevalence among public sector antenatal clinic attendees has risen from around 7% in 1994 to 27% in 2002. Current estimates suggest that 1 in 9 South Africans are infected with HIV (Dorrington, Bradshaw and Budlender, 2002). The Aids epidemic has also fuelled a tuberculosis (TB) epidemic, resulting in increased deaths due to pneumonia, diarrhoea and other related conditions. A burden of disease study by Bradshaw and Nannan (2003) reports that HIV/Aids accounts for 38% of the country's total premature death burden.[3]

Diseases usually associated with poverty and underdevelopment accounted for 25% of the total burden, while non-communicable diseases accounted for 21% and injuries for 16% of the years of life lost (Bradshaw and Nannan, 2003). The study also confirmed an extremely high burden owing to violence, homicide and road traffic accidents. Consequently, the care needs of patients have placed severe strain on services, often disproportionately on some of the most disadvantaged facilities in the public sector. Moreover, health personnel in the public sector cite low levels of job satisfaction, poor working conditions, despondency in the face of the HIV epidemic, and unsatisfactory management, as well as inadequate salaries, as underlying their dissatisfaction with working in the public sector (Ntuli and Day, 2004).

Table 5.2 illustrates the inequalities in access to basic amenities resulting from apartheid that still persist today. Africans account for 79% of the population, whites for 9.6%, coloureds for 8.9% and Indians for 2.5%. Unemployment is much higher among Africans, with 51.1% unemployed compared with 27% of the coloured population group, 5.3% of the Indian population group and 1.4% of the white population group.

---

[3] Premature death burden refers to years of life lost based on a standard life expectancy for the age of death, with future years discounted at a percentage and age weighting formulated by the Global Burden of Disease Project study.

Table 5.2

Racial inequalities

Subject	Measures
Population**	
Africans	79%
Whites	9.6%
Coloureds	8.9%
Indians	2.5%
Unemployment	
African	
Coloured	51.1%
Indian	27%
White	5.3%
	1.4%
Formal households	
African	55%
White	95%
Television (TV)	
African	
Coloured	44%
Indian	>70%
White	>90%
	>90%
No toilet facility	
African	
Coloured	17%
Indian	6%
White	<1%
	<1%
Electricity	
African	
Indian and White	66%*
	>90%

Source: Health Status (Bradshaw and Nannan, 2004) MRC Burden of Disease Research Unit

* About two thirds of the African households use electricity for lighting, but only a third use it for heating. Nearly a third of African households use candles for lighting.

**2001 census estimates the population to be 44.8 million.

Moreover, over half (55%) of the African households live in formal housing compared with 95% of white households. Access to facilities differs sharply; less than 1% of white and Indian households have no toilet facility, while 16.9% of African and 6% of coloured households have no toilet facility. Barely 44% of African households have TV, while over 70% of coloured and over 90% of Indian and white households have TV. The energy source is also alarmingly different according to population group. About two thirds of the African households use

electricity for lighting, but only a third use it for cooking or heating, while nearly a third of African households still use candles for lighting. On the contrary, over 90% of the white and Indian population groups use electricity for all household purposes. While the majority of households (84.5%) have access to piped water, for most African households water is not immediately accessible in the home, but in a yard or a nearby public facility (Day and Gray, 2003). These factors appear to have a direct bearing on the health status of the different population groups.

Generally, population health in South Africa has declined rapidly in the last decade, as evidenced by a decreasing life expectancy, largely as a result of the rapid spread of the HIV epidemic. Furthermore, demographic differences and pervasive disparities in basic amenities between the population groups persist even after 10 years of democracy (Ntuli and Day, 2004). As we will observe in chapter 7, these broader contextual elements had an implicit or explicit influence on the SST implementation process. The ongoing inequity between those with access to private medical care and those dependent on the public sector represents another major challenge for the South African health system. We review these issues in the next section.

5.2.4 The South African private healthcare sector

In this section I briefly explore the structure of the South African private healthcare sector. I will also consider important trends and changes such as deregulation, competition, and healthcare reforms.

From a structural perspective, South Africa's current health service fits somewhere between the predominantly fee-for-service, privately run, managed-care healthcare system in the United States and the predominantly socialised, non-profit system in the United Kingdom (Benatar, 1993; Benatar and Van Rensburg, 1995). Three interrelated functions of financing are crucial for healthcare systems: revenue collection; pooling of resources; and purchasing of interventions (World Health Report, 2000). South Africa's spending on healthcare of 8.8% of Gross Domestic Product (GDP) via its institutions is unusually high by international standards (Doherty, Thomas, Muirhead and McIntyre, 2003). The system of private care in South Africa is characterised by providers of healthcare services on the one hand and funding institutions that finance the use of private care on the other. Private healthcare providers comprise a range of resources from

general practitioners, traditional healers, specialist physicians, neurologists, cardiologists, surgery, hospitals, and pharmacies to pathology services. These institutions are typically characterised as being independent of one another, competitive, and driven by the aim to make profits for their shareholders from the services they sell.
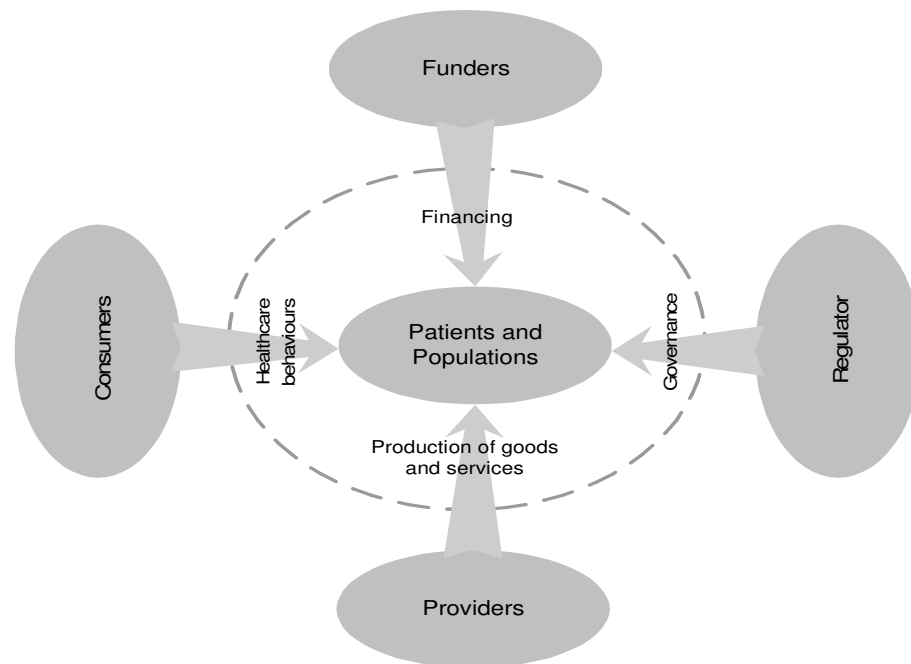
Figure 5.1 The multiple stakeholders in a health system
Source: Adapted from World Health Report (1999)

From a financing perspective, South Africa is following the trends of countries like the USA where a growing number of commercial health products are sold by insurance companies in the form of medical schemes. Within these private insurance schemes, revenue collection and pooling are typically integrated in one organisation and one purchasing process (World Health Report, 2000). The accumulation and management of revenues or pooling is done in such a way to ensure that the risk of having to pay for healthcare is borne by all the members of the pool and not by each contributor individually. In this way, pooling increases the likelihood that all patients can afford services. Recently, however, new designs in products have been passing the onus onto individuals so that they bear some of the risks from their savings or their own pockets.

Between 1974 and 1994, the membership of medical schemes in South Africa grew from 2.5 million to nearly 6 million people. This growth led to an increasing use of outsourced service providers, known as 'administrators', to manage the logistics of processing member claims and contributions. While by law medical schemes are supposed to be non-profit entities, meaning that any surplus generated that remains in the scheme belongs to members, their administration partners are profit-driven entities. The industry also includes an increasing number of managed-care organisations, some of which are owned by the very same administrators. Unlike their US counterparts, however, few schemes in South Africa have taken on the essential elements of provider networks and risk-sharing arrangements in their implementation of managed care. In the main, these organisations focus on utilisation rather than on risk-sharing and on advising schemes on hospital admission criteria, appropriate treatment protocols and cost containment (Doherty *et al*, 2003).

Linked to these trends were changes to the structure of medical schemes themselves. Historically, medical schemes were 'closed', meaning that membership was restricted to a certain group of people, for example employees in a particular firm. However, from around the mid '90s, there was growth in the number of 'open schemes', with membership open to anyone who could afford to pay the premiums. This rise in open schemes needed the support of another important group of intermediaries, known as brokers. During the '90s brokers were instrumental in selling medical insurance and increasing the membership of open medical schemes for which they earned a commission fee.

