Modelling Software Organisations

David Hales
Centre for Policy Modelling, Manchester Metropolitan University. UK.
dave@davidhales.com

Chris Douce
Feedback Instruments, Crowborough, UK.
chrisd@fdbk.co.uk

Keywords: POP-I A, group dynamics, team structure, POP-V B, research methodology, Agent Based Social Simulation (ABSS), Multi-Agent Based Simulation (MABS).

Abstract

We view software construction as a *social process* embedded within organisational and cultural structures. To increase our understanding of the process of software construction we attempt to model not just individuals or teams but organisational wide processes. How can we begin to model such complex dynamic entitles composed of many interacting individuals? We outline our initial steps towards such an enterprise (employing multi-agent based simulation) and argue that this is a worthwhile endeavour if we are to increase our understanding of software production in the real world.

1. Introduction

We aim to increase the understanding of the process of software construction, approaching this problem from a social (organisational) perspective. It would seem that all software construction (certainly at a large-scale) is fundamentally a social process. Programmers work within teams and groups which themselves are embedded within larger organisational structures and cultures. These social structures enable, constrain and shape the behaviour, knowledge and general conceptual repertoires of individuals involved in the software construction process.

The process is governed by and is situated within complex social institutions (whether traditional commercially driven hierarchical corporations or decentralised electronically mediated cooperative non-commercial enterprise). It may even be claimed that the organisation and its culture are more important in determining the quality and efficiency with which software is produced than the individuals that comprise it.

A method is outlined by which we intend to explore these organisation wide phenomena in a quasi-empirical way. We intend to use empirical work carried out within the Psychology of Programming (PoP) and apply it to the construction of an Agent Based Social Simulation model (ABSS). Construction of such a model will force us to be explicit about our assumptions and highlight gaps in our understanding and suggest novel hypotheses and future empirical work. Experimentation using the model will allow for the comparison of various organisation structures, cultures and models of programmer cognition[1].

Firstly, we briefly outline the ABSS approach then we give an overview of some relevant PoP literature. We discuss an initial model ontology, based on the concept of "informational artefacts",

[1] We want to make explicit that our goal is *not* to produce models that *replace* the function of a software organisation or developers. We are not here concerned with the *actual* construction, specification or delivery of any *real software systems*. This would be in the realm of software engineering and formal specification languages. We are proposing highly abstract models of some aspects of social processes related to software development within organisations.

and briefly describe how different kinds of organisational structures and cultural phenomena might be captured by such an ontology. Finally we summarise our approach as a complementary guiding method that can be combined with and profit from additional empirical investigation.

2. Agent Based Social Simulation

Agent Based Social Simulation (Moss & Davidsson 2000)[2] is an approach to modelling of social organisations based on the specification of individual agents and the analysis of the resulting emergent social phenomena that arise from their interactions over time. Roughly two kinds of approaches are used in this area. One approach, which might be termed the *artificial* approach, explores the behaviour of interacting agents with artificial worlds. The implementation of agent cognition is often based on intuition (or "off-the-shelf" techniques) and rarely grounded in any actual empirically based theory or evidence. These kinds of exploratory models are sometimes termed "artificial societies" and have some similarity with the models produced within "artificial life" (Langton 1994). Another approach attempts to ground agent cognition, behaviours and interaction scenarios on empirical observations[3].

This latter approach is increasingly being applied to real world scenarios in order to gain a deeper understanding of the kinds of complex interactions that shape real social phenomena[4]. The major use of such models isn't prediction in any quantitative sense (the complexity and contingency of most social phenomena make this a futile exercise) but to gain a deeper qualitative understanding of the process dynamics. Such models allow many "what if" runs in which different conditions, events and model assumptions can be compared (Carley 1998, Terán et al 2000).

What appears to be emerging from attempts to ground models in empirical observation is the need for scenario specific detail. Just how do agents in given scenarios perceive their roles, formulate their goals, interact with others and behave? It seems that no general theory can provide these answers. The only way forward therefore is the detailed study of the actual phenomena of interest. It would seem that some of the existing body of empirically based work that has been carried out in Psychology of Programming (see below) and Distributed Cognition (Perry 1998 and Hutchins 1995) studies could be of great use for informing an agent based model of the process of software construction as *a social process within organisations*. Increasingly, software management theory and practice is recognising the importance of organisational culture as *the* major factor in understanding and improving the efficiency of software organisations (Curtis 1998). As stated later, we will attempt to address the modelling of such cultures based on the conception of "informational artefacts" which may (in some circumstances) be compared to "memes" (Dawkins 1976).

