UNDERSTANDING SITUATED DESIGN AGENTS

UDO KANNENGIESSER AND JOHN S. GERO
Key Centre of Design Computing and Cognition
University of Sydney
Australia
{udo, john}@arch.usyd.edu.au

Abstract. This paper presents the beginnings of a formal framework to enhance our understanding of situated design agents. It builds on the function-behaviour-structure (FBS) schema to represent essential concepts of situated designing. Our framework covers different stages in the life-cycle of situated design agents, including their development, testing and usage.

1. Introduction

The notion of situatedness has become a recurrent theme in design science as its importance in characterising designing has increasingly been recognised. Aspects of situated designing such as reflection in action, situated cognition and constructive memory could be modelled or simulated using the construct of a computational agent. As a result, situated agents have become essential tools for researchers interested in situated designing, and a growing number of situated agent-based design systems have been developed.

Despite research progress in situated design agents (Gero and Brazier 2002), there is still considerable potential for their deployment on a larger scale. One reason for this is that research has not sufficiently investigated the conceptual underpinnings of situated design agents in terms of a set of formal representations for the fundamental concepts of situatedness and agency. Such representations are needed to improve the overall understanding of agents that encompasses not only the implementation level but also the conceptual level. This can significantly enhance the quality of agent-based innovations, as it is the conceptual stage of their development that must account for their performance across implementation, testing and usage.

2. Background

2.1. SITUATEDNESS IN DESIGN

Designing is an activity during which designers perform actions in order to change the environment. By observing and interpreting the results of their actions, they then decide on new actions to be executed on the environment. This means that the designer's concepts may change according to what they are "seeing", which itself is a function of what they have done. We may speak of a recursive process, an "interaction of making and seeing" (Schön and Wiggins 1992). This interaction between the designer and the environment strongly determines the course of designing. This idea is called situatedness, whose foundational concepts go back to the work of Dewey (1896) and Bartlett (1932). In paraphrasing Clancey (1997) we can summarise it as "where you are when you do what you do matters".

In experimental studies of designers related to the use of sketches, some phenomena which support this idea have been reported. Schön and Wiggins (1992) found that designers use their sketches not only as an external memory, but also as a means to reinterpret what they have drawn, thus leading the design in a new direction. Suwa, Gero and Purcell (1999) noted, in studying designers, a correlation of unexpected discoveries in sketches with the invention of new issues or requirements during the design process. They concluded that "sketches serve as a physical setting in which design thoughts are constructed on the fly in a situated way".

Situatedness also includes a concept known as constructive memory. It is best exemplified by a quote from Dewey via Clancey: "Sequences of acts are composed such that subsequent experiences categorize and hence give meaning to what was experienced before". The implication of this is that memory is not laid down and fixed at the time of the original sensate experience but is a function of what comes later as well. Memories can therefore be viewed as being constructed in response to a specific demand, based on the original experience as well as the situation pertaining at the time of the demand for this memory. Each memory, after it has been constructed, is added to the agent's knowledge and is now available to be used later, when new demands require the construction of further memories. These new memories can be viewed as new interpretations of the agent's augmented knowledge.

2.2. SITUATED DESIGN AGENTS

Agents are entities that are able to sense their environment and to act upon it without direct intervention of human or others (Wooldridge and Jennings 1995). This concept is called autonomy and implies the existence of goals and knowledge in the agent that direct its actions. In a stronger sense, autonomy includes the ability of the agent to learn from its interactions with the environment (Russell and Norvig 1995). As this interpretation of an agent intersects with the idea of situatedness, the emerging notion of a situated agent can be seen as a product of understanding situatedness using the agent paradigm. For example, Maher (2002) has defined situatedness as "the characteristic of a system that can construct its own representation of the context in which it operates based on its own knowledge, goals, beliefs, desires, intentions, biases etc." Gero and Fujii (2000) have proposed a framework to describe the processes involved in constructing these internal representations.

An important feature of a situated design agent is its rationality. Rationality constrains an agent's actions to conform to its current goals and beliefs (Cherniak 1986; Smith and Gero 2000). Combined with the above concept of autonomy this has the effect that the agent can potentially act in ways it has never acted before as it learns and acquires new experiences. One can view this as the agent's ability to use its augmented knowledge and goals to augment its state space of possible actions to be executed on the environment. This makes the agent highly adaptive even to situations it has not encountered before.