Despite these trends, for most of the period between 1995 and 2000, medical schemes suffered an overall deficit on operations (CMS, 2000; CMS, 2001). By the end of 2000 there were only 165 medical schemes, covering the lives of close to 7 million people (Doherty and McLeod, 2003). The nominal increase from 6 million in 1994 to 7 million in 2000 was as a result of previously disadvantaged groups, consisting mainly of non-white members, who began to experience relatively large increases in income levels (Doherty *et al*, 2003). This trend was aided by the increasing pressure exerted by the trade unions on employers to provide equitable cover for all employees. Despite the total number of schemes having declined steadily over the years, and schemes having larger risk pools, the management of risk and associated healthcare costs was still not contained. Instead, the industry was facing significant challenges, as demonstrated by the cumulative operating loss of more than R1 billion in 2000, despite a gross income from

contributions amounting to almost R31 billion (CMS, 2001). In addition, at the end of 2000, almost 50% of the members of registered medical schemes belonged to schemes with perilously low solvency levels: below 10%, against the minimum statutory solvency level of 25% (CMS, 2001).

The Board of Healthcare Funders[4] of Southern Africa attributed part of their poor performance to the Medical Schemes Act of 1998[5] (Kruger, 2002). The act, which was promulgated in 1999 and fully implemented in 2000, faced tremendous opposition from the insurance companies who were underwriting medical schemes. New legislation – such as the requirement of open enrolment, which compels medical schemes to accept all applicants; the introduction of community rating, which meant that schemes could not discriminate against a member on the basis of age or current health status; and the provision for prescribed minimum benefits – was among the major issues contributing to the controversy. As the industry tried to cope with these new regulatory demands, medical aid contributions rose for some schemes by as much as 18% in January 2000 compared with the 7% consumer price inflation (CPI) (Doherty and McLeod, 2003). Indeed, contribution increases in registered schemes had consistently outpaced inflation since the 1980s, sometimes even rising to double the consumer inflation rate in the 1990s. Notwithstanding, the rate of increase was more rapid in 2000 and 2001 than it had been for a long time, at around 11% (Doherty and McLeod, 2003).

Meanwhile, the regulator, in the form of the Council for Medical Schemes,[6] attributed the problems in the industry to deregulation and inadequate regulatory capacity,[7] fostered by the

---

[4] The Board of Healthcare Funders (BHF) is a non-profit organisation that represents more than 95% of all medical schemes in Southern Africa. It comprises members from South Africa, Namibia, Botswana and Zimbabwe. The BHF used to set tariffs for healthcare services as a 'guideline' to its seven million beneficiaries.

[5] The 1998 Medical Schemes Act re-regulated medical schemes in South Africa by re-introducing mechanisms that aimed to make health services more equitable.

[6] The Coucil for Medical Schemes (CMS) is a regulatory body that is tasked with monitoring and regulating medical schemes. This body aims to address historical imbalances in the sector and improve governance and financial oversight of medical schemes. It reports to the minister of health on issues related to medical schemes.

[7] Some experts believe that the apartheid government deliberately fostered an environment that is encouraging the growth of the private sector at the expense of the public sector. As a result, there was a steady drift of white South Africans to a private healthcare sector. Thus by 1960, 80% of whites had medical scheme cover. At the same time, 95% of black people were reliant on the public sector (Hassim *et al*, 2007). During the dying years of apartheid, the government introduced two significant acts to deregulate healthcare in the private sector. In 1989, an amendment to the regulations of the Medical Schemes Act removed the prohibition against risk-rating, thus increasing the contribution costs of vulnerable groups, mainly the sick and elderly. In 1994, a further amendment removed the statutory requirement of minimum benefits. As a result, medical schemes were allowed to provide healthcare services and run their own health facilities without state intervention. These reforms encouraged a drift towards selling health services for a profit. In this context, schemes took advantage of their new rights to risk rate and increasingly discriminated in admitting members. For example, by 1999, no open scheme enrolled individuals over the age of 55. Throughout the 1990s, most medical schemes also refused membership to people with HIV, or

previous apartheid government. According to the Council for Medical Schemes (CMS, 2003; CMS, 2004), some of the major problems in the industry included low solvency levels, high administration and other non-health expenditure, inexplicably high levels of net reinsurance losses, unmerited levels of medical scheme expenditure on health brokers, and excessive increases in expenditure on hospitals, medicines and specialists that had been outstripping inflation. Given these challenges between 2000 and 2003, there was a further 13% decline in the number of registered medical schemes, with 26 amalgamations taking place and 23 more medical schemes being wound up or dissolved (CMS, 2004). In 2004, the number of registered medical schemes decreased to 131 from 133. There were now only 47 open schemes and 84 restricted schemes (CMS, 2004). Although there are currently more restricted medical schemes than open schemes, the latter were generally far bigger, and accounted for more than two thirds of members covered by medical schemes. Despite the growing consolidation, the medical schemes industry as a whole was facing significant challenges.

Even with the growth in the number of open medical schemes, the insured population continued to shrink. The total number of beneficiaries belonging to medical schemes declined marginally from 2001 to 2002 by 0.89%. Private sector population shrunk from just under 17% of the population in 1997 to only 15.2% of the population in 2002 (Ntuli and Day, 2004). Meanwhile, the actual number of people dependent on the public sector had grown by 6.5 million people since 1995 (Ntuli and Day, 2004).

A number of experts claim that prohibitive costs made private healthcare too expensive and therefore inaccessible to most people (Fish and Ramjee, 2007). Furthermore, anti-competitive pricing of medicines, laboratory and specialist services, collusion among hospital groups, and the rising premiums of medical schemes, were labelled the main causes of high costs (Hassim *et al*, 2007). The distribution of private health facilities, mainly in urban and affluent areas, was deemed to be another deterrent to accessing private healthcare. This geographic inequality is reflected in a comparison between percentages of people covered by medical schemes in urban and rural provinces (CMS, 2004). For instance, according to the Council for Medical Schemes, 37.3% of medical scheme members reside in Gauteng (urban), whereas only 2.2% live in the

---

drastically limited their benefits. Thus the consequence of the 1989 and 1994 deregulation was the growing exclusion of the sick, the elderly, the poor, and people with chronic conditions. The further result was inadequate healthcare, a decrease in benefits, higher premiums, and an upward spiral in non-medical expenditure.

Northern Cape (rural). Moreover, approximately 50% of doctors were working to serve only 20% of the population covered by health insurance (CMS, 2004).

The high cost of private healthcare created further inequalities between members of the same medical schemes. For example, lower-income earners contribute disproportionately to medical schemes. This means that for a person with a low monthly income of R2 000 a month, up to 40% goes to paying the premiums of a medical scheme, while for a person earning over R15 000 a month, less than 10% goes to premium contributions (Fish and Ramjee, 2007).

Furthermore, non-health expenditure continued to grow. Between 1992 and 1998 there was a 243.5% real increase in non-health expenditure, largely in the form of broker and administrator costs. In 2003, the cost of scheme administration fees increased to R4.5 billion and this expenditure grew by 10.4% to R5.4 bn in the 2005 review period from R4.9 bn in 2004 (CMS, 2005). From 2000 to 2003, total non-healthcare costs increased by approximately 66.9%. The bulk of this expenditure was owing to administration costs, which represent close to 70% of non-health expenditure. In this period, administration costs increased by 80.4%. Managed-care costs escalated by 24.5%, broker fees by 152.6% and bad debts by 107.7% (CMS, 2003).

Like the medical schemes market, the administrator market is dominated by the top six administrators. While these six administrators controlled only 41% of the market in 1997, they controlled as much as 84% in 2000 (Doherty and McLeod, 2003). Despite the concentration of market share by large administrators, no real economies of scale appeared to be reflected in the level of administration costs. Another reason for the high administration costs is attributable to a system with multiple insurers, which is intrinsically costlier than a single-payer system. For insurers, it means multiple duplicative claims-processing facilities and smaller insured groups, both of which increase overheads. This kind of fragmentation also raises costs for providers, who must deal with multiple insurance products, forcing them to determine applicants' eligibility and to keep track of the various co-payments, scheme rules, referral networks, and approval requirements. Therefore, despite a number of regulations that are intended to decrease costs for consumers, increasing administration costs continue to have a negative impact on the cost of medical schemes and consequently have led to increased membership contributions.

