Psychology-based Agent Architecture for Whole-of-user Interface to the Web

Steve Goschnick[1] and Leon Sterling[2]
IDEA Lab[1] & Agent Lab[2]
The University of Melbourne
stevenbg@unimelb.edu.au; leon@cs.mu.oz.au

Keywords: Models of Mind, Psychology, Agent Architecture, User Model, Interface Agents,

Abstract

This paper argues that the user interface of a workstation connected continuously (24x7) to a network would be most effective with a sophisticated agent architecture embedded deep in the workstation system software. In the user's absence an embedded agent system could act as something more than a proxy for the user, the multiple sub-agents within it should act in concert as a Digital Self, one representing and empowering the user. Our proposed agent architecture, called Shadowboard, is based on a sophisticated model of the user drawn from the Psychology of Subselves, a modern stream of Analytical Psychology.

1. Introduction

What we want as users of computers connected to the Internet 24x7 (24 hours a day, 7 days a week) is an *interface to the network,* rather than simply a GUI (graphic user interface) interface to the computer. We want to be notified of information that is important to us, but be spared the distractions of the mundane. We want to be accurately *represented* in our global community as we sleep, and be able to take in information and process it subliminally from the net, as we go about our physical lives away from the wire. We want software to take the initiative and make decisions on our behalf, without waking us from our sleep or distracting us from the things we prefer to consciously focus on, but only those decisions that can be clearly made, eg.*"No I can't go to the Sydney concert, I'll be in Melbourne that day."*

In the past decade there have been many people building Interface Agents, Information Agents and Personal Assistant Agents (Maes, 1994), addressing many of these computer system shortcomings but simply adding them to the workstation operating system as auxiliary applications, without regard to architecture. In this paper we take the view that a sophisticated agent architecture, one suitably designed to include user interface functionality, should be embedded deep in the computer system software with as much status as, and well integrated with - the GUI itself. We have named our agent architecture Shadowboard.

We will build upon this architecture a central controlling agent together with numerous sub-agents, forming a singular *whole agent* that is a sophisticated representation of the user in 24x7. We have called this enacted complex agent the *Digital Self* (Goschnick, 2000; 2001). The sub-agents within the Digital Self may have visual personifications within the GUI - i.e. visually displayed agents within the genres known as Believable Agents (a.k.a. Anthropomorphic Agents) (Elliott and Brzezinski, 1998). Our research question has been one of finding a sufficiently complex model of mind for a Digital Self, then modeling an agent architecture upon it.

2. Background

Software agents come in many forms with numerous categorical names, including: reactive agents, deliberative agents, rational agents (a.k.a intelligent agents), BDI agents, interactive agents, adaptive agents, interface agents, believable agents (a.k.a anthropomorphic agents), emotional agents, mobile agents and multi-agent systems. The

origins of all of these categories, can be traced to one of two domains of computer science: AI system software (backend technology); user interface software (client-side technology). Both these fields have been interested in models of mind: the AI system people primarily in building synthetic intelligence; the UI client-side people in making software more usable for human users.

2.1. User Interfaces, Mental Models and Interface Agents

The modeling of a user's thought processes has been a central aspect of user-centric interactive research from early on. Norman (1988) proposed a seven stage *action model* of people *doing things* (see Figure 1). He termed this model an approximate psychological model rather than a complete psychological theory, conceding that in practice the stages are not discrete, nor necessarily sequential and that most behaviour does not go through all stages. Nonetheless, the model is well known and is the theoretical basis of several formalisms in HCI.

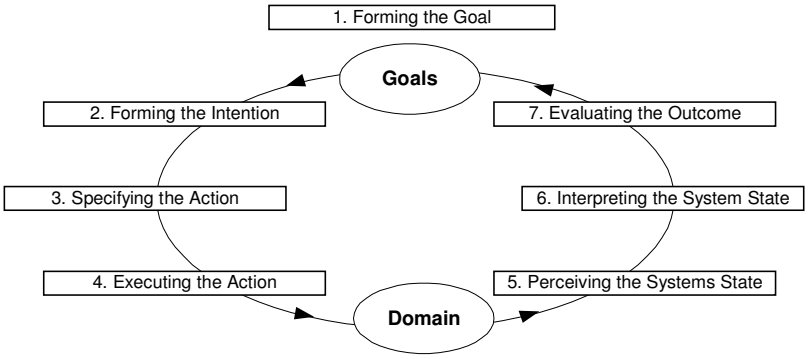

Figure 1. Norman's 7-stage Action Model.

