Identifying Theories of Agency In Information Systems

Reeva Lederman
University of Melbourne

Simon Milton
University of Melbourne

Robert Johnston
University of Melbourne

Follow this and additional works at: http://aisel.aisnet.org/amcis2004

Recommended Citation
http://aisel.aisnet.org/amcis2004/193

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2004 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
Identifying Theories of Agency In Information Systems

Reeva Lederman  
University of Melbourne  
reeval@staff.dis.unimelb.edu.au

Simon K. Milton  
University of Melbourne  
smilton@unimelb.edu.au

Robert B. Johnston  
University of Melbourne  
r.johnston@dis.unimelb.edu.au

ABSTRACT

Previous research has suggested that systems that are designed according to existing methodology fail in routine contexts whereas others that have evolved in such contexts without a traditional methodology work better. In previous work we have suggested that this is because the theory about purposeful activity that underlies the design of a system has an impact on the success of the system in particular task domains. In order to investigate this empirically we need to establish a method for determining the underlying theory of agency that informs an existing system’s approach to purposeful activity. This poses the following question: how do we recognize the theories of agency in systems? To investigate this, we propose a method for extracting the theory of agency from an existing system where we have tentatively assigned an underlying theory of agency to the system but want to definitively categorize it. This is a vital first step in developing an IS design methodology for routine information systems.

Keywords


INTRODUCTION

Arguably, all information systems are used for purposeful activity and are designed to achieve certain goals or carry out certain tasks. Therefore it should be important to study the nature of purposeful activity to determine how we can make systems more effective. There are two main theories of agency which address the nature of purposeful activity: one, the deliberative approach which asserts that people use an abstract model of the world to deduce actions that will achieve a goal (Vera and Simon, 1993), and a second, the situated approach (Suchman 1983, 1993), which emphasizes reactive responses to situations thrown up by structured environments.

In an earlier work it has been proposed that the theory of agency underlying systems’ design might effect the performance of the system in certain task environments (Johnston and Brennan, 1996, Johnston, 2001, reference withheld during review, 2003) and that situated theories of agency are more appropriate in routine environments. This paper is part of an ongoing research project that examines the appropriateness of a situated theory of agency as the basis for design in routine systems. In this research we examine an initial essential aspect of this project through the following research question: how do we uncover the theory of agency that influences the design of activities in an existing system? This is an important question that we must answer to complete the overall project because in order to test our claim that routine systems are more appropriately supported by the situated theory of agency we must be able to empirically test a system to determine which theory of agency is a basis for its design.

To investigate this question empirically requires the availability of a method for connecting systems design characteristics with theories of agency. In this paper we firstly examine theories of agency that explain purposeful activity, extract the issues that distinguish the different theories together with the commitments of the theories with regard to these issues. Then we look at two systems and make a cursory analysis of the possible theory of agency that might underpin their design. We then provide a methodology for attempting this task in a more rigorous and repeatable manner. We conclude with an analysis of the significance of this work for the production of better systems.
THEORIES OF AGENCY

Theories of Agency provide models for the way in which agents achieve goal directed behaviour in a system through purposeful activity. The deliberative theory of agency posits that an agent acts in a goal-directed way by building and maintaining an abstract model or representation of the world and applying deductive processes to this representation to determine what to do. This stance commits the theory to a certain position about the nature of the world that can be called 'objectivist' (Lakoff 1987, p159). The world consists of objects with properties. A particular object is of a given kind because it displays certain properties characteristic of that class of object, and differs from other objects of that class by virtue of its particular values for the properties. Objects stand in certain relations to one another. The world is characterised by a set of possible states of the world, which are allowable configurations of objects (including the agent) under these relations.

The representation of the world is of a particular kind, a “symbol/object” representation (Agre 1997), in which configurations of symbols, their attributes and relations are isomorphic to objects, properties and relations within states of the world. The agent uses sense data to define the current state within this world model and then applies deductive processes to find a sequence of symbolic "actions" that will transform the model current state to the model goal state. This sequence of symbolic actions is a “plan” that is handed to an effecter system that implements it in the real world.