Product innovations such as health savings accounts (HSAs), which medical schemes evangelise as providing consumers with affordable coverage for routine medical expenses, have also come under scrutiny. Critics say they have driven up medical inflation and failed to bring lower-income consumers into the private healthcare net (Fish and Ramjee, 2007). Many experts argue that because out-of-hospital costs are limited and patients run out of funds, they are now being hospitalised for treatments that doctors would previously have done in their practice rooms. Meanwhile, HSAs created a trend where many employers were moving towards a 'cost-to-company' payment model, leaving it up to employees to decide what proportion of their salary to use for medical insurance.

Further, there are surprisingly large numbers of corporate and individual brokers (roughly 6 000) whose business is enrolling clients with open medical schemes. The shift of members from restricted to open schemes in the mid '90s was largely a result of the actions of these brokers. For the regulator, the substantial expenditure on broker fees continued to be a matter of grave concern in an environment in which the overall number of beneficiaries remained fairly stable at around 7 million members. There was a dramatic increase in broker costs from R354 million in 2002 to R581 million in 2003 and R700 million in 2004. Including other acquisition costs, such as initiating, underwriting and selling a membership policy and distribution, total non-health expenditure increased by 22.0% to R939 m from R769 m. This was 14.6% of total non-health expenditure, up from 13.5% in 2004. These substantial increases in broker fees were clearly not being matched by increases in new members.

While non-health costs, including bad debts, were a source of major concern, healthcare costs, which accounts for some 80% of total expenditure of medical schemes, placed even greater pressure on member contributions. Private hospitals accounted for about two thirds of the above inflationary increases on expenditure by medical schemes. This is partly explained by the fact that the costs of healthcare provided by private hospitals have grown enormously. The Annual Report of the Council for Medical Schemes 2005 shows that since 1997, total hospital expenditure has increased by 103%. More specifically, ward fees increased by 85%, medicines increased by 122%, and theatre fees increased by 172% (CMS, 2006a; 2006b). The growth in the number of private hospitals also has serious consequences. It has created an overcapacity of hospital beds and a further concentration of healthcare services in urban areas.

Indeed, few schemes in South Africa have been able to engage in the managed-care arrangements with hospital groups that have been the trend in the US. The result is that mechanisms such as the use of provider networks and risk-sharing arrangements in an attempt to manage these costs are almost nonexistent. This may be as a result of difficulties experienced by the schemes to engage with these powerful provider groups (Doherty and McLeod, 2003). The problem is also attributed to the fee-for-service reimbursement system which has incentivised over-servicing and higher prices. In a well-defined managed-care relationship, a capitation arrangement typically exists where a provider is paid a fixed fee for the monthly care of a group of patients. The provider then has to cover the cost of care within the agreed amount, thus assuming some of the risk normally held by schemes alone. Some experts suggest that such arrangements will force providers to limit unnecessary treatment. Others express the downside of managed care, such as the tendency of providers to undersupply if they are not managed appropriately (Doherty and McLeod, 2003). Although entering into contractual preferred provider arrangements has the potential to result in significant cost control in the medium to long term, the power balance in the market appears to be making this an insurmountable task.

In 2000, the largest of the medical savings portion which equates to as much as 40% is being spent on medicines (Kruger, 2002). Medical specialists, GP consultations, and other health professionals follow this. Within the practitioner category, the growth in specialist expenditure has far exceeded that of other practitioners since the late 1980s. Thus, specialist expenditure accounted for 58% of practitioner expenditure in 2000, more than twice the percentage it was two decades earlier (Kruger, 2002). Evidently, less and less of each rand spent on healthcare benefits is going towards the more economical primary care services provided by general practitioners. This trend may also suggest that members are increasingly paying out of their pockets for general practitioner services.

Meanwhile, technology also began to drive medical bills ever higher. Not only was it expensive, given the poor performance of the South African rand against the US dollar and euro, but some of the expenditure was deemed unnecessary. Doctors were accused of frequently ordering tests without clinical evidence to justify their use. Worse still, it was claimed that many doctors were over-servicing patients. For example, many patients were admitted to hospitals overnight when they could readily be seen as outpatients (Doherty and McLeod, 2003). It was believed that

doctors were also over-prescribing by giving patients prescriptions for several expensive brand name drugs when one or two equally effective and cheaper generics would suffice.

Another reason for rising costs during the period between 1999 and 2003 was reinsurance (Doherty and McLeod, 2003). To protect themselves from unexpectedly high claims, some of the major medical schemes insured themselves with other insurers. While there are legitimate reasons for reinsuring, regulators accused the administrators of using this process as a loophole for channelling money out of a scheme to another company's shareholders as a profit-making arrangement, rather than back to the scheme's members.

In addition, the market for various health services such as hospital and pathology services is concentrated, which means that there is less competition among the providers of these services and therefore higher prices. There are also oligopoly situations where just a few suppliers of a product, such as modern hospitals, dominate an industry. For instance, three big cartels control more than 80% of the South African private hospital market. Despite regulations such as the Competition Act, No 89 of 1989, which seeks to restrict horizontal and vertical integration practices among dominant firms, imperfect competition among dominant firms is still adversely affecting other producers along the value chain and consumers. Nevertheless a Monitor Group (2004) global study revealed that South Africa is fourth best in the world in private hospital care, coming only after Switzerland, the Netherlands and Belgium.

At the same time, despite innovations in healthcare financing, South African households are spending more than double the amount on healthcare compared with 1999. In fact, consumers' out-of-pocket expenditure soared from a little over R1 billion in 1993 to R25 billion rand in 2006 (CMS, 2006a; 2006b). Out-of-pocket expenditure consists of healthcare expenditure by people who are not members of a medical scheme, either because they do not see the value of medical schemes or because they cannot afford the premiums. It also includes expenses which members of medical schemes have to pay because they have reached their scheme limits or because they are obliged to make co-payments in order to access certain services. Trends indicate that at the top-end segment of the market, consumers are opting out of medical scheme membership and are self-funding their medical expenses. On the other hand, lower-income earners, who constitute the bottom segment of the market, are facing a high barrier as a result of unaffordable premiums. The private sector blames these high barriers to entry on the regulator's insistence that the

medical schemes offer many prescribed minimum benefits. On the other hand, the regulator insists that the private sector is paying more attention to shareholder demands for profitability.

In 2004, the Competition Commission carried out an investigation into healthcare sector. It concluded that medical schemes, hospital groups and medical practitioners were acting in contravention of the Competition Act by colluding to fix prices of healthcare services. In addition to influencing the incessant prices hikes, these practices were preventing new service providers from entering the market as competitors.

As a result, a number of regulatory interventions have been tabled to make private healthcare more accessible to most South Africans. The introduction of prescribed minimum benefits (PMBs) meant that all schemes are legally required to provide their members with certain benefits. However, this could affect their financial sustainability and thus the access of poorer and sicker people to medical care if schemes go bankrupt. Therefore to achieve greater fairness among schemes after the implementation of the PMBs, the government intends to introduce a Risk Equalisation Fund (REF) in 2007 (CMS, 2006b). This means that schemes with a younger and 'healthier' membership profile with less risk will contribute to an REF that will be administered by the Council for Medical Schemes. The REF will pay amounts to those schemes with an older or more 'unhealthy' membership profile. The introduction of the REF means that all schemes will have to compete on delivery of quality care at lower costs, thus reducing their community rating, rather than relying on their membership profile.

Another policy issue that aims to create a better balance between the public and private sectors involves the reform of tax subsidies for medical aid. At present this formula covers a percentage of the cost, rather than setting a cap, which has effectively meant that the more expensive the schemes are, the higher the subsidies (CMS, 2006b). This means that better-paid employees are getting bigger subsidies, and it also means that the tax rebate is effectively paying for these services. In effect, it means that government is subsidising the use of private health services. This further benefits the private sector and disadvantages the public sector. The 2003 Taylor Committee of Inquiry reported that this tax subsidy to medical schemes equated to R7.8 billion or R1 000 a member. This is significantly more than the total annual health expenditure for a person in the public sector. In October 2005, the government published its proposals for tax

reforms. Among its recommendations was a proposal that the current system would be replaced with a monetary cap.

The regulator believed that by redistributing tax subsidies via the risk equalisation fund the medical scheme industry will collectively benefit from an increase in premium income. The extra R7,8 bn added to the industry's annual premium revenue of R37 bn in 2001 should drive down costs of a basic medical scheme package, enabling more affordable cover. The health department hoped that between 4 million and 8 million new people would be able to enter the private healthcare market.