An implication of an agent's rationality is that changes in the agent's actions indicate possible changes that might have occurred in its goals and beliefs. Two experiments can be conducted to examine these changes in situated design agents. In the first experiment two identical agents (i.e. clones) are exposed to different sets of design problems to interact with. It is expected that over the course of the experiment the situations constructed by each agent will diverge. The extent to which the agents' situations differ from each other at the end of the experiment can be tested by exposing them to the same design problem and examining the differences in their actions. In the second experiment two non-identical agents are exposed to the same design problem. After a number of interactions with that environment, it is expected that the situations constructed by each agent will converge. The extent to which the agents' situations assimilate can then be concluded from the remaining differences in the agents' solutions to the design problem. Both experiments investigate the relative importance of learning by interaction with respect to initial biases in the construction of an agent's situation.

3. Towards a Conceptual Framework for Situated Design Agents

3.1. AN FBS VIEW OF SITUATED DESIGN AGENTS

In our previous work, we have used the FBS schema to model situated design agents using the notions of function (F), behaviour (B) and structure (S) (Gero and Kannengiesser 2003). The general definitions of these three notions are as follows:

1. The *function* (F) of an entity is defined as its teleology, i.e. "what the entity is for".
2. The *behaviour* (B) of an entity is defined as the attributes that are derived or expected to be derived from its structure, i.e. "what the entity does".
3. The *structure* (S) of an entity is defined as its components and their relationships, i.e. "what the entity is".

Connections between F and B and between expected B and S are constructed through experience. Those between derived B and S are constructed through causality. Figure 1 shows these connections as mappings between the function state space, the behaviour state space and the structure state space. A state space specifies all possible states in terms of the variables and ranges of values that define these states. Function (F) is ascribed to behaviour (B) using teleological reasoning, and behaviour (B) is derived from structure (S) using causal reasoning. The F-B relationship is used to map the observer's view of the functions (F) of the entity on expectations of a set of behavioural states (B) or to construct a set of functions (F) based on actual behavioural states (B). The B-S relationship is used to synthesise an entity in terms of its structure (S) that is to bring about desired behaviour (B) or to explain or predict actual behaviour (B) by means of the structure (S) of an existing entity.

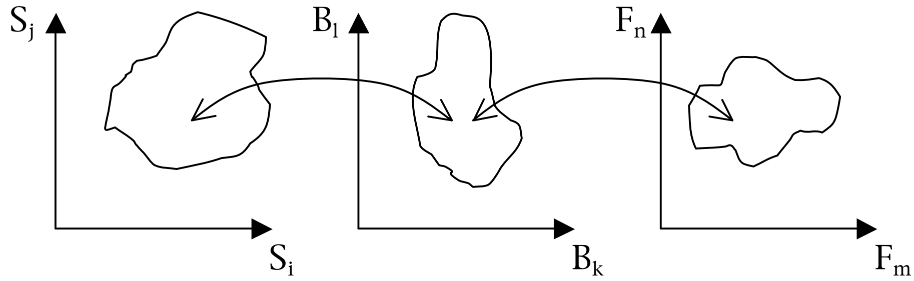

Figure 1. The function state space (F), the behaviour state space (B) and the structure state space (S) are interconnected.

A common interpretation of the function (F) of an agent is its role within some system (which can be a multi-agent system). The agent's behaviour (B) is usually viewed as the way it acts on its environment. Its structure (S) can be understood as the agent's components (i.e. sensors, effectors, architecture, machinery, wires etc.) or internal states and processes.

The connection between function (F) and behaviour (B) of an agent is constructed on the basis of the purposes of the observer that interacts with that agent. The teleological reasoning here does not differ from the one needed for entities other than agents. In contrast, the causal reasoning required to link behaviour (B) to structure (S) can be quite distinctive depending on the stance taken by the observer (Dennett 1987):

- *The physical stance*: This stance views the agent as a designed artefact and has detailed technical knowledge about all the structure (S) that causes its behaviour (B). Here the causal reasoning is based on physical laws or symbolic mapping rules.
- *The intentional stance*: This stance views the agent as an autonomous, rational entity and ascribes abstract, "mentalistic" states such as goals, desires and beliefs to the agent, which form the structure (S) underlying its behaviour (B). Here the causal reasoning is based on the principle of rationality that, similar to physical laws, governs all of the agent's actions (Newell 1982).