In the past decade or so researchers in diverse disciplines have argued against the efficacy of the deliberative theory of agency (Brooks 1986; Suchman 1987; Brooks 1991a; Johnston 1995; Hendriks-Jansen 1996; Johnston and Brennan 1996; Clancey 1997, Agre, 1997 #392) on the grounds that even if the world can be represented in such an abstract symbolic way (a debatable ontological assumption (Dreyfus 1992)) it does not provide a manageable engineering solution to the operation of artificial, autonomous agents, particularly in the face of time-constrained routine action. For instance, in a series of proof-of-concept robot designs, Rodney Brooks and co-workers (Brooks 1986; Brooks 1991b; Mataric 1992) have shown that goal-achieving behaviour is possible in robots that do not make use of deliberation or centralised symbolic representations of their environments of action. Their robots use more or less reactive behaviours to establish complex interactions between the agent and features of its environment. They have demonstrated that the performance of such agents in complex environments is more robust than that of deliberative agents and degrades more gracefully in the face of the unexpected. In a similar vein, Agre and co-authors (Agre and Chapman 1987; Chapman 1991; Agre and Horswill 1992) have argued for a theory of routine action based on the interaction of relatively simple action selection rules with the structure of environments and have demonstrated the possibility of robust goal-directed action in this way using computer-simulated agents.

Emerging from this technical work, is an alternative theory of agency with substantially different representational and ontological commitments to the deliberative theory. The key insight is the essential infeasibility of the continual translation, required by the deliberative theory, from the world of action to the world model and back again to action (Johnston and Brennan 1996; Agre 1997), as a principle for time-constrained action in complex environments. This translation is required because the world model is an “aerial” or “transcendental” objective view of the world from outside the agent and independent of its goals. Any real agent has available only an incomplete, skewed “ground” view from its sensors with which to construct and update its world model. Despite this difficulty, it is only by building this aerial world model that its own goal can be expressed and lead to action selection. In addition, the strategy of instantiating the world model independently of the agent’s goal means that it is difficult to bound the size of the world model, which typically leads to a computational complexity explosion during action planning (Chapman 1988).

The solution to these dilemmas proposed in “situational” theories agency is to provide an agent with more or less reactive responses based on sense data obtainable directly from the agent’s ground view of the world, and to introduce the agent's intentions and point of view early on in the theory at the level of the ontology and representation schemes. According to this theory, agents respond reactively to "situations" without deliberation. Situations, in contrast to "worlds states", are accounts of the world from the agent’s point of view. They are parsimonious descriptions because they need only include features of the world that are relevant to the agent’s purposes (Agre and Chapman 1987). These features consist of the relations of things to the agent given its goals. Actions are selected from a repertoire of more or less reactive rule-like responses to situations. Such a simple approach to action selection will only work (be goal attaining) if the action environment exhibits structure that makes complex planning unnecessary (Agre and Chapman 1987), and analysis and exploitation of environmental structure is an important part of designing situated agents (Agre and Horswill 1992; Hammond et al. 1995; Horswill 1995; Agre and Horswill 1997).

Because situations are agent-centred and intention-laden, representation of situations with a symbol/object isomorphism is neither possible nor necessary. However, aspects of situations do need to be marked in some way in order to fire situation-action responses. Agre and Chapman argue that the appropriate representational scheme should be “indexical” and
“functional” in nature (Agre and Chapman 1987). Indexical representations describe the disposition of things relative to the agent and functional representations select things according to their relevance for the purposes of the agent. An example of an indexical/functional reference is “the-keyboard-on-which-I-am-typing-now”, compared to the objective reference “keyboard (serial-number = “T4391”, position = "Kitchen 3 Pearson St. Richmond", time = "16-46 21/01/02")”. Agre and Chapman (1987) argue that the ability to rebind indexical/functional representations to situation features that are relevant to the agent’s goals on the fly, eliminates the computational complexity explosion inherent in using all encompassing aerial world models as a basis of action. The reliance of the situational theory on indexical/functional rather than symbol/object representation is one of a number of issues where we see differing commitments expressed in the two theories. Six of these commitments, synthesized from the literature discussed above, are summarised in Table 1:

<table>
<thead>
<tr>
<th>Issues</th>
<th>Commitments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Type of representation</td>
<td>Deliberative Theory: The symbol system has a one to one correspondence with objects external to the user.</td>
</tr>
<tr>
<td></td>
<td>Situational Theory: The representation of an entity depends on the occasions of its use and how it is used.</td>
</tr>
<tr>
<td>2. What represented</td>
<td>Deliberative Theory: Objects (Entities and properties) represented as though external to and independent of the user.</td>
</tr>
<tr>
<td></td>
<td>Situational Theory: Situations and aspects of situations relevant to goal-attainment.</td>
</tr>
<tr>
<td>3. Persistence of representation</td>
<td>Deliberative Theory: Persistent account of the state of the world from a dispassionate viewpoint.</td>
</tr>
<tr>
<td></td>
<td>Situational Theory: Transient representations intended only for initiating immediate action.</td>
</tr>
<tr>
<td>4. Underlying ontology</td>
<td>Deliberative Theory: Objective: makes objective claims whose truth or falsity is independent of what anyone thinks or feels about the matter.</td>
</tr>
<tr>
<td></td>
<td>Situational Theory: Perspectival: the world is modelled through entities as appreciated from the agents situated perspective.</td>
</tr>
<tr>
<td>5. Action selection</td>
<td>Deliberative Theory: Deduced from representations of the state of the world and resulting in plans and schedules for future execution.</td>
</tr>
<tr>
<td></td>
<td>Situational Theory: Immediate (reactive) rule-like responses to aspects of situations.</td>
</tr>
<tr>
<td>6. Significance placed upon environmental structure</td>
<td>Deliberative Theory: Structuring of the social and physical environment of work marginalised.</td>
</tr>
<tr>
<td></td>
<td>Situational Theory: Structuring of social and physical environment of work manifest as essential to system functioning.</td>
</tr>
</tbody>
</table>

Table 1: The Commitments of the Deliberative and Situated Theories

EXAMPLES OF SYSTEMS

Keeping the issues in Table 1 in mind, let us now have a look at some systems. A first system (System 1) is a typical resource planning system that one might find in highly computerised manufacturing environments. In a traditional computerised system a master production schedule represents the company-wide production goal. This is processed against a product structure database that describes how raw materials and bought components can be converted into saleable items. The current availability of products is checked against an inventory database and, by purely deductive means, schedules for buying and assembling component parts to meet the master schedule are produced and handed to operators for execution (Wright, 1981, MRPII). Completion of these schedules is recorded through transaction processing systems and the planning process is repeated periodically.

A second very different example (System 2) is an emergency dispatch system for ambulances described by Wong (2000) and Wong and Blandford (2001). It is a dynamic domain where it is essential to the process that operators are aware of the
goings-on in and around the area being covered by the ambulances and are able to predict the future state of ambulance control to cater for possible eventualities. The system is functionally divided into two areas: call taking and prioritisation and command and control of emergency ambulances. There is a single control room where call takers sit on one side and dispatchers on the other. The call takers prioritise calls and the dispatchers communicate with ambulance crews. "Details of calls for ambulances are printed onto tickets which become the means by which incidents are managed. Ambulances are assigned and this information and additional updates are recorded by hand on the ticket. The ticket is placed in a slot in the allocation box. The placement of the tickets in the box provides cues about the operation of ambulances." The physical layout of the room, seating arrangements of the staff and placement of the tickets is crucial to the effectiveness of the system.

Looking at these examples it seems, intuitively, that the first system is informed by the deliberative model in the sense that we see symbols which can be manipulated being used to represent entities properties and relationships in the system. Actions are initiated by a process of reasoning (planning) applied to a symbolic model of the world. All environmental structure is implicitly contained in the world model so that the structuring effect of environments for action is marginalised. In the MRP type of system it is assumed that human actors just receive schedules and act upon them.

In the second system, the emergency dispatch, we see an environment deliberately structured to provide support for the work being carried out with the help of rule based actions, which make use of transient and ephemeral representations. The dispatch officers describe information “coming from different sources… interleaved with information about different incidents…” which operators need to react to. The operator’s environment is not static but we see that they develop routines to deal with the information and allow new inputs from the environment to be assimilated. In the environment described there are tags such as the “spatial positioning of tickets ...(and) tickets that indicate handing over a job to another dispatcher...” illustrating the active use of environmental structure. These environmental inputs reduce the need for cognitive processing. Structure in the environment takes some of the cognitive burden away from the system operators (Agre and Horswill, 1997).

There is a significant paradigmatic difference between the systems designs. One is a system designed to support the world of action while the other abstracts information processing and communication with no regard for the world in which the system operates.

