Shadowboard: A Whole-Agent Architecture that draws Abstractions from Analytical Psychology

Steve Goschnick

Intelligent Agent Lab
Department of Computer Science
University of Melbourne, VIC, 3010, Australia.
gosh@cs.mu.oz.au

Abstract. This paper presents an intra-agent architecture called Shadowboard, one that takes abstractions from analytical psychology. The Shadowboard architecture is a foundation upon which to build a whole-agent - an individual autonomous agent no more, but one made up of many sub-agents. Such a whole-agent approach to modelling enables a psychologically sound, finer-grained approach to applying behavioural abstractions such as BDI, while incorporating the selection of capabilities and plans, together with learning and optimization. An individual agent built upon Shadowboard is also capable of collaboration and cooperation in a wider MAS system. The strong degree of *self-awareness* that a Shadowboard agent intrinsically has, not only allows it to improve its own performance and effectiveness over time, it also offers significant advantages in modelling other agents in an encompassing MAS system.

1. Introduction

The BDI Agent architecture calls upon the *mentalistic* notions of *Beliefs*, *Desires* and *Intentions* as course abstractions to encapsulate the hidden complexity of the inner functioning of an *individual* agent. BDI is a course-grained approach to the use of such mentalistic notions, and as such, a starting point for intra-agent modelling.

Others have expressed the desire for individual agents to be more psychologically sophisticated, more dependent on human psychology, so that they may function and interact more effectively within our human social systems [23, 12]. Watts' aim is for agents and humans to cooperate and otherwise socially interact, in more effective and useful ways. He would like to build agents that interact with humans and can stand in for humans - sophisticated Interface Agents and Intelligent Personal Assistant agents. Such a definition of agency, draws upon human social intelligence as an ideal rather than just as a metaphor, and hence upon an individual person as a psychological *archetype* for an individual agent.

In Society of Mind [14], Minksy takes a reductionist view of the human mind in which every describable process (eg. Add, Move, Grasp) is considered an agent, so that a brain would be made up of many millions of such agents. When he did discuss the higher constructs within the mind he generally referred to cognitive functionality,

such as: that large part of the brain concerned with vision. Nonetheless he did discuss higher levels of organisation of mind, such as the *Self* and the *Conservative Self*. Yet he was troubled with *'hidden'* aspects of mind, most evident in a section he titled: *Self Knowledge is Dangerous* - a point of view that is contrary to cross-discipline and cross-cultural wisdom.

Suppes et al [22] describe *representation* - in the sense of modelling within psychology - as a way of reduction of a more complex structure. They discuss two forms of *reduction* that employ representation: one characterizes a set of theoretical concepts with another, giving the example of Descarte's reduction of geometry into algebra; the second, describes (noisy) data in context of parameters that capture the main tendencies, in a pattern recognition fashion. BDI, which takes advantage of the reduction of *behavioural concepts* into *mental concepts*, is an example of the first. The approach outlined in this paper, incorporates both the first and second forms of representation, to reduce the complexity of modelling an agent - an approach based on finer grained patterns of *behaviour of subselves* within the whole self.

A finer grained *intra-agent* model based on psychological notions, should result in a much stronger degree of agent self-knowledge (*self-awareness*), which in turn should improve an individual agents performance within a social MAS environment. Western psychology has many rich branches from which one could draw models of agency: cognitive psychology, analytical psychology, humanistic psychology, developmental psychology. The psychological foundation of BDI is behavioural folk psychology [17]. Shadowboard, the agent architecture presented in this paper, draws upon *analytical psychology*, in particular upon contemporary refinements of Freudian [21], Jungian [11] and Assagiolian [2] concepts.

Shadowboard uses abstractions of mentalistic notions based on well documented and clinically supported psychology involving *subselves* (also known as *subpersonalities*), at work within the psyche of an individual. To broadly place this work in context of research of multi-agent systems: most multi-agent systems (MAS) can also be described as inter-agent systems; the Shadowboard theory and architecture by comparison is an intra-agent system. Nonetheless, Shadowboard has psychologically founded *handles* in the architecture that allow a whole-agent based on Shadowboard, to collaborate within a more general MAS infrastructure. Also, a *whole agent* built upon the Shadowboard architecture, should be seen as a fully *autonomous* individual agent compatible with existing definitions of agency, such as that by Wooldridge and Jennings [25]. This is in contrast to the inner sub-agents representing subselves, which are only semi-autonomous or even totally subservient to the *Aware Ego Agent*, which is the executive controller within a Shadowboard agent.