3. Considering a Programming Model

One of the legendry stories associated with a 'programming group' describes a situation where a coffee machine within an organization is removed. The coffee machine formed a place where software developers could informally meet, discuss issues and trade techniques. The removal of the coffee machine would remove an informal meeting place, potentially causing a negative effect on the performance of the group of developers due to the reduction in the level of communication.

This is presented to illustrate the need to consider the importance of the development environment and the range and variation of interactions that can occur between individuals within an organisation. Several studies have been performed to begin to understand how teams of developers function. A fine example being the work

[2] The ABSS community organise a Special Interest Group (ABSS-SIG) within the E.U. AgentLink II Network (see http://www.agentlink.org/activities/sigs/sig4.html).

[3] See the Journal of Artificial Societies and Social Simulation (JASSS) for examples of both kinds of model. The Journal is available free at: http://www.soc.surrey.ac.uk/JASSS

[4] For example, the on-going European wide FIRMA project on water usage. See http://firma.cfpm.org/ for details.

of Curtis (1988) who discusses software development in terms of the individual, team, project, company and business milieu. Curtis indicates that the development of large software systems comprises of the activities of learning, communication and negotiation. It was observed that difficulties became apparent when communication between particular components of an organization were inhibited due to organizational structure. Conflicts would arise and agents would hold opposing beliefs regarding system design. It is not only the organization that is significant, but individuals who make up that organization.

In the psychology of programming literature, programmers have naturally been the target of researchers scrutiny. A growing number of models of software comprehension have been posited and continue to be assessed and considered (Shneiderman & Carrol, 1988). Some researchers have attempted to examine how different categories of knowledge is used during comprehension (Pennington, 1987), and models of how strategy can be affected by domain knowledge. There are fewer models of software production or generation, some of which are described by Davies (1993).

To construct and comprehend software, a programmer is required to perform logical reasoning using short-term memory to encode and store information obtained from a variety of sources. One approach, applied to the study logical reasoning within Cognitive Science, has been to construct software models that aim to simulate logical reasoning, a technique not entirely unlike the modelling approach that is being described. The behaviour and performance of a software system would then be compared with the behaviour of a subject who has carried out similar tasks within the laboratory. Differences between the two can be used to provide evidence in support of a particular model of reasoning that has been essentially formalised in software code.

Understanding and construction of models within the Psychology of Programming can occur through two routes. One approach is the empirical route. Programmers and development teams can be studied both inside and outside the laboratory. Evidence of behaviour can be collected which can then be used to assist in the construction and development of new or existing models, or through the proving or disproving or hypotheses. The second approach could be called the traditional engineering method, where researchers and developers attempt to produce artefacts that are intended to develop and enhance the programmers working environment. Through the application of a new tool or technique, new phenomena or problems may be discovered. This may provide an insight into the needs of the community where the tool is used, suggesting further enhancements and improvements that could be made. Subsequent development incorporates proposed enhancements allowing the process to continue.

The modelling of a software organisation, and potentially, to a limited degree, the agents contained within an organisation is primarily an engineering activity. Models have to be designed and constructed. It is also an empirical process, since models that are built cannot easily be considered to be ‘correct’ in the software engineering sense. A model may exhibit wide variability in its behaviour based upon the parameters and its architecture.

Within the modelling community, models may be proposed and implemented. Results from models would be captured, analysed and then published. A result from one model is entirely at the mercy of its implementation, primarily since first implementation of a model may have guided its design. Since details of the model are published, just like any other empirical techniques, they can be subjected to scrutiny.

4. Information Artefacts[5]

A general rule-of-thumb for any modelling enterprise is to construct a model that is “as simple as possible but no simpler”. In this spirit initial thoughts on capturing the processes of software construction in an ABSS (in the simplest possible way) are presented.

An organisation can be viewed as a collection of agents, interaction possibilities and informational artefacts. By agents we mean individuals or possibly teams with some set of core skills or competences and some ability to learn and adapt. The relationships between the agents determine the

[5] This concept of “informational artefacts” is loosely based on Green (2000) although as presented here probably has more in common with the area known as “memetics” (for example see Speel H-C 1997 and Hales 2001).

interaction possibilities – for example, a manager agent may co-ordinate with other manager agents and delegate to lower level operative agents.