Prima facie it appears that these two systems can be categorised as deliberative and situated, respectively. These assignments are however, tentative. Our objective in this paper is to make these assignments more explicit and to establish a genuine correlation between the traits in a system and the commitments of the theory of agency that informs the system. We want to make a more rigorous association between the theories of agency detailed above and the built systems we observe. It is expected that the commitments to a particular theory will be manifest in observable traits. Thus, we need to develop a precise methodology for measuring this correspondence.

The methodology needs to address the following question: how can we extract commitments from the theories and map them onto the properties found in a system to determine the theory of agency that informs the design of a system?

DETERMINING THE THEORY OF AGENCY IN A SYSTEM

This section considers how designers build characteristics into systems that embody a particular theory of agency. Completed systems are comprised of both a design process and a product, the system itself. There are however, significant problems in extracting a design process from a completed system. When we look at the systems described above we imagine we can work out the theory of agency used by the designers. However, we aim to develop a methodology where there is certainty about our units of analysis, where the analysis is repeatable across systems and where the categorisation of a particular system is unambiguous.

We have two possible options: asking the designer his/her intention in developing the process or examining the built system without reference to the designer. The first option makes a presumption that designers consciously subscribe to a particular theory and map this onto traits that are left in the system. However, in the routine systems from the literature cited above we see systems that have largely evolved over time often with no clearly identified designers or with the participants in the system creating the system in an ad hoc manner (Wong 2000, Zuboff 1983, Bentley et al 1992). As in biological systems, here the true “designer” is evolution. For example, in Zuboff’s paper mill the workers are integral to the evolving design of the work system in which they participate and are required to “control certain parts of the process… and make proper adjustments for achieving the best results.” In this system and others in the literature systems participants evolve a system over time while not necessarily being able to clearly articulate the mode of purposeful activity needed to achieve goal attainment. In other cases in the literature designers are also difficult to identify and the way in which they intend activity to
be carried out cannot be clearly stated. Thus, a second approach is to look for concrete manifestations of design commitments about agency in the systems themselves. In a sense, this is “reverse engineering” the position of the designer whether the designer is a person or an evolutionary process.

The second option is more feasible: look at the built system and extract the position of the designer. We propose that theories of agency are displayed as objective commitments of a system that we can find in the system without interviewing the designer. The literature suggests that designers’ choices can be difficult to extract from the artefact itself (Agre, 2001). While it may be difficult to reverse-engineer compiled code or extract intention from closed robotics systems the possibility exists where the system contains concrete manifestations of the design process and in some socio-technical systems it is possible to suggest a method for doing so. If a designer has a specific theory of agency it will appear as some of the commitments to a theory that we saw in Table 1 and we can find these commitments in the system without reference to the designer. To examine this we need to inspect the system artefact and be able to map its features against the set of expected features that reflect the commitments of a particular theory. We will demonstrate how to do this in the next section for the case of Systems 1 and 2 previously described.

**FINDING THE THEORETICAL COMMITMENTS IN A SYSTEM**

It is possible to map the commitments of the theories shown in Table 1 onto the examples (Systems 1 & 2) discussed previously and see if evidence of either the deliberative or situated approach is present. Are there features in the two systems that can be viewed as physical manifestations of either theoretical commitment? Conversely, are we certain that traces of the alternative theory are not present?

The first issue considered in Table 1 is the type of representation expected given either theoretical commitment. In System 1 the type of representation is of the form expected under the deliberative theory, symbol/object, expressed as a computerised database. A computer system such as this takes a symbol token as input and uses it to gain access to a referenced object in order to affect or to be affected by it in some way. In System 2 we see a completely different type of representation, the indexical/functional representation we expect in a situated system. With this type of representation meaning is not constant as we see with the physical tickets that mean different things as the situation changes. Conversely, the fixed fields with constant meanings that we find in System 1 could never be seen as having the ephemeral qualities of a situated system. Similarly, the tickets in System 2, which change meaning with changes in their spatial positioning, could never be seen as representing the fixed truths that are the hallmarks of systems designed under a deliberative theory.

The second issue focuses on what type of entities are represented in the system. In System 1 we see a clear physical manifestation of the commitment to the deliberative theory where we find fixed records and fields representing objects and their properties. In System 2 what we see represented are aspects of the situation, like the change of a dispatcher that is represented by the physical handling of a ticket. We can see no manifestation of the situated approach in System 1 because what is represented is completely external to and independent of what is happening around the system. In System 2 these independent objects on which one may deliberate do not exist but rather the situation or aspect of a situation is crucial to goal attainment.