It is evident from the literature on multi-agent systems, particularly amongst those papers based on the practicalities of building MAS systems, that inner sub-agents are often constructed and intra-agent communication is then supported, in some of those systems [15,3]. As with the sub-agents in Shadowboard, those sub-agents are also either semi-autonomous or totally subservient to a primary agent. Whereas the sub-agents in these other systems have been introduced for practical implementation reasons, or for representing specific distinct Personal Assistant agents, in Shadowboard, sub-agents are *first-class entities* based directly on subselves - mentalistic entities in analytical psychology.

The name Shadowboard draws from a tool-based metaphor for user interface. (see Figure 1). It is included at this point, as a conceptual aide in understanding the architecture. However, Shadowboard doubles as a psychological metaphor with regard to subselves, as will be described further down. Note. The user-interface (UI) aspects of Shadowboard as an agent-oriented UI metaphor, are the subject of another paper [8], in which the Shadowboard architecture is presented as an Agent-oriented Model-View-Controller (AoMVC) UI architecture (see Figure 2).

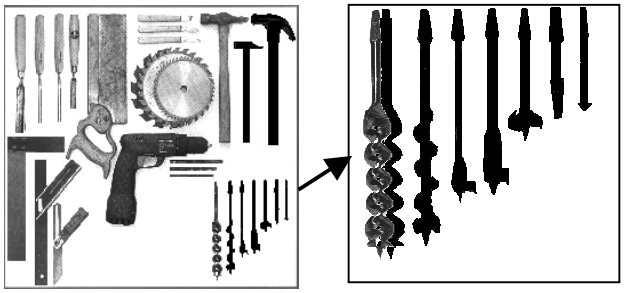

Fig. 1. A workshop tool-based shadowboard, as metaphor.

In section 2, a necessary overview of the *psychology of subselves* is presented, as it effectively represents the *method* behind the architecture - hopefully described in enough detail to lay bare the roots of Shadowboard, yet laymen enough to be comprehended without necessitating further reading in psychology. In section 3, I describe the elements of the Shadowboard architecture, resulting from a mapping of the psychological terms and concepts, to the agent terms and concepts of the Shadowboard architecture.

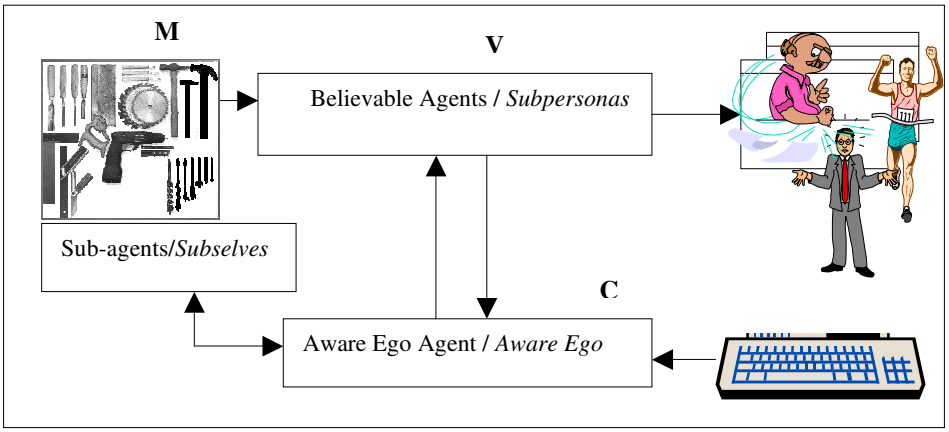

Fig. 2. Shadowboard as an *Agent-oriented* MVC Architecture (AoMVC)

In section 4, I describe the *Aware Ego Agent* within a Shadowboard-based whole-Agent, as the *executive decision maker* and discuss its other primary functions. In section 5, I compare the resulting intra-agents with agents in a multi-agent system (inter-Agent activity), drawing upon some facets of multi-agent research such as:

inter-agent cooperation; joint goals; joint intentions; plans; roles; social commitment - and examine the fit of those ideas to the Shadowboard multi-sub-agent architecture.

In conclusion, we revisit the discussion above with respect to agent modelling and the modelling of mind, looking afresh at Watt, Minsky, Suppes and others, in light of the results in the form of Shadowboard.