By informational artefact we mean to capture a whole raft of phenomena in a single abstraction. Effectively, we wish to label anything within an organisation that directly affects the performance of an agent in achieving its goals (but that is not another agent or a relationship between agents) as an informational artefact. By this we mean things such as documentation, methods of coding, fixes and workarounds that can be communicated between agents either formally via persistent documentation or informally via advice, guidance or habitual work patterns. For a rich synthesis of recent work, specific examples and a typology of artefacts related to real work practices see Perry (1998), however, we have almost certainly expanded the notion here to cover a wider spectrum of phenomena (further analysis may lead us to restrict it again).

For agents to perform tasks (such as programming) they need to combine intrinsic skills with the correct set of informational artefacts. More precisely, given an agent with some task *T* to perform, the task may require less effort (time) given some associated informational resource (artefact). However, it will also require effort to locate the resource or even to ascertain if such a resource exists.

In the context of the proverbial coffee machine, this can be viewed as an interaction catalyst that facilitates the interchange, trading and replication of valuable informational artefacts. The moral of the story is that the informal nature of the interactions (around coffee breaks) provides an ideal setting for the exchange of informal artefacts. Such informal knowledge may be just as important (if not more so) than formal documentation or procedures.

An idealised hierarchical model of a software organisation might work by submitting a high level task in the form of a set of requirements. The task would be decomposed and passed to agents lower in the hierarchy. This process would continue until a level was reached at which agents with the intrinsic skills and access to the required informational artefacts could complete the task.

Alternatively, consider a more defuse and diverse bottom-up organisation. Here no overall high-level task (or goal) would be submitted at all, but individual agents at the lowest level would collaborate and work on tasks suited to their skills and situation. From this a higher-level goal or task may emerge. Agents identifying the higher-level task may become responsible for bringing together the necessary artefacts to complete the task. The final result of such a process is unpredictable yet highly productive.

In a model of a software organisation we envisage a model where individual tasks may be represented as some set of skill requirements and time based effort requirements (these could be specified as simply as pairs of single values). Agents may have diverse and/or multiple skills. Additionally informational artefacts may be represented as pairs of values indicating associated skill and effort reduction factor. So an agent may complete a task more quickly given access to a relevant informational artefact. Without such an artefact the task may still be completed but with more effort (time). So artefacts may be stored independently of agents, others may only survive by repeated replication through agents' memories.

Some artefacts may be so prevalent and permanent as to become 'ambient'. Examples of ambient information artefacts might be software tools, methodologies version control and configuration management systems. Every developer must have a common appreciation of such tools since they are pervasive throughout the organisation.

In the simplest form, a model could boil-down to sets of value pairs used to represent required skills and effort. Within a model, a specification could be decomposed, relevant skilled agents located and informational artefacts utilised. Particular kinds of organisation structures and individual agent characteristics would affect the success and performance of the organisation in completing the task. The model could be refined to capture more details of a task by increasing the expressive power (using more than simple pairs of values).

This simplistic initial sketch points to a major issue to be resolved: what are the minimal representation requirements for informational artefacts (in the model) to produce any interesting

emergent behaviour (from the model)? Are simple "value pairs" sufficient to produce anything interesting or counter-intuitive when applied to various organisational designs? Perhaps such questions can only be resolved by application to a scenario specific (real world) organisation. It would seem that *a priori* speculation can not be guaranteed to produce answers to that kind of question. We are in the process of examining existing studies to help answer this question or perhaps build new empirical studies. Perhaps the social nature of our perspective requires analysis "in the wild" (Hutchins 1996)[6]. In this context our focus would be on identifying and tracking informational artefacts within organisational practice: When are certain resources used and by whom? For some set of agents in a team what would they rank as their most necessary informational resources for the performance of their on-going work practice? How do agents learn to utilize or create new artefacts? We obviously need to develop or locate appropriate elicitation methodology to take this forward – and this is on-going work.

We distinguish our currently proposed direction from existing modelling frameworks (e.g. the "Brahms" system – see Clancey et al) in that our emphasis is on agent adaptation (learning) and informational artefacts. However, from our initial survey of this work, we can learn much from the approach, particularly the methodology by which real empirical data may be incorporated within an agent based model.