Other avenues that were being explored by the regulator to redress obstacles to equity include the possibility of a low income medical scheme (LIMS). In 2005, a consultative process began with the private sector and government to investigate this possibility. The vision of the LIMS is to open schemes to lower-income earners who are either formally or informally employed (Hassim *et a.*, 2007). This legislation could have the effect of increasing the potential lives covered from only 7 million lives to 20 million. For instance, while only 2% of the entire population are covered by fully comprehensive medical insurance and a further 8% by limited co-payment and managed-care schemes, a further 30% of the employed population could be insured (Kruger, 2002). This has spurred intensified competition between medical schemes to grow their markets by capturing this lower-income market segment. It remains to be seen whether these schemes can succeed in these low-margin, high-volume segments.

There are also discussions in government circles in the longer term to introduce social health insurance (SHI). In this approach, all employers will be required to contribute to a national fund that will be used to pay for the cost of medical care for all employers. This could double the size of the privately insured market, creating large risk pools and providing new opportunities for cross-subsidisation. This would in turn improve negotiating power with providers. However, the trade unions are concerned that this will serve only to increase the cost of labour, thus resulting in further unemployment.

Other interventions are in the process of being been implemented (Doherty and McLeod, 2003; Gilbert, 2004). These are aimed at controlling the cost of medicines, such as generic prescribing and substitution, limiting dispensing by private practitioners, and creating a single 'exit' price for

pharmaceuticals as they leave the manufacturer. Meanwhile, government's attempt to limit the supply of private hospital services through a moratorium on new beds has failed to have an impact on the industry (Doherty and McLeod, 2003).

By 2002 the regulator was also seeking to curb the unseemly behaviour of brokers. According to Doherty and McLeod (2003), brokers were systematically raiding younger members from restricted schemes and were enrolling them into selected open schemes where they could earn high commission fees. Consequently, restricted schemes were left with a much older beneficiary profile. Recently, as the market became more static, brokerage fees continued to escalate as a result of brokers moving members from one open scheme to another. The latest regulation restricts brokerage fees to a small percentage, 3% of contributions paid, with a maximum monetary cap of R50 paid monthly (Doherty and McLeod, 2003). The purpose of this restriction is to remove incentives for brokers to 'churn' members and to sell only the premium priced products.

By 2004, draft regulations for a transparent pricing system and limits on distribution and dispensing fees, were implemented by the health minister. The market expected the prices of medicine to drop by between 40% and 70% in May. The regulations, part of the Medicines and Related Substances Control Amendment Act, and public comments were to be heard over the next three months. The draft required that yearly price rises would be limited to a ceiling that would be determined by the government, taking into account inflation and the exchange rate. If the savings were indeed passed on to consumers, it could save medical schemes R4 bn a year. Meanwhile, the pharmaceutical industry believed that the health department's plans would beggar the sector. Several multinational companies were threatening an exit strategy. There was concern that multinational firms might decide to scale back the products offered to the South African market, to the detriment of patient care.

Since the first democratic general election in 1994, the issue of economic empowerment had become increasingly more important against the racial imbalances and unequal opportunities of the past. For instance, in 1996 about 94% (8 401) of directors in South Africa were still white men and only 5% of companies listed in the JSE were black-controlled (Du Toit, 1998). By 2004, as a result of the Broad-Based Black Economic Empowerment (BEE) Act, the financial health insurance sector was seeking to transform its shareholding to black-owned equity. During

this period and under enormous political pressure, most of the medical schemes were looking to conclude BEE transactions with black-owned firms. Many critics argue that it was not the noble desire to overcome inequalities of the past in terms of racial discrimination, but government's decision to develop a single medical scheme for public service that catalysed the industry into implementing this policy. As the largest single employer, the public sector represented an enormous financial opportunity for these firms.

While the regulator seeks to apply interventions in the industry to promote equity, efficiency and the quality of care, healthcare insurance organisations are now being forced to seek more cost-effective and innovative approaches for delivering healthcare. This is a direct threat to their profitability and hence their survival. Furthermore, these regulatory trends are starting to shape the organisational strategies of the dominant healthcare insurance companies. Some of these trends are leading to strategies to devise products and services that would be more suitable to their markets. It has also started to become more apparent to these organisations that existing practices were incapable of tackling the root causes of cost escalation, namely the over-servicing and insidiously high prices charged by providers and the disproportionate contributions represented by non-healthcare expenditure. As a result, more elaborate product designs that were adaptable to a fairly unstable regulatory environment with short time-to-market started to become more vital to the strategies of healthcare insurance organisations.

5.2.5 Summary

A number of concerning issues face healthcare globally. It is clear that the startling inequities between the haves and have-nots is a major contributor to this problem. From a South African perspective, racial discrimination against non-whites affected their health in many ways. These included social conditions of ill health, the segregation of health services, the unequal spending on health services, and the failure of the professional medical bodies and civil society to challenge apartheid health. More than 10 years later, after its first democratic elections and despite a number of regulatory reforms, South Africa is still suffering from the serious inequities in health that took place systematically under apartheid laws and policies. Given a turbulent regulatory environment, healthcare insurance organisations are now being forced to seek more cost-effective and innovative approaches for delivering healthcare. A crucial component of resolving the healthcare crisis is via appropriate funding strategies. In the next section, I review

some of the major trends on the design, structure, and healthcare funding systems in the healthcare insurance market.

5.3 The changing healthcare insurance market

5.3.1 Introduction

In this section, I focus on the critical challenges facing the healthcare financial market worldwide. The explosion in scientific knowledge, its costly application to healthcare, the concomitant rise in inflation, and new models to alleviate problems in healthcare financing are explored.

5.3.2 Inflationary environment

It is widely acknowledged that healthcare costs globally are continually rising. Some of the key reasons for this phenomena are located in the area of outpatient and physician services. Rising costs are attributed mainly to expensive, high-tech treatments or diagnostic procedures such as radiation, MRIs, X-rays, lab tests, laparoscopic surgery, and body scans (Kronick, 2005; Cigich, 2002). For instance, while new imaging technologies are very effective at diagnosing and treating diseases at the cellular level, the short-term impact has been unfavourable. As an example, consider how the combination of positron emission tomography and CT is being used, not only to find cancers, but also to evaluate in microscopic detail which treatments are working. While these technological breakthroughs have long-term potential for saving lives and developing new drugs, they result in short-term problems such as creating higher demand for 'high tech' treatment from consumers, and healthcare providers who are obliged to offer these treatments to remain competitive (Benatar and Fleischer, 2003). This has a definite effect on healthcare inflation. In countries like the United States there is also the tendency for healthcare providers to conduct more expensive diagnostic tests for liability protection (Cigich, 2002).

Another force that has encouraged inflation is the managed-care backlash[8] (Gabel, 2003). Since 1996, health plans have retreated from closed or narrow provider networks, utilisation review programmes, provider risk-sharing and limited treatment options. In general, managed-care systems offered cheaper products because they restricted the choice of the patient. This movement away from a 'narrow' managed-care system set off a resurgence in healthcare inflation. This was largely brought about by a consumer backlash against network restrictions and utilisation review processes. Furthermore, providers added to the problem (Robinson, 2004) by consolidating their local markets, demanding rate increases, litigating over delays in payment and denials in authorisation. For example, in the US, premium increases accelerated every year after 1996 to a figure of 12.7% in 2002.

Other cost drivers include the hospital-building boom and widespread inefficiencies in the healthcare system. Furthermore, total spending on healthcare administration accounted for $261.2 billion, 21.6% US healthcare expenditure. According to Woolhandler, Campbell and Himmelstein (2003) private insurers have high overheads in most nations: 15.8% in Australia, 13.2% in Canada, 20.4% in Germany, and 10.4% in the Netherlands. Functions essential to private insurance, but absent in the public sector, such as underwriting, claims processing and marketing – accounting for about two thirds of private insurers' overheads – are also contributing to rising costs. As a result, private insurance companies are seeking novel ways to control the cost of financing healthcare.

5.3.3 Consumer-driven healthcare

In an attempt to curb the cost of inflation, a number of experts, including both regulators and private sector firms, are suggesting that there is currently an overemphasis on managing the supply side in the provision of healthcare services and are calling for a shift towards demand management (Labiris and Niakas, 2005). The managed-care approach alluded to earlier was a classic attempt to constrain the supply of care. However, many experts argue that the inability of the managed-care era to bring to an end the persistence of the fee-for-service reimbursement system continues to create incentives for providers to oversupply services. Patients have

---

[8] Managed care can be defined as the use of selective networks of contracted providers by health insurance companies which encompasses some means of incentivising members to use the networks, and some degree of risk sharing with those networks (Robinson, 2004).

relatively little information or personal authority with which to challenge care decisions made by providers, while the third-party payment system (whereby an administrator settles claims) shields both providers and patients from a full awareness of the cost of services (Doherty and McLeod, 2004).