2. Psychology of Subselves (sub-personalities) as Methodology

The sub-agent approach in Shadowboard, is modelled on sub-personalities within the individual - a contemporary approach in psychology to understand the whole personality - in order to model consciousness, deliberation and action [2,19,20]. This section aims to give a brief description and background to the *psychology of subselves*, as it represents the *method* underpinning this research. Note: the term sub-personalities is used interchangeably with subselves in this paper and elsewhere.

Consider the voice (and gestures, and language) a person uses when talking to a child, then compare them to those they use when taking to one of their own parents. In such exchanges it is often possible to glimpse the facets of two of the subselves within the psyche. The subselves within, often accompany different roles a person has in their outer life, but not always. Also, we get to see very few of the subselves, in the external persona of an individual. Sliker [19:32] noted: *'Jung described in detail the persona, the personality mask developed to meet the world safely. Subpersonality knowledge reveals the extreme limitations of the persona. Usually one or two subpersonalities perform the function of persona, while perhaps the rest of the personality is rarely seen in public.'*

The lineage of subselves in psychology is scantly represented in Table 1 below, not to under-appreciate it, nor to strictly categorise one against the other, but to indicate that the psychology of subselves underlying the Shadowboard architecture, though based on relatively modern developments in psychology, has been evolving since the start of modern western psychology.

Sliker points out that in 1907 Freud, then aged 51, Jung, then aged 32 and Assagioli, then aged 19 all met, and she observes that: *'Although their careers overlap, Freud, Jung and Assagioli quite literally represent three generations of thought on subpersonality.'*

Psychosynthesis

While Freud and Jung were basically dealing with pathology in their practising lives, Assagioli was most interested in the human development of *'healthy'* individuals. The advocates of each, still lean in those directions: Jungians are generally occupied with building *strong foundations* by uncovering flaws and misplaced energy in the psyche, while Assagiolians are more intent on *'building splendid skyscrapers'* upon assumed solid foundations, under the heading of Psychosynthesis. Although Assagioli expressed his ideas of a synthesis of the subselves in the psyche as early as 1909, they only found a ground-swell in adoption within the humanistic psychology movement of the 1960's and 70's.

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
	Id (Repression)	The Shadow		Several *Disowned Selves*
	Klein, Fairburn:	Anima/ Animus	Many sub-selves	Inner Critic, Pusher, Pleaser, Parental selves, many other subselves
	Mental Objects	Archetypes	Evolved subselves	

Table 1. Lineage of subpersonality exponents and some of their divisions of the psyche.

Given that pathology is not something one would normally build into an agent, it might seem tempting to base an agent model simply upon the Assagiolian principle of subselves - one devoid of any aspect of Jung's concept of the *Shadow*. To emphasise the point, in her coverage of Psychosynthesis, Sliker sees the goals of Psychosynthesis as the transformation of the subselves into *well-polished tools* that can bring about effective action. On the surface then, Psychosynthesis seems like a good psychological metaphor to match with the graded tools-oriented metaphor represented in Figure 1.

Voice Dialogue - a Psychology of Subselves

Jung's concept of the Shadow, the part of personality that Assagiolians choose to gloss over - represents facets of personality *repressed* or *projected*. It is not only an interesting aspect of people with respect to explaining and predicting their overall behaviour, it is also the engine-room behind relationship-building between individuals, dysfunctional or otherwise, in the social world. Therefore, with respect to *inter*-agent activity between individual agents based on Shadowboard, an Assagiolian approach is not enough. So, the model of subselves adapted for Shadowboard, is from a psychotherapy technique called *Voice Dialogue* developed by Stone and Winkelman [20], which is founded on Jungian analytical concepts, but has also drawn refinement from Psychosynthesis, in an effective synthesis of those earlier bodies of psychology. [Note: We are not interested here in the therapy aspect of any of the psychologies, just the *models of mind* they give us with respect to deliberation, cooperation and action.]

Stone and Winkelman identify and name several generic subselves that they can readily identify in most people: *Protector/Controller; Pusher; Inner Critic; Pleaser; The Perfectionist; Inner Child; Parental Selves*. Without going into the explicit definitions here, the names themselves are sufficient to allude to their functions. There

are many other less dominant subselves a given person will have developed, or is still prototyping, to fulfill the roles they carry out as they carefully negotiate their way in the outer world.