The third issue looks at how persistent is the representation and again we see an unambiguous commitment to a deliberative approach in System 1 and a situated approach in System 2. In System 1 again we see a clear physical manifestation of the former approach: a structured, constant database that persists despite any transformations of the actual data or changes in the world in which the database exists. In System 2 however the representations are transient, as the tickets, for example, take on a new meaning when their spatial positioning changes. These ephemeral representations are typical of situated systems, and not evident in the fixed fields with constant properties found in systems like System 1 (eg. field: Component).

When we examine the underlying ontological commitments of each theoretical approach, issue 4, we also see a commitment to the deliberative approach evidenced in System 1. That is, we find elements in the database that are in direct correspondence to real world objects described from a dispassionate and uninvolved viewpoint. In System 2, on the other hand, the ontological commitment is clearly non-objective: tickets do not correspond to real-world objects but are subjective and the “reality” represented is highly perspectival. The world in System 2 is modelled through situations that take on a meaning entirely dependent on the agent’s position in the environment.

The fifth issue is action selection. Clearly System 1 involves creating a model of the world and deducing what to do through deductive planning. It is an approach that negates the situated theory that would suggest that factors in complex changing environments influence action selection and make fixed plans unfeasible. Conversely, when action selection takes place in
System 2, the environment is a vital component. Actors in the system react in a rule-like manner depending on environmental cues. To do this they use the existing schemata of possible problem situations against which presented cues (eg. the next emergency) are compared.

In System 1 the environmental structure is implicitly contained in the database and there is a tendency for managers to view attention to the physical environment as mere practical wisdom (Johnston and Brennan, 1996). This clearly illustrates a commitment to the deliberative theory in regard to the sixth issue, which looks at how much significance is placed upon environmental structure. In System 1 actions are taken based only on the nature of the objects contained within the database and the environment of the system is irrelevant. In System 2, however, there is a high significance placed on the way the environment is structured (eg it is vital to prioritise call taking). We see in System 2 an intensely time constrained process which creates a high level of need for situation awareness. This makes the physical placement of aspects such as status screens important – an issue that would be unimportant in a system like System 1.

In the analysis above we have examined the two systems to look for physical manifestations of the commitments to the two theories. We have clearly found the manifestations of a deliberative approach in System 1 and of a situated approach in System 2. Conversely, we have shown how it would be difficult to claim that the opposite theoretical commitments could be applied to either system.

Based on this analysis our proposed methodology is to map the systems examined onto the extracted theoretical commitments and find a correspondence that dictates evidence of either a situated or a deliberative approach. This is done by firstly describing each of the systems through the typical activities undertaken using the systems. Following this, each description is examined for evidence of the commitments to the two opposing theories of agency. By comparing both systems to each theory of agency we can see the categorisation is un-ambiguous as the commitments of the alternative theory are shown not to be evident. The methodology should be repeatable across other open systems whose traits are readily observable.

CONCLUSION

In this paper we have argued the proposition that one can tell the theory of agency that informs the design of systems and have set out the first steps in a methodology for doing so. This methodology, which proposes examining systems to uncover material manifestations of indicative commitments to the two most prevalent theories of agency, has provided a novel approach to viewing systems through the theoretical lens of commitment to agency. In the two cases that we have examined in depth we have reflected on the qualitative differences between the systems and produced evidence of a commitment to a particular theory of agency as well as evidence that the alternative model does not apply. Through this we have devised a defensible case for proposing that a particular system appears to have been designed following the presumptions of a particular theory by making a rigorous association between the theories of agency and the built systems described. This work is an important step in establishing the on-going research proposition that a situated theory of agency provides a better methodology for the design of routine systems. Given what has already been found about the effectiveness of routine systems based on a situated theory of agency, it is an important starting point in the development of tools for producing more effective systems.

One limitation of this approach is that the style of analysis is only possible on systems that expose some degree of their workings in a concrete form. In some closed systems it would be difficult to extract the features of the system as we have done with the systems selected. However, because we have been able to unambiguously ascribe one theoretical commitment to each system with the other obviously not applying, it is clear that these early results are promising and that the methodology may be feasible in some circumstances.

Our next task is to refine the lists of commitments and the methodology for deciding whether these dimensions are, in fact, the correct units of analysis to uncover the specific theories of agency as well as to determine how repeatable this analysis is across other systems.
REFERENCES