The intentional stance is an abstract perspective of an agent's structure (S) that is independent of its technical realisation in the physical stance. It is typically applied in the early conceptual stages of developing agent-based systems, during which behavioural specifications (B) of agents are linked to high-level, "mentalistic" descriptions of their internal structure (S).

Situated agents can construct new knowledge structures (S) and consequently exhibit new behaviours (B) as a result of their interaction with the environment. These new structures and behaviours complement pre-programmed structures and behaviours necessary to spawn these agents and execute lower-level processes. We can therefore distinguish between situated structure (S^s) and fixed structure (S^f) (Gero and Kannengiesser 2003). Situated structure (S^s) includes all modifiable parts of the structure state space, which can be viewed as new internal symbols (physical stance) or new goals, beliefs, etc. (intentional stance). Fixed structure (S^f) includes all non-modifiable parts of the structure state space, which can be viewed as all persistent components of the agent's architecture, such as sensors, effectors, memory system, reasoning system, etc., and hard-coded parts of the agent's knowledge (in terms of symbols or "mentalistic" concepts).

Newly constructed situated structure (S^s) can potentially bring about new behaviours (B), which in turn can drive the construction of further knowledge that is added to the agent's situated structure (S^s). In the context of situated design agents, this is known as "interaction between making and seeing" (Schön and Wiggins 1992). Figure 2 shows how this interaction can be modelled as the "co-evolution" of B and S^s, a notion borrowed from Maher and Poon's (1996) model of designing. The agent uses its current structure at time t (S^s (t)) to produce a rational action that is modelled as its behaviour at time t (B (t)). An example is a design agent producing a sketch of its current design. The results of this behaviour can be used by the agent at a later point in time t+1 to reformulate its knowledge or goals leading to a modified structure (S^s (t+1)). Here it is possible that a useful yet previously unintended concept emerges from the sketch and is included into the agent's current set of design goals. This evolved structure can then cause a new rational action resulting in evolved behaviour (B (t+1)), e.g. in terms of a new sketch describing the modified design.

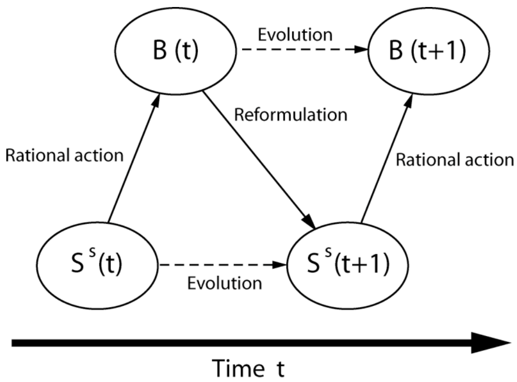

Figure 2. Co-evolution of the structure (S^s) and behaviour (B) of a situated agent (adapted from Maher and Poon (1996)).

This model of co-evolution does not apply for non-situated design agents: The relation between their structure (S) and behaviour (B) is uni-directional, i.e. they are only capable of rational action that does not feed back in a way to significantly alter their structure (S). As a result, their structure (S) state space is fixed and is not able to produce a novel behaviour (B) that is outside their routine behaviour (B) state space.

3.2. INTEGRATING THE OBSERVER'S PERSPECTIVE

The FBS view described so far has adopted the perspective of an omniscient observer that has access to all details of the internal structure (S) of an agent, which includes its situated and fixed parts. This perspective is usually available only to the designer of the agent. Once that agent has been designed and put into an environment within which it can operate, its internal components and states are usually opaque to the observer. The main source of information about the agent then are the changes it brings about in its environment, providing the grounds for readily modelling the agent's behaviour (B) and externally visible parts of its fixed structure (S^f) such as its sensors and effectors. Other portions of the agent's structure (S), in particular its situated structure (S^s), can only be inferred from the agent's behaviour (B). The result of this inference process strongly depends on the experience of the individual observer.

This bounded perspective of the observer poses a problem for differentiating between situated structure (S^s) and fixed structure (S^f). As most of the agent's initial structure state space is unknown, it is often not clear if a behavioural (B) change has been caused by a modification of the structure state space or by moving within the existing structure state space. In design, it often remains an open question if a particular design feature was intended from the outset of the design task or invented during the interaction of the agent with external design representations. One way could be to track the complete sequence of design interactions of that agent. The availability of such a comprehensive history, however, is usually unrealistic for an observer.