In addition to their own agendas, the inner subselves often work in cooperation, in small teams, for example: the *Inner Critic* might make a person feel bad about not knowing enough, then the *Pusher* will step in by providing a reading list to help the person improve themselves. The Inner Critic is also using *archetypes* as ideal versions of a subself, against which to measure and judge the individual's worldly actions and thoughts. In this sense, Jung's archetypes, are like mentors towards which an individuals beliefs and actions, are tailored.

Stone and Winkelman realised that the various facets of Jung's 'Shadow', were better understood as a group of *Disowned Selves* - subselves that have been disowned through bad experiences in the past: via behaviour that wasn't socially rewarding; or behaviour which exposed the individual to danger in the past. It is not always bad behaviour, it could have been something that a person was naturally good at, but which caste them into the limelight (eg. dancing), which in turn made them vulnerable (eg. peer-group pressure). So, *disowned selves* are the result of an individual surviving and evolving within the social system that is the society about them.

However, these old aspects of self are not just *disowned*, they are also *projected* onto other people. The disowned selves with a negative (anti-social) energy, are usually projected onto people that an individual strongly dislikes, someone who exhibits the disowned traits. The positive disowned energies are most often projected onto a partner or other friend, and hence play an important part in relationships. Continuing the example of the person with a *disowned dancer*, it might be projected onto a partner or potential partner who is very comfortable at such performances. Another *disowned self* sometimes projected onto a partner, is the *Inner Critic*. We will see further down, that the concept of disowning a subself, is a very useful analogy for the way in which sub-agents are put on hold, in Shadowboard. Disowned selves also give us a psychologically inspired handle, for inter-agent relationship building.

Aware Ego

One of Stone and Winkelman primary advances is on *awareness*. Historically, in psychoanalysis the Ego is seen as the executive function of the personality - *the decision maker* - but it is also seen to be in control of awareness, at least during consciousness. Stone and Winkelman saw the need for individuals to actively separate the *awareness* aspect of mind from the *control* part. To them, awareness is clear space, that just *witnesses*, not attached to outcomes. They see a successful result of their work on a person, as someone who has developed an *Aware Ego* - an executive decision maker, that calls upon a purely awareness function of mind, one in full knowledge of all the subselves and able to call upon them individually or in teams, to negotiate the challenges of life in an optimal manner.

3. Applying the *Psychology of Subselves* to Shadowboard

Shadowboard represents the *result* of mapping the *psychology of subselves* as a *methodology*, upon agent concepts, to achieve a new agent architecture. This section discusses that mapping.

The metaphor of Shadowboard is drawn from the concept of a shadowboard in a workshop, used to store, locate and return tools. It is a physical representation of the saying: *a place for everything and everything in its place.* Referring to right-hand-side of Figure 1, is a class of tools named drill-bits. Only one of the drill-bits is present, the rest are off somewhere presumably being used for current tasks. The Shadowboard architecture has classes of sub-agents, analogous to classes of tools such as the drill-bits class. The psychological equivalent to a *sub-agent* is the *subself*. A class of sub-agents as such, represents an *envelope of capability* that a whole Shadowboard agent has. In Figure 2, which represents Shadowboard as an AoMVC - an Agent-oriented Model-View-Controller (MVC) UI architecture - you can see the following mappings: the data Model is a repertoire of sub-agents, which represent the subselves; the Controller is a primary sub-agent called the *Aware Ego Agent* (covered in the next section) which maps to psychology's *Aware Ego*; while the Views, which could be as simple as standard GUI windows or sophisticated Believable Agents [7], represent *subpersonas* - the personality masks of the inner sub-agents, as they chose to present themselves on the screen to human users.

Whole classes of Shadowboard sub-agents are enacted as required for a given agent being built. For example if an Agent has taken on a primary role as an Engineer in some larger scheme, it might include a *mechanical engineer sub-agent*, a *materials engineer sub-agent*, a *chemical engineer sub-agent*, a *construction engineer sub-agent*, an *information engineer sub-agent,* etc. Within each type of engineer sub-agent, there may be a further *envelope of capability*, for example: one *construction engineer sub-agent* may be optimised for speed; while another, optimised for quality of workmanship. To the *quality control sub-agent* concerned with overall quality of resulting workmanship, the sub-agent optimised for speed will look unpolished. Whereas to a *time-management sub-agent*, the one optimized for speed may look like a polished version of one oriented for quality.

In personal development work in psychoanalysis, disidentification is aimed for, by stepping back from the personality, describing it and thereby objectifying it. Disidentification in agent terms, equates to differentiation in capabilities and functions of sub-agents. Classification of individual sub-agents within the architecture, is largely done via their built-in capabilities.

