



## SCHOOL OF INFORMATION SCIENCES AND TECHNOLOGY THE PENNSYLVANIA STATE UNIVERSITY

# Steps Towards Including Behavior Moderators in Human Performance Models in Synthetic Environments

Frank E. Ritter and Marios N. Avraamides

Technical Report No. ACS 2000-1 19 May 2000

> Ritter@ist.psu.edu ritter.ist.psu.edu

Phone +1 (814) 865-4453 Fax +1 (814) 865-5604

School of IST, 504 Rider, 120 S. Burrowes St., State College, PA 16801

### Steps Towards Including Behavior Moderators in Human Performance Models in Synthetic Environments

Frank E. Ritter and Marios N. Avraamides ritter@ist.psu.edu marios@psu.edu School of Information Sciences and Technology The Pennsylvania State University University Park, PA 16801-3857

> ACS #2000-1 17 May 2000

#### Abstract

There are many factors that influence human behaviour that are not often included in models of human behaviour used in synthetic environments. These factors include external moderators such as heat, internal moderators such as various personality traits, and task-based moderators such as task difficulty and recent successes and failures. We provide an initial list of these moderators drawn from previous theories, particularly those that have been implemented, and from a review of experimental work. Implementing this list is considered with respect to several existing and new behaviour architectures. Including these moderators as part of an architecture such as Soar or ModSAF would be the most natural approach, but benefits could also be gained by implementing them in a separate system that moderates another architecture's output. Once implemented, the resulting model should be compared to relevant human behaviour to validate it.

#### Acknowledgements

Simon Russell and Bob Wray have provided useful comments. This work was supported by DERA (Portsmouth, UK) under contract CU003-0000001096. The conclusions reported here are solely the responsibility of the authors.

#### **Table of Contents**

1.	Overview of report	. 1
2.	The case for including behavioural moderators	. 1
	2.1 Extending and reusing effects within an architecture	. 1
	2.2 Filtering information with emotions	. 2
	2.3 Unified theory of personality and individual differences	. 3
3.	Review of existing models of moderators: Primarily emotions	.4
	3.1 PSI	. 5
	3.2 Architectural ideas behind the Sim_Agent Toolkit	. 5
	3.2.1 Cognition and Affect	. 5
	3.2.2 Sim_Agent and CogAff	.6
	3.2.3 Summary of the Sim_Agent Toolkit	.7
	3.3 Miscellaneous examples	.7
	3.4 Further use of the moderators	. 8
4.	Internal and External Moderators of Behaviour	. 8
	4.1 Types of moderators	.9
	4.2 External moderators	. 10
	4.3 Internal moderators	. 11
	4.3.1 Perception	. 12
	4.3.2 Decision process	. 13
	4.3.3 Multitasking variables	. 14
	4.3.4 Action	. 15
	4.3.5 Personality variables	. 15
	4.4 Task-based moderators	. 16
	4.5 Other moderators and interactions	. 17
5.	Possible implementation testbeds	. 19
	5.1 A human behaviour server	. 19

5.2 Modifications to ModSAF	. 20
5.3 Modifications to SAFs that ModSAF uses	. 21
6. Testing and analysis: An experimental programme	. 22
6.1 Experimental goals	. 23
6.2 Experiments to demonstrate and test changes	. 23
6.2.1 Measuring behaviour in a known environment	. 23
6.2.2 Showing that moderators lead to 'looking' like a human	. 24
6.2.3 Detailed comparisons of models and human behaviour	. 25
6.3 What testing will mean	. 25
7. Summary and future possibilities	. 26
8. References	. 27
Appendix 1. Example moderator interaction diagrams	. 31
Vibration	. 31
Temperature and exposure time	. 32
Visual fatigue	. 32
Level of training	. 33
Boredom	. 33
Cold and additional clothing	. 34
Humidity, heat, and fatigue	. 35
Grief	. 35
Anger and rage	. 36
Anxiety and noise	. 37
Intelligence, fatigue, and anxiety	. 38