To address this issue, let us distinguish between a *strong notion of situatedness* or *s-situatedness* and a *weak notion of situatedness* or *o-situatedness* ("observer"-situatedness). Parts of the structure (S) of an agent are s-situated if the designer of that agent (or any other omniscient observer) assumes them to be situated based on knowledge about the agent's implementation and test results. Parts of the structure (S) of an agent are o-situated if there is an external observer with bounded rationality that assumes them to be situated based on observations of the agent's interactions and the observer's prior experience and expectations.

Situated design agents have mostly been viewed using the strong notion of situatedness. The weak notion of situatedness now introduces a view of situated design agents that is itself situated, as it integrates the bounded perspective of an external observer. This allows superimposing the observer's situatedness on the agent's situatedness and thus maintaining the distinction between a situated and a fixed structure. As the weak notion of situatedness does not constrain the knowledge and expectations of the individual observer, any part of the agent's knowledge may be regarded as situated by that observer. However, we assume that all external observers share the same stance – the intentional stance – when viewing an agent's knowledge. The weak notion of situatedness can therefore include all "mentalistic" ascriptions of knowledge to an agent, without the agent necessarily being s-situated at all.

3.3. INTERACTIONS OF SITUATED DESIGN AGENTS

The weak notion of situatedness provides a pragmatic way of modelling agents independently of their implementation and thus shifts the focus from an omniscient design view to an external view based on the agents' interactions. It stresses the importance of behaviour (B) and assigns to structure (S) the role of supporting the explanation or prediction of behaviour (B).

Besides the intentional and the physical stance that establish the B-S relationship, Dennett (1987) has proposed a third stance to infer the behaviour (B) of an entity or agent: *The design stance* uses expectations about the way an agent is likely to behave in order to realise its assigned role. For example, one can expect a computer's behaviour (B) of starting up after pressing the "on" button based on the assumption that this is how the computer has been designed to perform its function of getting into service for a user. We can map the design stance on the F-B relationship. Figure 3 shows how the FBS view of a situated design agent – using the weak notion of situatedness – can summarise Dennett's (1987) three stances to infer an agent's behaviour (B).

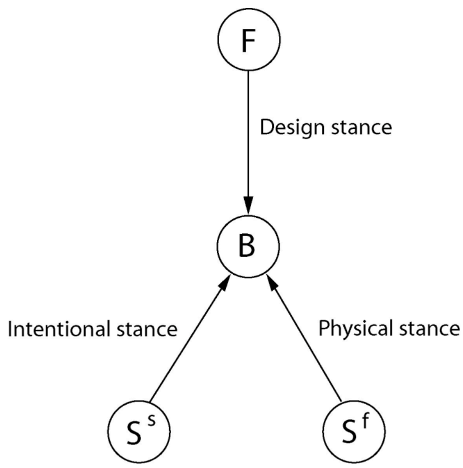

Figure 3. Dennett's (1987) three stances to infer an agent's behaviour using the FBS view.

An observer's FBS view of a situated design agent has to be kept consistent throughout the agent's interactions with its environment (that the observer may or may not be part of). In particular, this maintenance is required when unexpected behaviour (B) occurs. Behaviour (B) can be unexpected because the observer's expectations have been insufficient or incorrect or because the agent has altered its behavioural patterns (B) as a result of modifications in its situated structure (S).

Most changes in the observer's FBS view of a situated design agent concern the agent's situated structure (S^s). The process of revising the FBS model then resembles the above model of co-evolution, Figure 4. However, considering that we are dealing here with the weak notion of situatedness, there are two differences: First, this situated structure (S^s) and its connections to behaviour (B) are immersed into a "sea of assumptions" that complement the directly observable results of behaviour (B). The only assumption that can usually be taken for granted is the rationality of the agent's action choice. Second, the co-evolutionary process is now driven by observations of the behaviour at time t+1 (B (t+1)) that are inconsistent with the behaviour at time t (B (t)). Figure 4 alludes to this role of behaviour (B) as the driver by shifting its time-dependent representations (i.e. B (t), B (t+1) etc.) slightly forward along the time axis (as opposed to Figure 2 where they are shifted backward). The behavioural observations together with the assumption of rational action choice are then used to abductively infer an evolved model of the structure (S^s (t+1)) that accounts for the new behaviour. The new FBS model then holds until a new behaviour (B) is observed that is not consistent with that model.