At the micro level - the level of the sub-agent - Shadowboard agents are BDI. The subselves in psychology, contain personal histories: the individuals memory of war stories and the emotions that go with them, at the subself level. They have inclusive safety rules, based on those past histories for use in the future. Following suite, the *beliefs, intentions* and *goals* (goals taken as: *desires with intention*) of a Shadowboard agent, are stored down in the lowest level of entity: the sub-agent - yet they are all accessible globally by the Aware Ego Agent. The behaviour of subselves is pattern oriented. In the agent, this equates to *stored plans* at the sub-agent level, but which the Aware Ego Agent also has access to at the global level, in addition to those of its own.

The concept of *disowning a su*bself is mapped to the way a Shadowboard agent puts a sub-agent on hold. Those sub-agents not in use for any current task, are put back on the shadowboard, effectively on standby (note: Shadowboard agents are capable of concurrent execution of sub-agents, and teams of sub-agents). None of the sub-agents within a Shadowboard agent are considered malevolent – they are known and have been incorporated into the whole agent, for what they are; and their particularly skills are appreciated for whenever the *Aware Ego Agent* needs to enact them - discussed fully in the next section.

Disowned selves also give Shadowboard a psychologically inspired handle for *inter*-agent relationship building, when the whole agent is encompassed in a MAS system. If the Aware Ego Agent is not confident that its *best-fit* sub-agent (the sub-agent with capabilities closest to matching the task) is up to a task, then it may go outside the whole agent, looking for a more skilled external agent. In both cases, whether for inactivity or as a relationship hook, the concept of a *disowned self* is a useful psychological metaphor to the Shadowboard architecture to account for the suspended state of a sub-agent. And by embracing the concept, Shadowboard maintains integrity with the underlying Jungian psychology with respect to the Jung's concept of the *Shadow*, both in name and in function.

4. The *Aware Ego Agent*: Shadowboard's Executive Director

The *Ego* in the traditional psychoanalysis sense, is the *executive director* of the psyche, the decision maker. In Voice Dialogue, the Ego is viewed as a combination of dominant subselves acting in collaboration, for example: the Protector/Controller; the Pusher, the Pleaser; Responsible Father, and possibly others. Such a dominant cluster of subselves is working with a limited, incomplete set of survival goals. Other subselves of which the Ego is unaware, are probably working at a subconscious level only. Thus an unaware Ego will be inefficient and suboptimal in most scenarios faced in conscious life. Alternatively, an *Aware Ego*, one fully aware of all a persons subselves, including previously disowned selves, is one making optimal choices of subselves, individually or in combination, to handle any given situation.

The *Aware Ego Agent* in Shadowboard, is the agent equivalent of the attained Aware Ego in a person: by definition a fully *self-aware* entity. It is the primary sub-agent within a Shadowboard-based whole Agent, with a full knowledge of all sub-agents which make up the whole agent. It is the executive decision maker responsible for selecting individual sub-agents, or teams of sub-agents to handle specific tasks with appropriate sub-actions. It is responsible for coordination of multiple activities.

The *decision making process* it may use, is not fixed by the Shadowboard architecture, for example it may employ a Naturalistic Decision Making process [16]. However, the default strategy, is to base Shadowboard decision making on a Generalised Constraint Solver [13], one that locates multiple constraint solvers down at each sub-agent class level, to best handle domain specific problems. Also located at each sub-agent class level, is an *archetype* for that class - an ideal version of a sub-agent, against which to measure and judge the chosen sub-agents actions and successes. These archetypes are represented by a set of beliefs, goals and intentions, of an idealised version of a sub-agent for that class, and act as the default set of beliefs

and goals. A given sub-agent attempts to uphold these defaults from its archetype, but drops back to its own implicit beliefs and goals, if it cannot satisfy the current task it is attempting to complete. If it still cannot complete its task satisfactorily, the constraint solver will take it off the job and find another sub-agent that may be successful. If the whole Shadowboard agent is part of a MAS, the Aware Ego Agent also makes decisions on whether to use internal sub-agents, or to *disown them* with regard a specific task, effectively contracting-out for new or improved capabilities.