Norman's model is one of *action*, of the human user getting particular things done framed in the context of the user interface. However it lacks *deliberation*, except in the sense of evaluating past, less than successful actions. It is a model about *process* and as such it lacks a model of the human mind at any deeper level, instead assuming that a human mind is ever present throughout such situations.

When a user translates their intentions into actions played out within a virtual environment on the screen, they have a *mental model* of the application and of how their physical actions put plans into effect within the virtual environment. When the HCI specialist talks of *mental models*, it is usually these *models of the application* the user holds in their head that concerns them. The mental model we aimed for within the Digital Self is the agent's *mental model of the user*.

2.2. The Agent Oriented Paradigm and BDI Agents

Developments on the system software side of agents grew out of the desire to built autonomous systems, synthetic intelligence, in particular - robots. Though autonomous agents need not be constrained by the physical body of a robot, they can still be thought of as software robots. The following contemporary definition of a deliberative agent by Wooldridge and Jennings (1995) is well subscribed to in the AI/Agent community:

> *An agent is an entity situated in a changing environment that continuously receives perceptual input and reacts by taking actions. Its choices of action are rational and are autonomously determined by calling upon internal mental state, consisting of beliefs, plans, goals and intentions.*

In short, agents are: persistent, autonomous, reactive, pro-active, flexible and social (interacting).

How do software agents differ from other programs? Where a traditional program has *inputs* and *outputs*, an agent has *perceptions* and takes *actions*. Where the program holds facts, the agent holds both *facts* and *beliefs* and is therefore capable of execution with incomplete knowledge. Where the program usually stops or concludes when inputs cease, the agent runs continuously in an open system, autonomously, not just reacting to inputs. An agent has an internal model of its world (the environment it is situated in), and it expects that the actions it takes will change the environment in a manner that matches its own goals. It is not only the environment that is dynamic, the agent's own internal state is open to change, including its beliefs, goals, intentions and how to achieve them. Figure 2 below represents the BDI (Beliefs, Desires and Intentions) agent model.

The BDI agent architecture (Rao and Georgeff, 1992) is one of a number of architectures that enact deliberative agents. BDI calls upon the mentalistic notions of *Beliefs*, *Desires* and *Intentions* from folk psychology, as abstractions to encapsulate the hidden complexity of the inner functioning of an *individual* agent. Folk psychology is the name given to everyday talk about the mind, the vocabulary we use to explain ourselves such as: *beliefs, desires, intentions, fears, wishes, hopes*. As people, the use of such language gives us an efficient method for understanding and predicting behaviour.

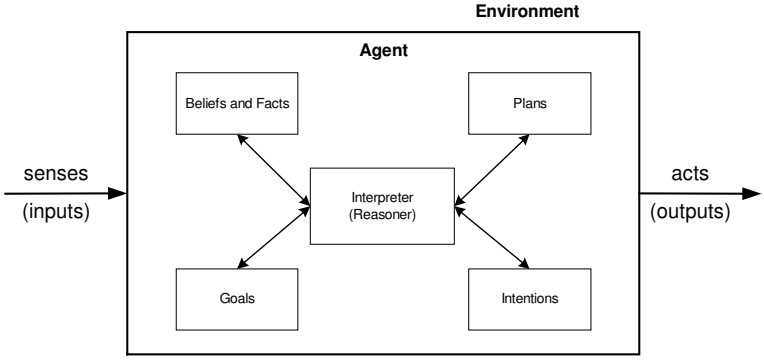

Figure 2. BDI Agent Architecture.

A BDI agent is an operational computation of folk psychology, with the intention of giving us an efficient method of building agents that are understandable and predictable.

3. Psychology of Subselves as Model

Theoretical design in HCI has been driven by cognitive psychology, ecological interface design and other psychological aspects of interaction. What we have aimed for here is a complex *mental model of the user*, one that is complete enough to manage and empower the user in most aspects of their life, in a 24x7 mode of operation. The design theory needed here is a higher order, as the technology it enables is *affective*. The BDI model of mind drawn from folk psychology is too course-grained, acceptable for building a simple synthetic agent but too simple to model a human user for our purposes. Nonetheless, taking direction from both HCI and Intelligent Software Agents, we looked to psychology as fertile ground for further models of mind.