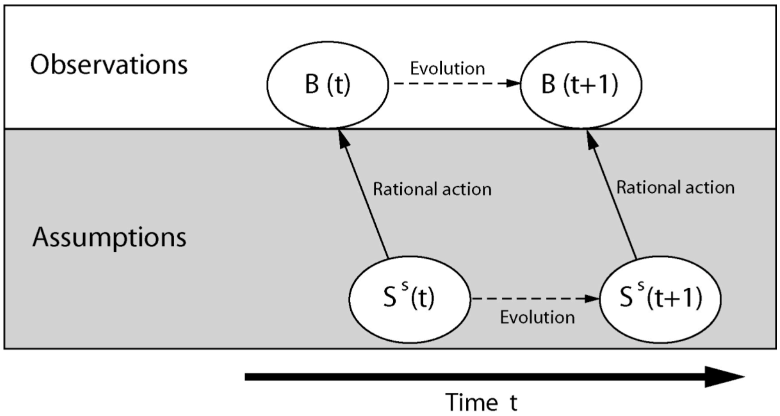

Figure 4. Co-evolution of the structure (S^s) and behaviour (B) within an observer's FBS view of a situated design agent.

3.4. TESTING SITUATED DESIGN AGENTS

The stage of testing can be seen as the interface between the development and usage of situated design agents. Testing involves putting the agents into well-defined test environments and observing or analysing particular behaviours. These behaviours are then evaluated by comparing their values with pre-set performance requirements. Depending on the outcome of this comparison and on the type of performance requirements, this process is used to verify or validate the implementation of the agent.

In terms of our FBS ontology, testing involves observing the system's behaviour (B) based on the interactions of its internal structure (S) and the test environment. Situated design agents are typically tested for the extent to which they can adapt to their environment, as adaptability is a behavioural (B) and thus observable consequence of the degree of situatedness of an agent's structure (S). The two experiments outlined in Section 2.2 are good examples: They observe the evolution of the different behaviours (B) of two situated design agents and analyse the relationships between these behaviours for divergence or convergence, Figure 5. Finally, they infer the evolution of their situated structures (S^s) by mapping differences in behaviour (B) directly on differences in situated structure (S^s).

This test may be seen as a demonstration of o-situatedness. However, the tester, as the external observer of the agents, is quite well informed about the agents, in particular about (the differences in) their initial structures (S). In addition, the tester has control over the sequence of interactions of the agents by systematically modifying their environments. As a result, the structures (S^s) that are to be inferred from observed behaviours (B) are easily accessible. For other observers such as peer agents, these structures (S^s) are located on deeper levels in the "sea of assumptions" mentioned in Section 3.3, as they often cannot control the complete environment of the tested agents.

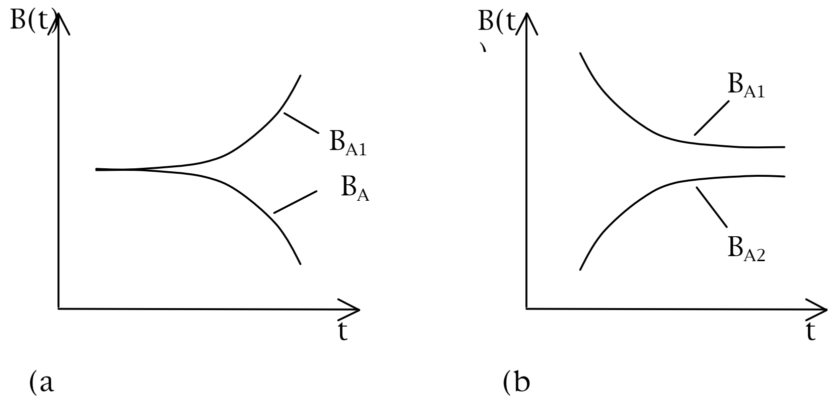

Figure 5. The expected behaviour (B) of two situated design agents (A1 and A2) during the course of two experiments:
(a) A1 and A2 have identical initial structures (S) and are exposed to different environments
(b) A1 and A2 have different initial structures (S) and are exposed to the same environment

A good way for an external observer to infer the evolution of the agent's internal structure (S^s) is to make the agent tell that observer about every change in its structure (S). This is done via communicative actions, which are a form of behaviour (B) that is the carrier of representations of the agent's knowledge and goals (S^s). We can model communicative actions as $B(S^s)$, and the inference of the structure contained in this behaviour is a rather trivial task, provided the agent can be trusted (Brazier and Wijngaards 2002). Protocol studies, in which designers are asked to "think aloud", are an example for this way of conducting experiments with situated design agents.