5. Analysis: Shadowboard Agency, MAS and Autonomy.

Autonomy and Agency

When analysing the difference between an inter-agent (MAS) and the intra-agent architecture of Shadowboard, the first obvious divergence, is the treatment of autonomy and the notion of continual execution. In the Shadowboard architecture, the Aware Ego Agent alone has complete autonomy. It may grant autonomy to sub-agents, which may or may not run concurrently, but that granted autonomy is only with regard to sub-goals and delegated tasks, directly related to its recognised function and capabilities. Even so, the Aware Ego Agent can shut down any of the sub-agents at any time, and may either pass their tasks on to another sub-agent, or out to an external agent. Alternatively, it may let a sub-agent run concurrently with other sub-agents, feeding it updated information regarding goal revision and intentions.

Within a class of sub-agents, one agent does not supersede another. Each sub-agent has a different level and mix of capability and efficiency for a given type of task. One is not hierarchically superior to the other, unless the Aware Ego deems it to be so, while completing a particular job. Even though one sub-agent may have evolved from a second sub-agent, both currently in the same class, the earlier, less evolved agent may be the more appropriate to use in particular circumstances.

The clear distinction of a Shadowboard sub-agent from a fully operational agent in a more formal definition of agency [10], relaxes a number of characteristics that a sub-agent may have. While a sub-agent may be as complex as a fully-fledged agent (including its own subself-agents, in a recursive fashion), it may also be as simple as an *Active Object* [15], or an *Expert System* in the form of a Personal Assistant [12].

With the significant differentiation in Shadowboard between the *Aware Ego Agent* and the other *sub-agents*, we can draw upon different deviations of agency, for the different parts of the whole agent. BDI is applicable to the sub-agents each having a sub-set of the Beliefs, Goals and Intentions of the whole Shadowboard Agent; while Shohams [18] specific mentalistic notions of beliefs, capabilities, choices and commitments, in addition to BDI, are more suited to the Aware Ego Agent. Shoham alerts us to the divergence of meaning of agency from the original meaning of the word, of *'acting on behalf of someone else'*, to much more extensive and diverse definitions. The sub-agents of Shadowboard, adhere to that original meaning of the word *agent*.

Intra-agent communication

One of the advantages of modelling sub-agents within a whole agent superstructure such as Shadowboard, where it is feasible to do so over a MAS approach, is that sub-agent communication can be direct, implicit and efficient. Where a MAS will generally use an inter-agent communication language such as KQML, intra-agent communication can be via either: implicit complimentary pointers; or some form of a blackboard system. It is worth noting that some implementers of MAS systems have already used blackboard-based systems for communication between sub-agent like *internal agents*: the ARTIS system [3] uses a blackboard model for communication between their so-called *in-agents*; and Nikolaos [15] uses an *active message board* for what he calls *intra agent communication* in the April++/ALFA system.

Ontologies

In [26], Zini and Sterling point to an advantage of a MAS approach in that it enables the developer to decompose a complex task into easier-to-implement sub-tasks. Each agent within a MAS can be allocated *sub-tasks* according to its *role*. They differentiate task-based organisation of a MAS and knowledge-based organisation. They then make a case for explicit ontologies to handle the knowledge sharing aspects between various agents. That use of ontologies within a MAS applies equally well to the sub-agents of Shadowboard. With respect to use of ontologies, the difference between Shadowboard sub-agents and agents in a MAS, is one of granularity only - so Zini & Sterling use of ontologies are as valid to the sub-agents of Shadowboard as they are to a MAS.

Situated-awareness, self-awareness, social-awareness

A part of agency has included the notion that the agent has *situated-awareness* [10] – the agent is aware of its immediate surrounding environment. In the Shadowboard architecture the whole Agent has situated-awareness, while the sub-agents need have no idea about the surrounding environment, unless their role needs to know it.

Substantial amounts of research into multi-agent systems, particularly with respect to collaboration and cooperation - such as investigating shared goals and joint intentions – can be collectively viewed as making an agent *socially aware*. Castelfranchi [4] puts forward the idea that an agent socially committed to another, looses some autonomy. Cavedon and Sonenberg [5] investigate *roles* and *relationship between roles* in multi-agent systems. They see roles as an abstraction that aides the specification of agent behaviours. If you consider the closeness between a *role* and a subpersonality in psychological terms, there is some parallel between the Cavedon and Sonenberg agents and the sub-agents of Shadowboard, but there are also significant differences: theirs is a multi-agent system; their roles dictate obligations towards others; they rely on goal and intention substitution to get collaboration, whereas Shadowboard can simply swap or enliven sub-agents to get the result it needs; their roles have a supervisor/subordinate hierarchical relationship such that the superiors goals will always have a higher priority than the underlings, whereas there is no such

rigid structure amongst the sub-agents of Shadowboard – apart from the all powerful Aware Ego Agent with control over all sub-agents.