Consequently, some healthcare insurance companies are now modelling their business on the so-called consumer-driven healthcare (CDH) concept. The healthcare companies frame this emerging concept as one in which the employees play a greater role in making decisions on their healthcare, have better access to information to make informed decisions, and share more in the costs (Cannon and Tanner, 2005). Consumer-driven healthcare plans seek to give employees greater options to choose their providers and their level of benefits. This translates to more point-of-service plans, higher deductibles and co-payments, and, in general, policies in which employees have more financial responsibility for their healthcare and employers and healthcare insurance firms provide better information to employees on their choices. The choice of the employees determines the premium for their health plan, and the employees disburse the difference, should that premium exceed the employers' contribution. A number of healthcare insurance firms are using Web-based tools to enable employees to make decisions such as selecting their network of physicians and hospitals, determining their co-payment levels, among other decision dimensions. From a financial perspective though, this increases an employee's share of total medical spending and resultant risk.

As alluded to earlier, the premise of consumer-driven healthcare is that employees will make more prudent decisions about their healthcare, which implies that patients are now making tactless ones. To support their argument, proponents of consumer-driven healthcare argue that most patients readily accept their doctor's treatment suggestions without question, including which drugs to take or specialists to see (Cannon and Tanner, 2005). They assert that when the patient's own money is at stake, he or she will be more likely to evaluate the best price for a routine visit or to forgo a costly brand name drug for a cheaper generic one. Advocates of consumer-driven healthcare are convinced this behaviour will in turn force pharmaceutical companies, hospitals and physicians to compete more aggressively on quality and cost. To support this claim, healthcare insurance companies are touting numerous studies which show that employees with greater access to information make more effective decisions about their

healthcare. Furthermore they claim that when given a greater burden of the cost, employee spending decreases.

On the other hand, detractors of the consumer-driven healthcare approach argue that employers are helping to drive this vision because it is in their interest to reduce their role in health insurance (Lee and Zappert, 2005). This is especially so, given the recent inflationary spiral in healthcare costs and their increasing liability in the way the medical benefit programs are currently being administered. Some employers have sought to cap the amount that they pay for health benefits through what is now termed a 'defined contribution'. 'Defined contribution' means that an employer pays a fixed amount towards each employee's health benefits, rather than whatever it costs for a fixed package of benefits (Robinson, 2005). In the view of critics, consumer-driven healthcare plans make it easier for employers to evade premium hikes and legal problems.

A majority of employers surveyed by Deloitte and Touche (2003) are seriously considering consumer-driven healthcare as a solution. Nearly 50% of companies surveyed in the US believed that consumer-driven healthcare would be part of their plans by 2003. Despite this, 63% of the companies surveyed stated that current consumer-driven plans have complex and confusing designs. In addition, another 41% reported that they are concerned that the plans benefit only healthy employees. Although plans are being put in place to move towards offering these healthcare solutions, some employers have voiced significant reservations, including a negative impact on healthcare inflation and whether less than healthy employees will be losers under the plan. Employers are particularly concerned about their poor employees who, when confronted with co-payments, may reduce the amount of services they demand, but not necessarily in a more rational way as proponents of consumer-driven healthcare assumes.

Apart from employers, private healthcare funders have enlisted the support of their governments in enlivening consumer-driven healthcare products. Consumer-driven healthcare firms have demonstrated their political power by the ease in which they have been able to frame their political-economic discourse in the media. Indeed, influenced by these health insurers, some governments are supporting further deregulation in social health insurance to encourage healthcare innovation towards this direction (Lehmann and Zweifel, 2004). In his State of the Union address President George Bush made the following remark:

We must work together to...control those costs and extend the benefits of

modern medicine throughout our country … Starting this year, millions of

Americans will be able to save money tax-free for their medical expenses,

in a health savings account. State of the Union address, January 2004.

Recent trends in particular show a shift in government policy towards further deregulation. For example, in the US the health savings account (HSA) remains a cornerstone of the Bush administration's agenda to control health insurance costs. In December 2003, the Medicare Reform Law in the US allowed consumers to set up tax-free savings accounts to cover their out-of-pocket medical expenses. Here again, the idea is to give consumers a stronger incentive to seek the most cost-effective care.

Many healthcare practitioners who are frustrated with the restrictions and complexity of managed care are supporting consumer-driven health plans (CDHP). They hope that these plans will reduce the intrusion of health insurance into the practice of medicine and restore independence to the doctor-patient relationship.

One of the innovations in a consumer-driven healthcare plan is the health savings account (HSA) mechanism. According to Robinson (2005), although HSAs impose greater cost-sharing on members, they permit broader choices than the health maintenance organisation (HMO) plans of the managed-care era. Healthcare savings accounts works very much like a bank account which the consumer uses to pay for non-hospital expenses.[9] Typically members pay a monthly contribution towards their HSA. This means that the larger the HSA, the higher the member's contribution, although the maximum amount for the HSA is normally limited by the employer or regulator. Unlike a bank account, some non-hospital expenses are subject to limits, such as spectacles or prescribed medicines on which consumers can only claim up to a certain amount a year. In addition, chronic medication is subject to limits based on an insurance scheme's rules. In most cases, any money accumulated in the HSA is carried over to the next year. Similar to banks, HSAs with a positive balance may earn interest, while HSAs with a negative balance pay interest. Furthermore, unspent balances belong to the account holder instead of the employer, and can be moved when the member leaves his or her job.

---

[9] Non-hospital expenses refer to costs when a patient is not admitted to hospitals, such as consultations with healthcare practitioners and approved medication.

HSAs challenge the notion of the 'use it or lose it' principle of insured services, where the financial benefit goes to sick members of a pool who incur claims in excess of their premium payments. In the HSA approach, the funds remain the asset of each member. By reformulating the principle to a 'use it or save it' principle, it is assumed that the member's responsibility for costs incurred in own care increases, but decreases his or her responsibility for costs incurred for the care of other members within the pool. For instance, consumers are being incentivised to request doctors to charge medical aid rates rather than private rates. The differential between private rates which are normally in excess of medical aid rates is the responsibility of the consumer. When consumers run out of funds from their savings account, they will have to pay out of their pockets to providers. In addition to contributing to an HSA for routine care, consumers purchase a catastrophic health insurance policy to cover the costs of major illnesses.

HSAs thus shift the locus of rights and responsibilities for financing day-to-day healthcare expenses from governments and employers to individual consumers (Cannon and Tanner, 2006). In doing so, the language of individual ownership weakens society's sense of collective responsibility for its most vulnerable members, and instead emphasises the importance of individual effort in generating the economic resources that underlie any system of care (Robinson, 2003). According to Robinson (2003), HSAs moves us closer to the notion of a personalised and privatised healthcare system. From a macro perspective, according to this theory, if patients decide whether the care they are receiving is worth their money, the invisible hand of the market will determine the proper level of resources and the efficient allocation of them.

However, critics argue that HSAs may have a negative impact on the outcome of the healthcare of financially vulnerable patients because these patients will choose not to seek care or not to adhere to prescribed medication regimens. Furthermore, other critics are challenging the notion that consumers are capable of assuming most of the responsibility for making decisions about their own healthcare. For example, Lee and Zapert (2005) found that the likelihood of patient care worsened as patients realised they were paying for effective preventative care such as mammograms and Pap smears. To counter this behaviour, a number of CDH firms are offering comprehensive cover for preventative services to motivate plan members to schedule routine care and thus avoid costly treatment in the future (Sharon and Donahue, 2006). Moreover, others have noted that the HSA approach tends to be more attractive to younger, healthier individuals

and are concerned about the impact of these members opting out of conventional insurance pools.

Detractors continue to argue that despite these innovations, healthcare costs will continue to rise. Opponents of consumer-driven healthcare are providing recent evidence that healthcare inflation persists, even though the much-vaunted consumer-driven healthcare movement is supposed to be driving down costs (Chiappetta, 2005). Some experts are arguing that while there is merit in making people aware of the cost of healthcare procedures, there are limits to the extent that cost and information sharing can check overall expenditures. In the first place, the vast majority of spending is on a relatively small subset of the population, such as those with chronic illnesses whose costs far outstrip the out-of-pocket limits in consumer-driven plans. Second, the march of new technology, including hospital construction, continues to experience significant growth. While many critics acknowledge that market forces are effective in improving consumer welfare in markets for many types of products (such as household appliances and personal computers), they maintain that these same market forces will harm consumers of healthcare because healthcare is different. Low-income people in particular may find it difficult to pay for added out-of-pocket costs necessary under CDHP. Third, some critics argue that market competition creates waste in the form of duplication of facilities and diverts resources from patient care to administrative costs and profits. For these reasons, a number of experts are proposing instead a national commitment that addresses systemic changes in disease management, preventive care, clinical research, and information systems, to help manage healthcare more efficiently. They suggest that conventional sources of finance or novel models alone cannot solve the healthcare problem (Afsaw and von Braun, 2005). The more cynical critics are simply viewing these new health plans as another mechanism for giving tax breaks to wealthy citizens.