4. Conclusion

This paper has outlined the beginnings of a formal framework for representing situated design agents throughout all the stages of their life-cycle. This extends the previous emphasis on only the development or design stage. It also re-directs the focus onto the conceptual underpinnings of situated design agents. This can enhance an overall understanding of this emerging area within design research. In addition, it provides researchers with a uniform terminology and guidance for developing, testing and using situated agents in their design systems. This can potentially increase consensus on research ideas and methodologies in this field and produce more and better agent-based design systems.

Acknowledgements

This research is supported by grants from the Australian Research Council and the University of Sydney.

References

Bartlett, F. C.: 1932 reprinted in 1977, *Remembering: A Study in Experimental and Social Psychology*, Cambridge University Press, Cambridge.

Brazier F.M.T. and Wijngaards N.J.E. (2002) The role of trust in automated distributed design, *in* J.S. Gero and F.M.T. Brazier (eds) *Agents in Design 2002*, Key Centre of Design Computing and Cognition, University of Sydney, Australia, pp. 71-84.

Cherniak C. (1986) *Minimal Rationality*, MIT Press, Cambridge, MA.

Clancey W.J. (1997) *Situated Cognition: On Human Knowledge and Computer Representations*, Cambridge University Press, Cambridge.

Dennett D.C. (1987) *The Intentional Stance*, MIT Press, Cambridge, MA.

Dewey, J.: 1896 reprinted in 1981, The reflex arc concept in psychology, *Psychological Review* **3**, 357-370.

Gero J.S. and Brazier F.M.T. (eds) (2002) *Agents in Design 2002*, Key Centre of Design Computing and Cognition, University of Sydney, Australia.

Gero J.S. and Fujii H. (2000) A computational framework for concept formation for a situated design agent, *Knowledge-Based Systems* **13**(6): 361-368.

Gero J.S. and Kannengiesser U. (2003) Function-Behaviour-Structure: A model for social situated agents, *in* R. Sun (ed.) *Workshop on Cognitive Modeling of Agents and Multi-Agent Interactions,* International Joint Conference on Artificial Intelligence 2003, Acapulco, Mexico, pp. 101-107.

Maher M.L. (2002) Summary of group working session, *Workshop on Agents in Design 2002*, Massachusetts Institute of Technology, MA.

Maher M.L. and Poon J. (1996) Modelling design exploration as co-evolution, *Microcomputers in Civil Engineering* **11**: 195-210.

Newell A. (1982) The knowledge level, *Artificial Intelligence* **18**(1): 87-127.

Russell S. and Norvig P. (1995) *Artificial Intelligence: A Modern Approach*, Prentice Hall, Englewood Cliffs, NJ.

Schön D. and Wiggins G. (1992) Kinds of seeing and their functions in designing, *Design Studies* **13**(2): 135-156.

Smith G.J. and Gero J.S. (2000) The autonomous, rational design agent, *in* H. Fujii (ed.) *Workshop on Situatedness in Design*, Artificial Intelligence in Design'00, Worcester, MA, pp. 19-23.

Suwa M., Gero J.S. and Purcell T. (1999) Unexpected discoveries and s-inventions of design requirements: A key to creative designs, *in* J.S. Gero and M.L. Maher (eds) *Computational Models of Creative Design IV,* Key Centre of Design Computing and Cognition, University of Sydney, Sydney, Australia, pp. 297-320.

Wooldridge M.J. and Jennings N.R. (1995) Intelligent agents: Theory and practice, *Knowledge Engineering Review* **10**(2): 115-152.

This is a copy of the paper: Kannengiesser, U and Gero, JS (2005) Understanding situated design agents, *in* A Bhatt (ed), *CAADRIA'05* Vol 2, TVB, New Delhi, pp 277-287.