As opposed to the socially-aware nature of MAS architectures, the emphasis in Shadowboard, is very much about the *self-awareness* of an agent. Any social-awareness of a Shadowboard agent must come via an encompassing MAS system.

Mutual beliefs, stereotyping, prejudice and belief revision in a MAS

Disowning a sub-agent for the services of an external agent, involves commitment to an outer agent. In terms of such cooperation of multiple whole agents, where each is based on the Shadowboard model, mutual belief could be enacted by an agreement to strive for the same *archetype*, for a given capability envelope (class of sub-agent). Changing archetypes can also be used to effect *belief revision*.

In addition to locating a cooperating agent with mutual beliefs via shared archetypes, the concept of archetype may also helps us build internal models of other competitive or uncooperative agents. Building an internal model of the other, involves estimating where on the scale of levels of each *class of sub-agents*, does the others comparable sub-agent lie? How close to the archetype is it? This uses the known internal sub-agents, initially as a *stereotype*, to categorise the competitive agent - something that we humans often do, before getting to know a person more thoroughly. However, there are dangers associated with adopting stereotypes, with respect to belief revision. It is most well put by Allport in *The Nature of Prejudice* [1]:

> *To stereotype is to place a newly encountered entity into a preestablished category to save oneself the effort and time in getting to know this entity and in having to think about it. To stereotype is to shortcut thought, an economy measure we all take. However, not to allow facts to change the stereotypes we hold, is to be prejudiced.*

To avoid prejudice in modelling other agents, Shadowboard will need to dynamically adjust the categorisation of the competitor sub-agents, as experiences shows up inaccuracies of the current categorisations.

7. Conclusion

In the research and development of Personal Assistant Agents, semi-autonomous agency is generally assumed. In the MAS world, agents are generally considered autonomous, although less so under the influence of social obligations. The redefinition of an individual agent with a robust, analytical psychology based structure of sub-agents, brings those two domains closer - as it does, the concept of agency within BDI agents and that within Shohams Agent 0.

Watt's aim - to see agents and humans cooperate and otherwise socially interact, *more fully* - would certainly be realised if agents were each based on the Shadowboard architecture. Within Shadowboard, humans are being used as an archetype, something that Watts strove for.

It is worth noting, that while Minsky mainly dealt with very small sub-mind components, in a bottom-up manner when modelling the mind - when he did discuss some higher aspects of mind, he touched on several of the concepts of the psychology of subselves. When he discussed the Self, he used the term *'self-images'* in a similar way that subselves are used here, and he used the term *'self-ideals'* in a similar manner to the way *archetypes* are used here. He discussed an inner-self he termed *'The Conservative Self',* which closely fits the Protector/Controller concept of Voice Dialogue. However, he had a significant problem with a Central Self in a control mode, in the manner that the Aware Ego Agent is in control of a whole Shadowboard agent. He does talk of hidden selves, selves unknown, influencing ones course of action, much in the way that the disowned selves of Voice Dialogue do, in an individual not aware of their full compliment of subselves. Yet he was never able to identify those hidden selves in any systematic manner. Instead he talked of *'tricks'* to get things done - to effectively sidestep the negative effects of disowned selves. I would contend that his distrust of a *central control*, in the manner that the Aware Ego agent controls, stemmed entirely from the lack of an available technique for becoming aware of the hidden selves within the human psyche. Voice Dialogue is a technique now available to do just that.

Within psychology, Suppes et al describe two forms of *reduction* that employ *representation* in the modelling of complexities. Both are gainfully employed in Shadowboard: the use of sub-selves represent observable patterns drawn from analytical psychology; while the use of BDI (at the sub-agent level, including at the Aware Ego Agent level), is the sort of reduction that represents one as a set of theoretical concepts, with another - mapping *behaviour* into *belief, desires and intentions*.

Drawing extensively upon a refined model of mind as the Shadowboard architecture does, its holds a lot of promise in the construction of sophisticated whole-agents, both individual agents and MAS systems that encompass them.