However, the HSA mechanism is not the only innovation inscribed in the design of CDH products to manage costs. In the next section, we review some of the other key steps healthcare insurance firms are taking, such as integrating elements of wellness into their design of CDH products and services.

5.3.4 Wellness programs

Most CDH providers have also integrated elements of wellness into their product design and are making these available through Web-based portals and traditional channels. Examples of these wellness approaches include online health risk appraisals, preventative care guidelines, lifestyle modification modules and disease management programs. Some organisations have ventured further down this path and are providing comprehensive wellness programs that include on-site fitness centres, managed clinics or pharmacies, intensive health coaching and disease management.

This novel approach has been taken by consumer-driven healthcare insurance companies to curb costs by focusing on wellness rather than on illness and the resultant design of their products on the basis that 'prevention is better than cure'. These typically take the form of proactive health assessment processes where details of the patient's lifestyle habits – such as drinking, smoking, stress levels, nutrition, sleep patterns, and physical exercise – are recorded and analysed online or by a qualified health practitioner. Once analysed, a confidential and comprehensive report is typically sent back to the patient. The report is meant to alert the patient to looming health problems, from predictable problems such as being overweight to the unpredictable risk of developing diabetes. The report would analyse the patient's lifestyle and provide guidance or advice on how to make the patient healthier. Patients would then be entitled to consult specific experts such as a dietician to help modify eating habits, a biokineticist to recommend an exercise regimen, or a psychologist for stress counselling. Some healthcare insurance companies are offering incentives such as discounted gymnasium fees for maintaining a healthy lifestyle or an elaborate points scheme to reward appropriate health-related behaviours. While components of CDH products such as the HSA make consumers accountable for the financial consequences of their decisions, the wellness component seeks to reward them for appropriate behaviours.

Although the predominant view held by clinicians and healthcare decision makers is that prevention inhibits rising healthcare costs, recent empirical evidence suggests that most clinical preventive services may actually increase societal costs (Gandjour and Lauterbach, 2005). Mathis (2005) succinctly argues why healthcare costs will continue to rise, and preventive care and better treatments will be major causes of that rise rather than the panacea:

The medical establishment cured the infectious diseases that took the young lives of so many: yellow fever, typhoid, typhus, smal pox, diphtheria and whooping cough. Those who would have died from those diseases were spared to live on into middle age to develop heart disease. Our physicians and hospitals have made remarkable, but costly, strides in combating that killer, sparing many to live on to develop cancer. Now many cancers can be effectively but expensively treated or cured, thus sparing many to live on to develop Alzheimer's disease at great continuing expense to our healthcare system. All of the buzzwords – prevention, early detection, screening, chronic case-management – are nothing more than prescriptions for increasing the overall cost of healthcare.

In any event, the WHO report mentioned in section 5.2 already states that an enormous part of the crisis in healthcare can be attributed to individual behaviours related to risk, such as food intake, smoking and sexual behaviour. It may be simplistic to assume that strong monetary incentives for prevention will resolve the complex problem of consumer health. Despite the mixed views about the efficacy of wellness programs in the literature, the Internet is deemed an important ally in this process. The role of the Internet in consumer-driven healthcare is explored in the next section.

5.3.5 Summary

To summarise, advances in medical care and the ability to improve the duration and quality of life, combined with the growing expectations of both doctors and patients that new modalities of treatment should be implemented in everyday practice, and high administration costs of healthcare insurance firms are among the major reasons for modern healthcare services becoming so expensive. Many frustrated employers and struggling health insurers are adamant that patients should assume greater responsibility for containing healthcare spending and absorbing cost increases. While there are mixed views on the likely outcome of consumer-driven healthcare as the panacea, this steadily growing model which seeks to manipulate the behaviour of the demand and supply side of the market is bound to change consumer behaviour and alter traditional patient-provider relationships. The well-publicised shortcomings of managed care have prompted employers and insurers to promote the attributes of consumer-driven healthcare.

They paint a picture of patient empowerment, highlighted by greater flexibility, wider options and more authority in decision making. In return, patients are expected to make better and more cost-effective decisions. Despite a broad range of challenges, the Web is seen as a suitable tool to help consumers make these decisions.

In the next section, I take a closer look at the ICT sector in South Africa and its ability to enable the implementation of the Web for these kinds of self-service applications.

5.4 The ICT sector and the healthcare insurance industry

5.4.1 Introduction

In earlier sections, I established how, over a period of 60 years, the apartheid policy had skewed the development of healthcare in South Africa along racial lines. In section 5.3 I also recognised that ICT is central to an information intensive industry such as healthcare insurance. In this section I will establish that ICT and the Internet in particular are at the heart of the new consumer-driven healthcare approach to healthcare financing. I will also show how the same legacy of apartheid applies to access to basic ICT infrastructure. I also sketch the broad range of challenges facing organisations wishing to implement financial services via the Internet in the South African context.

5.4.2 ICT in the financial services sector

ICT is of paramount importance to the healthcare insurance market. It plays a crucial role in the sales, underwriting, and claims procedures, procedures that require the collection, manipulation, storage and retrieval of massive amounts of data. Furthermore ICT solutions are being extended to channels such as 'high-tech' call centres to make it easier for members, medical practitioners, hospitals and pharmacies to communicate with the firm. With rising medical costs, private insurance companies are looking to reduce their administration costs and improve their service levels. The Internet in particular presents new opportunities to distribute products and services more effectively and efficiently. During the dot.com zenith, the Internet held aloft what was in hindsight the somewhat exaggerated promise of a revolution in commerce. In this section, I trace

the status of the ICT sector in South Africa over the period of the case study and how it influenced the use of the Internet over the same period.

During the period from 2000 onwards, the telecommunications sector in South Africa continued to be characterised by buoyant growth. However, this growth was accompanied by relatively high retail prices, super-profits, job losses, licensing delays and minimal new foreign investment in the sector. The overall growth of the sector in 2003 rose to a total revenue of R74 billion and an increased GDP contribution of around 6%. Furthermore, relative to other African countries, South Africa had a fairly advanced telecommunications infrastructure and boasted the highest teledensity in Africa (Information Technologies Group, 2000).

While there is currently a lack of publicly available data on users of ICT in South Africa and it is even more difficult to ascertain how well South African enterprises are using their ICT facilities, a study completed in 1996 attempted to assess the effectiveness of ICT use in the four major economic sectors (Young and Ridley, 2003; Hodge and Miller, 1996). At the time, the financial services and retailing sectors appeared to be making the most effective use of ICT and in several cases could be considered world-class users. For instance, all the major banks offered a full range of on-line banking facilities. In addition, the four largest banks were placed among the top 500 largest banks in the world and the IT support in the country for financial services was regarded as comparable with or better than that of many developed countries. Furthermore, the local market for call-centre products continued to grow rapidly.

Table 5.3

Examples of Web-style commerce

Segment	Web-style commerce
Financial sector examples	Web-based electronic banking
	Online bill presentation and payment
	Asset financing
	Mortgage applications
	Online share dealing
	Unit trusts
	Insurance product sales
	Insurance claims processing

Source: BMI-TechKnowledge (2004)

The high investment in ICT by the South African financial services sector represents an interesting dichotomy. For instance, the manufacturing and mining sector, which remains the lifeblood of the South African economy, was clearly behind. A reliable estimate put the IT spend of the manufacturing sector at 1% of turnover, compared with a figure of 4% in Europe at the time. Nevertheless, all of South Africa's major financial institutions were using countrywide electronic networks. These institutions were also engaging in significant levels of e-commerce in various forms (see table 5.3). While some of these observations provide an optimistic view, these performances mask the serious underlying problems in the sector. I will review this position by paying particular attention to the Internet.

5.4.3 The Internet and consumer-driven healthcare

A number of researchers have found that it is becoming a common trend to look to the Internet for information about personal health issues (Larkin, 2001). These range from self-maintenance and personal health insurance to medical information that will enable one to understand one's own symptoms (Lang and Collen, 2005). According to Rainie (in Lang and Collen, 2005:43), a member of the health sites panel that is discussing the building of trust on the Web:

> *... about 80% of U.S. Internet users have gone online to get health diagnoses or to*
> *find out what's wrong with them; to get second opinions ... about what they have and*
> *how to treat it; to check out doctors and check out hospitals they might be going to;*
> *to go to support groups.*

The use of the Internet by consumers for healthcare services is increasing. Therefore the role of the Internet as a potential source of patient information and support can make a significant contribution to the delivery of healthcare services (Laing, Hogg and Winkleman, 2004). This behaviour is especially relevant for consumer-driven healthcare firms, where the process of engaging their members in managing costs is central to their philosophy of consumer-driven or consumer-centric healthcare. As a result, a number of CDH firms are implementing online self-service applications to reduce the cost of servicing customers and to improve customer relationships (Kolsky and Bivin, 2001). These online self-service applications provide

consumers with interfaces to access a CDH firm's data, enabling consumers to serve themselves with the information they need or the service they require.