Western psychology has many rich branches other than folk psychology from which one could draw models of agency including: *cognitive psychology, analytical psychology, humanistic psychology, developmental psychology*. Shadowboard, the agent architecture we have developed draws upon *analytical psychology*, in particular, a contemporary stream encompassing refinements of Freudian (Sliker, 1992), Jungian (Jung, 1964) and Assagiolian (Assagioli, 1965) concepts. The approach is known as the *Psychology of Subselves,* an attempt to understand the whole personality of an individual - in order to model consciousness, deliberation and action (Sliker, 1992; Stone and Winkelman, 1989).

Main exponent:	**Sigmund Freud**	**Carl Jung**	**Roberto Assagioli**	**Hal Stone & Sidra Winkelman** (Psychoanalysis & Psychosynthesis)
Technique	Psychoanalysis		Psychosynthesis	Voice Dialogue
Model divisions of the human psyche	Ego	Persona	Centre	Aware Ego
		Self, self	Self	
	Super Ego	Higher Self		Protector/Controller Inner Critic
	Id (repression)	The Shadow		Several *Disowned Selves*
		Anima/ Animus	Many sub-selves	Pusher, Pleaser, Parental selves, many other subselves
		Archetypes	Evolved subselves	

Figure 3. Lineage of subpersonality exponents and some of their divisions of the psyche.

Figure 3 above represents a scant overview of the divisions of the psyche that different influential analytical psychologists identified and named. Beginning with Freud about 100 years ago, analytical psychologist identified substrata within the psyche of an individual. As we move across the columns in Figure 3 from left to right, we move forward right up to contemporary theory on the sub-selves within an individual psyche. The most easily identified sub-selves of a person usually align with the *roles* they play in their lives – teacher, parent, protector, and so on. A comparative description of the different psychologies we considered and the subsequent development of Shadowboard is detailed extensively elsewhere (Goschnick, 2001).

4. Architectural Overview of Shadowboard.

Figure 4 below, depicts the Shadowboard architecture, collectively representing an individual *whole agent* made up of numerous sub-components – the structural implications are inferred from the Psychology of Subselves. In the centre of the agent is the *Aware Ego Agent* – the dominant sub-agent in the whole cluster of sub-agents. In the figure, the Aware Ego Agent is surrounded by eight first-level sub-agents, diagrammatically draw as circles the same size as the Aware Ego Agent. Five of these example sub-agents are not nested any deeper (i.e. sub-agents can be clustered recursively), while the other three have clusters of circles within them, representing a second-level of sub-agents, grouped into numerous *envelopes-of-capability*.

Each *envelope-of-capability*, of which there are arbitrarily eight in the figure, represent different areas of expertise that a particular whole agent embodies. As such the whole agent could perform a number of consecutive and diverse tasks, depending on what goals via what roles it has taken on in the outer world.

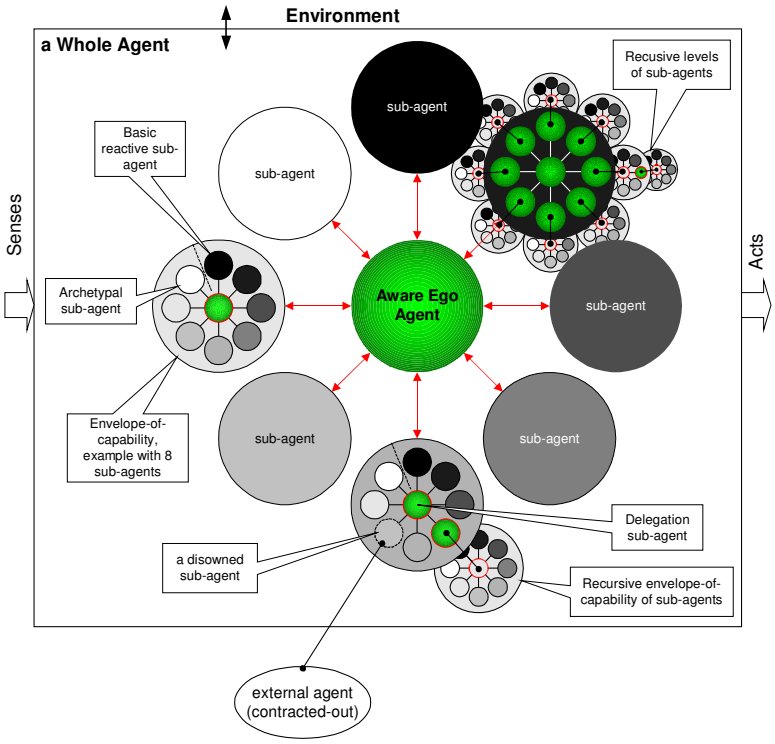

Figure 4. The Shadowboard Architecture.