5. Summary

It is proposed that agent based computational modelling of organisations could be used as an exploratory technique to study the activities of software developers. The process of constructing a model requires its designer to formalise many potentially intangible processes that occur and concepts that may be used during the software development. The process of considering each element forces the modeller to address these issues in depth, which may potentially necessitate investigation in areas that may not have previously been considered.

For the software development community, the purpose of studying the modelling of software development should be clear. Experimenting with real groups of programmers, however ecologically valid, is incredibly expensive necessitating laboratories, recording equipment, experienced software developers and projects, see Shneiderman and Carrol (1988). Experimenting, however abstract, with a computer-based interactive model is substantially less expensive but may result in behaviour that may be somewhat removed from the real programming laboratory.

Modelling of software development activities and studying of programmers within a laboratory should be considered as two complimentary empirical approaches, where each technique could suggest or provide a guiding direction to the other.

Acknowledgements

We would like to thank the PPIG workshop referees for their detailed and very helpful comments on a previous draft of this paper. Their pointers and suggestions have already been of great use to us. Thanks also go to our respective institutions for their support.

References

Carley, K. M. (1998) Organization Adaptation. Annals of Operations Research, vol. 75, 25-47.

Chancey, W.J., Sachs, P., Sierhuis, M., & van Hoof, R. (1998) Brahms: simulating practice for work systems design. Int. J. Human-Computer Studies, 49, 831-865.

[6] Although our initial investigations of Hutchins work suggest our perspective (at least presently) is less radical than his.

Curtis, B. (1988) A field study of the software design process on large systems. Communication of the ACM, vol. 13, no. 11, 1268-1287.

Curtis, B. (1998) Which Comes First, the Organization or Its Processes? IEEE Software, Nov. 1998, pp. 10-13.

Davies, S. P. (1993) Models and theories of programming strategy. International Journal of Man-Machine Studies, vol. 39, 237-267.

Dawkins, R. (1976) The Selfish Gene. OUP. Oxford.

Green T. R. G. (2000) Instructions and Descriptions: some cognitive aspects of programming and similar activities. Invited paper, in Di Gesu., V., Leviadi, S. and Tarantino, L., (eds.) Proceedings of Working Conference on Advanced Visual Interfaces (AVI 2000). New York: ACM Press, pp21-29. Available at: http://www.ndirect.co.uk/~thomas.green/workStuff/papers/

Hales, D. (2001) Tag Based Cooperation in Artificial Societies. Unpublished PhD Thesis. Dept. of Computer Science. University of Essex. UK. Availible at: http://www.davidhales.com

Hutchins, E. (1995). Cognition in the Wild. MIT Press.

Langton, C.G. (1994) Editor's introduction. *Artificial Life Journal*, Volume 1, Number 1/2, pages v-viii, The MIT Press, Cambridge, MA.

Moss, S., Davidsson, P. (2000) Multi-Agent-Based Simulation. Lecture Notes in Artificial Intelligence 1979. Berlin: Springer-Verlag.

Pennington, N. (1987) Stimulus structures and mental representations in expert comprehension of computer programs. Cognitive Psychology, vol 19, 295-341.

Perry, M.J. (1998) Distributed cognition and computer supported collaborative design: the organization of work in construction engineering. Unpublished PhD Thesis. Department of Information Systems and Computing, Brunel University, UK.

Shneiderman, B. & Carrol, J. M. (1988) Ecological Studies of professional programming. Communications of the ACM, vol 31, 1256-1258.

Shneiderman, B. & Mayer, R. E. (1979) Syntactic/semantic interactions in programmer behaviour: a model and experimental results. International Journal of Computer and Information Sciences, vol 8, 219-238.

Speel, H-C. (1997) A Memetic Analysis of Policy Making. Journal of Memetics - Evolutionary Models of Information Transmission, 1(2). `Available at:` http://jom-emit.cfpm.org/1997/vol1/speel_h-c.html

Terán, O., Edmonds, B., and Wallis, S. (2000) Mapping the Envelope of Social Simulation Trajectories. In Moss, S., Davidsson, P. (Eds.) Multi-Agent-Based Simulation. Lecture Notes in Artificial Intelligence 1979. Berlin: Springer-Verlag (available at: http://cfpm.org/cpmrep72.html).