References

1. Allport, G.W. *The Nature of Prejudice*. 25th Anniversary Edition. Adison-Wesley, 1979.
2. Assagioli, Roberto. *Psychosynthesis*. New York. Viking, 1965.
3. Botti, V, Carrascosa C., Julian V. and Soler J. Modelling Agents in Hard Real-Time Environments, pp 63-76 in *Multi-Agent Systems Engineering*, Garijo F. J. and Boman, M.(eds), 9th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW'99. Springer, LNAI 1647. 1999.
4. Castelfranchi, C. Commitments: from individual intentions to groups and organisations. In *Proceedings of International Conference on Multi-Agent Systems* ICMAS'95, 1995.
5. Cavedon, L. and Sonenberg, E.A. On Social Commitment, Roles and Preferred Goals. In Proceedings of the 1998 International Conference on Multi-Agent Systems ICMAS'98 Paris, Demazeau, Y., (ed), pp 80-87, July 1998.
6. Cohen, P.R. and Levesque, H.J. Intention is choice with commitment. Artificial Intelligence, 42(3), 1990.
7. Elliott, C. and Brzezinski, J. Autonomous Agents as Synthetic Characters, *AI Magazine*, American Association for Artificial Intelligence, pp13-30, Summer, 1998.
8. Goschnick, Steve and Sterling, Leon. Shadowboard as metaphor for an Agent Oriented Interface. Paper submitted to: *OZCHI 2000 Conference on Human-Computer Interaction*, Sydney, Australia, 2000.

9. Griffiths, Nathan. and Luck, Michael. *Cooperative Plan Selection Through Trust.* In Multi-Agent System Engineering, pp 162-74 in Multi-Agent Systems Engineering, Garijo F. J. and Boman, M.(eds), 9th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, MAAMAW'99. Springer, LNAI 1647. 1999.
10. Jennings, Nicholas R. and Wooldridge, Michael J. *Agent Technology: Foundations, Applications and Markets*, 1995.
11. Jung, Carl G. *Man and his Symbols*. Aldus Books, 1964.
12. Lashkari, Y., Metral, M., and Maes, P. Collaborative Interface Agents. In *Proceedings of the 12th National Conference of Artificial Intelligence*, 444-450, 1994.
13. Marriott, K. and Stuckey, P.J. *Programming with Constraints: an Introduction.* MIT Press. 1998.
14. Minsky, Marvin. *The Society of Mind.* Simon and Schuster Inc. 1986.
15. Nikolaos, Sharmeas. *Agents as Objects with Knowledge Base State*. Imperial College Press, 1999.
16. Norling, E., Sonenberg, E.A. and Ronnquist, R. *Enhancing Multi-Agent Based Simulation with Human Decision-Making Strategies.* Paper presented at the Fifth Australasian Cognitive Science Conference, Melbourne, Australia, Jan. 2000.
17. Rao, A.S., Georgeff, M.P. Modelling rational agents within a BDI-architecture. In, Allen, J., Fikes, E. and Sandewall (eds) *Proceedings of the Second International Conference on the Principles of Knowledge Representation and Reasoning*. Morgan Kaufmann Publishers, San Mateo, CA, 1991.
18. Shoham, Y. Agent-oriented programming. *Artificial Intelligence* 60:51-92, 1993.
19. Sliker, Gretchen. *Multiple Mind - Healing the Split in Psyche and World.* Shambhala Publications Inc. 1992.
20. Stone, Hal and Winkelman, Sidra. *Embracing Ourselves - the Voice Dialogue Manual.* New World Library. 1989.
21. St. Clair, M. *Object Relations and Self Psychology*. Brooks/Cole Publishing Co., 1996.
22. Suppes, P., Pavel, M., and Falmagne, J. -Cl. Presentations and Models in Psychology. In Porter, L.W. and Rosenzweig, M.R., editors, *Annual Review of Psychology*, 45:517-44, 1994.
23. Watt, S. Artificial Societies and Psychological Agents. In *British Telecom Journal, Special Issue on Intelligent Agents*, Autumn 1996.
24. Wooldridge, Michael. Agent-Based Software Engineering. In IEE Proceedings on Software Engineering, 144(1), pp 26-37, 1997.
25. Wooldridge, Michael and Jennings, Nicholas R. Intelligent Agents: Theory and Practice. *Knowledge Engineering Review*, 10(2):115-152, 1995.
26. Zini, Floriano, and Sterling, Leon. *Designing Ontologies for Agents*. Dept of Comp. Science, University of Melbourne, Technical Report 1999/15.