In the main, self-service capabilities through a Web-enabled portal are allowing employees to monitor their claim activities, access provider information, use tools for self-care and personal health appraisal. Some CDH firms also offer online help sources and access to personal assistance in the form of customer-service representatives and nurse lines (Abbot and Feltman, 2002).

Communication, education and easy access to the right information through the use of Web-enabled technology are critical elements of the processes relevant to CDH firms (Abbot and Feltman, 2002). Proponents of CDH believe that consumers feel empowered because they have direct inputs into decisions about their healthcare, specifically with the online knowledge and tools they need to make those decisions. Consequently, the traditional process of health education has taken on new dimensions with the advent of the Internet and is being introduced in many e-health initiatives (Lang and Collen, 2005). For example, obesity is a growing problem in most developed countries, and the number of people using the Internet to learn more about weight reduction or maintenance options is also growing (Larkin, 2001; Tate, Wing and Winett, 2001). Furthermore, some CDH firms have started to offer tools for comparing hospitals and physicians on the Internet such as physician scorecards.

However, the efficacy of the Web for decision making depends on the transparency and reliability of provider cost data. It is currently very difficult for a consumer to shop for services without information that reflects the cost and quality of those services. Furthermore, it is possible that potential risks can emanate from the use of irrelevant or inaccurate information (Beaulieu, 2002; Crocco, Villasis-Keever, and Jadad, 2002). To empower consumers, one of the challenges is to provide reliable and comparable information (Scheffler and Felton, 2006, Eysenbach, Powell, Kuss, and Sa, 2002). While it is not clear how prevalent Internet use for healthcare really is, or what impact it has on healthcare utilisation (Baker, Wagner, Singer and Bundorf, 2003), proponents of consumer-driven healthcare believe that health insurance companies that have already invested heavily in e-commerce will be in the best position to service this segment.

According to some experts, the key to serving a consumer-driven healthcare plan is the ability to provide personalised, high-value service to the member at any time of the day or night through multiple communication channels (Abbot and Feltman, 2002). Firms should be able to authenticate members over the Internet and present them with a complete transactional history. Furthermore, they should be able to offer personalised educational information in the context of actual transactions processed. E-commerce capabilities should integrate these with customer relationship management tools that allow the delivery of targeted, consumer-focused educational components such as disease management or prevention and early intervention in chronic diseases.

Aetna, a leading CDH provider in the United States claims that its members were more aware of the actual costs associated with their care owing to increased usage of Aetna's Internet portal tools. Members were more likely to view their explanation of benefits (EOB) online, check claim status, use the provider search function and review their HSA balance (Sharon and Donahue, 2006). However, survey data collected by Harris Interactive provide little evidence of an emergence of the 'market-driven healthcare' culture that is critical to the success of CDH plans. On the contrary, they believe there is little evidence that there is a culture in which consumers actually use data on costs and quality to choose their hospitals and doctors. A nationally representative telephone survey of 1 000 adults conducted in 2001 and 2005 found low rates of use of information on the quality of hospitals, health plans, and physicians, and no sign of an increase over this period.

While enrolment in CDH plans is still low, it is expected to increase rapidly in coming years. Some insurers are optimistic that information technology (IT), specifically the Internet, will also spur overall growth of the CDH market. Although IT has been associated among some experts with reduced administrative costs, there is little evidence that it is effective as a sales channel and is thereby contributing to greater enrolment of consumers. While the Internet is altering traditional distribution channels, agents and brokers appear to remain the dominant channel for healthcare insurance products and are perceived to be the core of success in this market. Proponents of conventional channels to healthcare insurance sales argue that brokers and agents are able to interpret the market to potential consumers and thus affect what products they demand (Buntin, Marquis and Yegian, 2004; Schultze and Orlikowski, 2004). Despite attempts

by healthcare insurance companies to the contrary, it appears that consumers are more receptive to the idea of the Internet as a service channel than as a sales and marketing channel.

Another key challenge is catering for the broad array of users with very different information needs (Lang and Collen, 2005; Adam and De Bont, 2003). Furthermore, it is unclear whether providing information to patients will transform them into consumers that are actually able to improve the outcomes of their care. Nevertheless, a Forrester study (see Johnston, Brown, Molvar and Twist, 2001), found that most healthcare insurance organisations are implementing online self-service portals to help them achieve critical business goals such as improved customer service, greater member satisfaction and reduced administration costs.

5.4.4 The status of e-commerce

The Internet is presenting new opportunities to the healthcare insurance industry to distribute products and service more efficiently. During the dotcom boom at the end of the '90s the Internet presented opportunities, such as operating cost reductions, and threats, such as the entrance of new competitors. These opportunities and threats drove many healthcare insurance firms to regard the Internet as a new distribution or service channel. However, according to Goldstuck (2004), at the end of 2002 only 3.1 million South Africans had access to the Internet (see table 5.4). This translates to only 1 in every 13 South Africans. This was marginally up from 1 in 15 at the end of 2001 compared with 3 out of 4 in the United States during the same period. Growth in Internet access showed a dramatic slow-down between 2000 and 2003. Growth in 2002 was around 7%, the slowest since the Internet became available to the public in 1993, and the first time it had been below 20%. Growth in 2003 was set to be only 6%, with 3.28 million South Africans expected to have access to the Internet. However, specific developments were expected to boost growth in 2004, such as the roll-out of services by the Second Network Operator (SNO), which had finally been granted a licence to operate, the roll-out of high-speed or broadband wireless, and the healthy rand-dollar exchange rate, which had dramatically brought down the cost of equipment for rolling out infrastructure.

On a more positive note, the mobile subscriber base in South Africa since 2000 grew by 20% over a five-year period. By March 2004, the estimated size of the active mobile subscriber market was around 14.5 million subscribers (Goldstuck, 2005). The key reasons for the rapid

growth are the limited roll-out of fixed lines and the high cost of fixed-line rentals offered by the Telkom monopoly. Nevertheless, South Africa moved out of the top 20 most connected countries in the world, falling to 34th position (see table 5.5). This dwindling performance was supported by a Global Networked Readiness Study, which assesses the degree to which a nation is *prepared* to participate in the networked world; South Africa was ranked a paltry 40th out of 75 nations.

Table 5.4

Internet access in South Africa, 2004

	2000	2001	2002	2003	2004*
Dial-up	782,000	960, 000	1008,000	1038,000	1,100,000
Corporate	1,274,000	1,501,000	1,640,000	1,775,000	490,000
Academic	360,000	425,000	450,000	470,000	1,890,000
Total	2,416,000	2,886,000	3,098,000	3,283,000	3,480,000

2004 – Estimated

Source: The Goldstuck Report: Internet Access in South Africa, 2004 (Goldstuck, 2005).

Note: Internet access is growing, but mostly among elite users, who are generally wealthier White South Africans.

Table 5.5

South Africa's Internet connectivity ranking

Date of Survey	Number of Internet	World ranking
January 1999	144,445	21st
January 2000	167,635	25th
January 2002	238,462	28th
January 2003	298,035	34th

Source: DNS Survey

Note: July 1996 14th

While this drop in position was largely attributed to the more industrialised and richer nations being late adopters of the Internet, many local barriers were slowing the growth of the Internet.

5.4.5 A review of the key barriers to e-commerce

In considering the main barriers to e-commerce, one of the key aspects is the status of telecommunications industry. During the period of political transition in the 1990s, telecommunications in South Africa was governed by the 1958 Post Office Act under which a state-owned company, Telkom, functioned as a telephony infrastructure and services monopoly. Despite the restructuring and corporatisation of the telecommunications sector, the monopoly was re-enforced by a parliamentary act in 1997 that gave Telkom monopoly over South Africa's telecommunications infrastructure until 2002. Since there was no direct competition, market effects prevalent in first-world countries did not play a significant role in bringing down the cost of telecommunication. This discrepancy is demonstrated by the fact that 30 million users of fixed line services were obligated to only one telecommunication company. Therefore, regulations allowing Telkom to maintain its monopolistic position until 2003 caused telecommunication costs to rise even further in a non-competitive market.