Each envelope-of-capability contains a number of sub-agents with similar capabilities, but each different from it neighbours in some specialised way. At the two far ends of an envelope-of-capability are two diametrically skilled agents - in Figure 4 the two are adjacent, one is white the other is black, separated by a dotted radial line:

One is the *basic reactive sub-agent* a purely reactive agent, with a hard-wired rule-action mechanism and no deliberative capability. The basic reactive sub-agent is called upon within a particular envelope-of-capability when time or other resources are severely constrained.

At the other end of the scale is an *archetypal sub-agent*, one that has maximum deliberative ability used when time and other resources are plentiful.

The *aware ego agent,* as well as each of the other sub-agents (*delegation sub-agents*) that are shaded as spheres, have *knowledge about the capabilities* of sub-agents in their envelope-of-capability. They use this knowledge to select the appropriate sub-agent to achieve the particular goal that has been sent their way, from higher up the recursive hierarchy.

When a sub-agent has been found lacking in capability to achieve a specific intention, or when an external (and available) agent matches the particular specialty better than any internal sub-agents, an external agent can be called upon as if it were an internal sub-agent. This process is termed *disowning a sub-agent*.

4.1. Sub-Agents as Subselves

The mentalistic notion of a *subself* at work within the psyche of an individual, is a metaphor for the *sub-agent* of a Shadowboard agent. To broadly place this work in context of research upon multi-agent systems (MAS) – the closest comparative field of study - most multi-agent systems can be described as *inter*-agent systems. In contrast, the Shadowboard theory and architecture is an *intra*-agent system, one enabling the incorporation of many components that together represent one *whole* agent, albeit a very sophisticated one. Such a *whole agent* built upon the Shadowboard architecture – e.g. the *Digital Self* – seen from without, should be seen as a fully *autonomous* individual agent, one compatible with existing definitions of agency, such as the definition in Section 2.2 above.

Unlike the autonomous whole agent, the inner sub-agents are semi-autonomous or even totally subservient to the *Aware Ego Agent* - the executive controller within a Shadowboard agent. Sub-agents may themselves be sophisticated agents capable of their own semi-autonomous work, or they may be Active Objects or even conventional application programs. This is a significant relaxation on the need in most MAS systems for each individual agent to be a fully functioning autonomous agent. This flexibility in capability is provided within a Shadowboard agent, is attained by making all sub-agents ultimately subordinate to the *Aware Ego Agent*.

The inclusion of sub-agents and envelopes-of-capability of sub-agents within the Shadowboard architecture, allows it to be populated with all manner of domain specific and types of sub-agents, allowing for open-ended expansion of capability and knowledge.

We used the XML DTD language (eXtended Markup Language Document Type Definition - for defining XML schemas) to render the *structural model* of Shadowboard into syntax widely readable by increasing numbers of both humans and systems. This enabled us to *declare* a Digital Self agent based on the Shadowboard architecture using an *.xml* file. The *Shadowboard.dtd* file and the full declaration of a *DigitalSelf.xml* running to many pages, are outlined elsewhere (Goschnick, 2001). That document also outlines the system implementation strategy for realizing the architecture, which is currently well advanced.

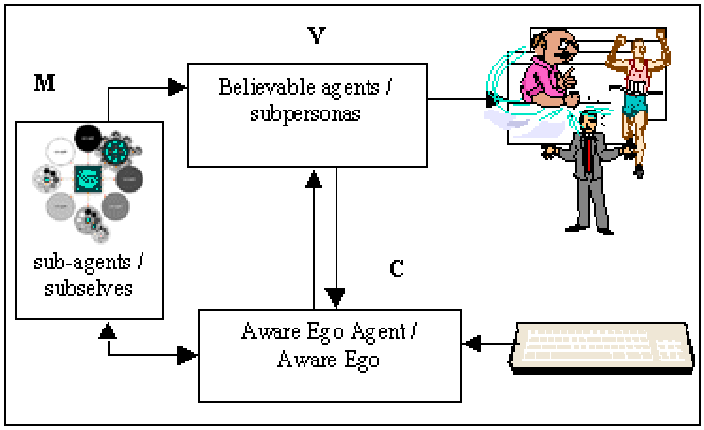

Figure 5. Shadowboard Architecture portrayed as AoMVC.