Allied to this barrier were the high equipment and Internet service provider (ISP) costs which contributed to low levels of PC penetration. South Africa imports most of its computer hardware from abroad. According to Goble (2004), the initial set-up cost of hardware of this type may be as high as $1500,00. Consequently, hardware costs were unaffordable for the average South African. Another survey conducted by Milne (2001) found that the cost of connecting to an ISP was significantly higher in South Africa than in other developing countries. Goble's (2004) study also compared the US's cost of broadband services with the equivalent bandwidth in South Africa. At the time, the comparative costs for 512 kbps download speed and 128 kbps upload speed from AT&T was $200 per installation and $79 per month, while the equivalent bandwidth in South Africa would cost a R1 275 installation fee plus a monthly rental fee of R1 500. In other words, the relative cost of bandwidth was reported to be 42 times more expensive in South Africa compared with the United States. The year 2003 saw the birth of broadband in South Africa, with the roll-out of the first commercial ADSL service by Telkom, but growth in 2005 was expected to be no greater, with broadband services resulting mostly in upgrades of current dial-up connected users rather than new users.

In addition to purchasing hardware, hardware must be configured and set up correctly, which requires some basic knowledge and training. A rudimentary knowledge of computers is required

before a user is able to connect to the Internet and browse this resource. According to a comparative study by the World Bank (2001), the average illiteracy rate in South Africa in 1999 was 15% of the total population. According to a study by Hodge and Miller (1996), South Africa has a computer literacy percentage of 7.7% of the population. While this may have changed to a degree, it serves to highlight the problem that the majority of South Africans are unable to participate in e-commerce because they do not possess sufficient skills to do so. The World Competitiveness Report (Institute for Management Development, 1993) also showed that South Africa rates behind India, Singapore and Chile. While this report may be subjective, it gives a fair reflection of South Africa's position in relation to the rest of the world. The comparison clearly demonstrates that access to the Internet in South Africa is further restricted because of the lack of basic computer literacy skills among its citizens.

Table 5.6

Comparison of computer literacy rating

Country	Rating
Singapore	7.6
Japan	7.3
Chile	5.9
USA	5.7
Malaysia	5.3
UK	4.9
India	3.4
South Africa	3.1
Brazil	3.1
Pakistan	2.1

Source: World Competitiveness Report (Institute for Management Development, 1993)

The combination of adult illiteracy and computer illiteracy significantly raises the barriers of entry to the Internet and excludes the majority of the South African population. The relatively high illiteracy level implies that a large portion of the population must first learn to read and write before attempting to become computer literate.

While training and literacy are important, language poses an equally high barrier to entry to the Internet. A widespread lack of English as a medium of communication prevents the majority of

the South African people from participating in the Internet. Even today, English dominates the world of computing and information technology. After all, most Web pages are written in English. Therefore, people who were unable to read, write or speak English are virtually excluded from communicating with this medium. In South Africa, the most commonly spoken language is Zulu, spoken by 22.4% of the population, followed by Xhosa, 17.5% ,then Afrikaans, 15.1% and finally English 9.1% (see table 5.7). While all schools in South Africa require English to be taught at least as a second language, Hodge and Miller (1996) maintain that this was insufficient to meet the demands placed on acquiring and using knowledge on the Internet. Clearly then, apartheid's separate development policies had left the majority of the population at a distinct disadvantage.

Table 5.7

Distribution of home languages in South Africa

Home language	Percentage
Afrikaans	15.1
English	9.1
Afrikaans/English	0.2
isiNdebele	1.5
Sepedi	9.8
Sesotho	6.9
siSwati	2.6
Xitsonga	4.2
Setswana	7.2
Tshivenda	1.7
isiXhosa	7.5
isiZulu	22.4
Other	1.8

Source: Central Statistical Service(2004)

Compounding these barriers were the slow and unreliable connections made to ISPs because of South Africa's poor infrastructure, particularly in rural areas. These factors have limited the entry into this market only to those who are able to afford it. Another barrier is that the speed at which South Africa connects to peering points on other continents is dictated by the poor infrastructure the monopoly provider has in place. The demand for international bandwidth is high, as most sites are hosted in the USA and Europe. The baud rate of analogue telephone lines determines the bandwidth that the majority of South African Internet home-users have. It is seldom more

than 56 kb/s and in reality is between 19.6 and 28.8 kb/s. This is caused by the poor quality of telephone lines and the high cost of ISDN. European and American users have a higher quality connection, because they have better-quality telephone lines. These users also have access to higher connection speeds at lower cost. ADSL and cable connections are the norm in Europe and America, while in South Africa analogue telephones are currently the primary means available to households to connect to the Internet.

In South Africa, Telkom began offering home users speeds of up to 128 kbs using ISDN lines and modems. The majority of South Africans cannot afford ISDN, as this is too expensive. Because the connection speed to an ISP is, on average, much slower in South Africa than in developed countries, the time expended in accessing resources on the World-Wide Web is therefore much higher in South Africa than in Europe or America.

Moreover, supporting the diffusion process and fostering growth of ICT requires a skilled ICT labour pool. This is especially relevant since, like many developing countries, South Africa suffers a brain drain, whereby several hundred ICT-skilled professionals are leaving the country every year. The inequities created in South Africa due to apartheid are also highly visible in the education system that is meant to supply these skills. The pass rate with university exemption fluctuated over the period between 1992 and 1996, and the 1997 pass rate of 12.4 % was actually lower than in 1993 (15.8%). Fewer than half of pupils (44.2%) who obtained a Senior Certificate passed with mathematics and physical science. A mathematics and physical science pass on higher grade are usually the minimum requirements for persons who want to obtain a university qualification in one of the key skill areas in IT. Technikons usually demand at least a mathematics pass on standard level for IT-related qualifications (HSRC Telecommunications Study, 1998:26). Therefore the HSRC survey concluded that supply will not be able to meet demand for computer science occupations and an undersupply of between 54% and 60% was estimated for the dotcom period between 1998 and 2003.

Given these barriers in ICT in South Africa and e-commerce in particular, the exclusion of poor communities by healthcare insurance institutions became a common phenomenon. At the same time, the urban elite or the more affluent and large businesses continued to dominate in their interactions with these institutions. Many healthcare insurance institutions were able to justify exclusive financial service delivery on the grounds that the current ICT infrastructure made it

costly to service the poor communities. While economically justified, the continued 'financial services divide' was becoming socially and politically unacceptable (Maumbe, 2006). In any event, the ultimate success of South Africa's financial services sector may depend on how these institutions adapt ICT to penetrate previously neglected and untapped markets in poor communities, and markets created by the burgeoning new black middle class and elite black community.

5.4.6 Summary

To sum up, the Internet remains inaccessible to the majority of South Africans. Policies of the past with regards to the major elements of the country's ICT infrastructure and education have effectively excluded the majority of the population from the economic and social benefits promised by e-commerce. Furthermore the quality of the infrastructure and the costs prohibit even those elite users who are able to access the Internet. At the time, it looked unlikely that the South African health insurance sector could take advantage of e-commerce due to the low level of PC penetration, generally low Internet usage, and restricted broadband infrastructure. These observations had far-reaching consequences for healthcare insurance organisations hoping to take advantage of the opportunities presented by the Internet during the 2000–2005 period. In the next chapter, I will link the implications of these trends and challenges to our case study organisation.

5.5 Conclusion

In this chapter, I presented a brief background to global healthcare issues and healthcare issues in South Africa in particular, such as the pervasiveness of inequality in healthcare access and the challenges of poverty and HIV/Aids. I also described some of the major challenges facing healthcare insurance firms, especially an inflationary environment of 'crisis' proportions. These inflationary pressures have recently become a major catalyst for the development of innovative healthcare financing models. Particular attention was drawn to private sector organisations in the healthcare insurance industry, and the methods they are instituting in a turbulent regulatory environment. This led to the centrepiece for further discussion – a new healthcare funding model

being proposed under an umbrella concept – 'consumer-driven healthcare'. The consumer-driven healthcare model was shown to rely intimately on information systems, particularly the Internet. Thus, the Internet, and the deployment of online self-service applications in particular, has been viewed by supporters of consumer-driven healthcare as an important ally to their success.

The chapter reviewed the status of the Internet in the light of social, political and infrastructure issues in South Africa during the period of the case study. It alluded briefly to some of the challenges facing healthcare insurance organisations wishing to deploy online self-service technologies, in what can be summed up as a context of startling paradoxes. Understanding at the broader social level has implicit or explicit bearing on developing a better understanding of IS implementation. Therefore the trends and developments described in this chapter have important implications for the implementation of the online self-service technology, at the time of the case study. The goal of the next chapter, chapter 6, is to provide a natural language description of the case study organisation and their SST implementation efforts, prior to a formal analysis in chapter 7 and chapter 8.