4.2. Shadowboard agent system meets the GUI

Existing operating systems have GUI's (Apple Mac OS X, Microsoft Windows, MOTIF) with which the Shadowboard architecture needs to co-exist. To explain how Shadowboard meets the GUI we mapped the

psychological model behind Shadowboard into the well-known MVC (Model, View, Controller) model, and arrived at a representation of Shadowboard that has hybrid UI/Agent descriptive features, dubbed Agent-oriented MVC (AoMVC). See Figure 5 above.

The MVC user interface architecture first appeared as the model of interaction for the Smalltalk80 language (Goldberg, 1983). It was devised as a particular strategy for representing, displaying, and controlling information. The more recent Java language uses the MVC model to implement its *pluggable look-and-feel*. The MVC architecture consists of three main conceptual elements as follows: Model -*stores structure and state data;* View - *deals with display of data;* Controller - *synchronizes data between model and view, schedules interactions, and enacts events.*

5. Progress and Reflections

We have modelled the Shadowboard architecture upon the Psychology of Subselves in order to get a sophisticated model of the user, and have drawn many structural implications from it. We have specified the Shadowboard architecture in an XML DTD file and we have declared a simple generic Digital Self in XML which conforms to that Shadowboard.dtd which runs to seven pages elsewhere. We have a Systems Requirement Specification to implement Shadowboard upon an existing system event-driven system called SlimWinX, replacing the event queuing system with a goal and constraint queuing system. These milestones are fully detailed in Goschnick (2001). Development of the Shadowboard implementation is proceeding well and is expected to be operational in the third quarter of 2002.

The degree that a user's life will be genuinely empowered by this agent system is significantly dependent on two things: the availability of numerous sub-agents which genuinely service the broad range of needs and interests of an individual; and secondly, the extent of self-knowledge that the user has.

The first, which at first glance seems like a significant limitation, can be overcome by using third party agents and other applications as sub-agents – as Shadowboard sub-agents do not have to be autonomous or individually sophisticated: it is the collective *whole agent* which conforms to the conditions of agency set out in Section 2.2. The second issue has wide variance across individuals, but the very process of specifying and then using an extensive Digital Self, will probably lead the user on a path of growth. The contemplation of one's possible subselves, the selection of particular sub-agents one would employ and then the circumspection of them in action, will probably lead the user to some of these following things that agent researchers have been catering for in their agents: *goal revision*, *belief revision*, *intention revision* and *plan selection* – as this is clearly affective technology. There is a complementary dichotomy between agent learning from humans and human learning from agent simulations. With Shadowboard the user learns about Self from an operational simulation of Self.

6. References

- Assagioli, Roberto (1965) Psychosynthesis. (Viking: New York).
- Elliott, C. and Brzezinski, J. (1998) Autonomous Agents as Synthetic Characters (*AI Magazine*, American Association for Artificial Intelligence, pp13-30, Summer)
- Goldberg, A. (1983) Smalltalk-80: The Interactive Programming Environment (Addison-Wesley)
- Goschnick, S.B. (2001) Shadowboard**:** an Agent Architecture for enabling a sophisticated Digital Self (Dept. of Computer Science, University of Melbourne: M.Sc.Thesis, 180 pages)
- Goschnick, S.B. (2000) Shadowboard: A Whole-Agent Architecture that draws Abstractions from Analytical Psychology (Proc. of Third Pacific Rim International Workshop on Multi-Agents *(PRIMA 2000)*: Melbourne)
- Jung, C.G. (1964) Man and his Symbols (Aldus Books)
- Maes, P. (1994) Agents that reduce workload and information overload (Communications of the ACM, 37:7)
- Norman D. (1988) The Psychology of Everyday Things (Harper Collins Publishers)
- Rao A.S. and Georgeff M.P. (1992) An Abstract Architecture for Rational Agents (Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning, 439-449)
- Sliker, Gretchen (1992) Multiple Mind - Healing the Split in Psyche and World (Shambhala Publications)
- Stone, Hal and Winkelman, Sidra (1989) Embracing Ourselves (New World Library)
- Wooldridge, M. and Jennings, N.R. (1995). Intelligent Agents: Theory and Practice (Knowledge Engineering Review, 10:2, pp115-52)