#### 1. Overview of report

This report explores steps towards including behaviour moderators in human performance models in synthetic environments. The report will then note preferred options (best informed guesses) for implementing Extended Human Behavior Representation. The 'extended' is intended to cover any (and all) appropriate extensions that would make the behaviour more representative of the behaviour population modelled. It examines possible behaviour moderators, related models, and outlines a partial design for including the most important behaviour moderators into a behavioural (or cognitive) architecture, in particular, including such behaviours in a behavioural architecture that could be incorporated into or used in conjunction with ModSAF (Loral, 1995) and its derivatives (ONESAF, etc.). The idea is not necessarily to augment ModSAF, which we will also examine, but to design a behavioural architecture that could be incorporated into or used in conjunction with ModSAF and its derivatives. In the latter case, the architecture would be implemented in 'external' agents working with the SAFs (e.g. providing behaviour for entity placeholders in the SAF such as the UK command agents in UKSTOW97).

This report starts by reviewing how models have included behavioural modifiers. This has primarily been done with respect to models of emotions. After a review of models of emotions, factors that can be used to implement emotions are explored.

We have included an appendix noting in diagrammatic form the behavioural effects of eleven sets of moderators. They demonstrate how emotions and other moderators can modify behavioural models to make them more realistic. The emotional effects chosen for implementation include effects caused by the agent's external environment, the agent's internal state, and effects caused by recent mental and physical tasks that the agent has performed.

The results of these changes need to be tested and demonstrated by running the modified model and analysing and summarising the changes in behaviour with respect to military outcomes. We note here some suggestions for how to proceed in this area. The results can be used to suggest future areas and scope of work.

This report draws heavily on Pew and Mavor (1998) and particularly on Ritter Shadbolt, Elliman, Young, Gobet, and Baxter, (1999).

#### 2. The case for including behavioural moderators

There are a large number of reasons for including further moderators into behaviour models in synthetic environments. We note several here.

#### 2.1 Extending and reusing effects within an architecture

All of the arguments for creating a unified theory of cognition (Anderson, Matessa, & Lebiere, 1998; Newell, 1990) also apply to creating a unified theory of behaviour that includes the known behaviour moderators as well. The effects of behavioral moderators are by definition not task specific, so their implementation belongs in the architecture, not in the task knowledge.

Theories of emotions should be implemented within a uniform reusable architecture, and because they are related to cognition this will require a cognitive or behavioural architecture as well. Some models of emotions, as an example type of moderator, have been included in cognitive architectures (Bartl & Dörner, 1998; Belavkin, Ritter, & Elliman, 1999; R. Jones, 1998; Rosenbloom, 1998).

#### 2.2 Filtering information with emotions

The role of cognition is to process sensory information, assign meaning to it, and then decide upon a plan of actions in response. It is a real-time process in which new sensory information arrives continuously. The plan must therefore be dynamically reconfigurable and will often be abandoned in favour of a better plan mid-way through its execution. In a way similar to Rasmussen's (1998) stepladder framework of behaviour, Elliman writing in a joint authored work (Ritter et al., 1999) presented a speculative view of the role of emotions in cognition, which makes the following assumptions:

- (a) The amount of sensory data available at any moment is too large for attention to be given to a more than a small fraction of the data.
- (b) The conscious consideration of the results of perception is an expensive process in terms of the load on neural hardware and is also time consuming.
- (c) Most sensory processing is unconscious in its early stages in order that expensive conscious processes need consider only the *results* of perception. These results might include labelled objects with a position in space, for example, 'a tank moving its turret in that clump of trees'. Conscious processes might well add further detail such as the type of tank and the range of its gun.
- (d) Attentional mechanisms are needed to direct the limited high-level processing to the most *interesting* objects. These objects may be novel, brightly coloured, fast moving, or potentially threatening.
- (e) Planning is an especially heavy computational process for the human mind, and one that is difficult to carry out effectively under combat conditions. (Perhaps military doctrine is useful in that it distils the best generic practice and trains the soldier to behave in a way that might well have been a chosen and planned behaviour if the individual had the time and skill to formulate the action himself. The danger is that no doctrine can envisage all scenarios in advance, and on occasion the use of doctrine in a rigid manner may be harmful.)
- (f) From an evolutionary perspective this system of unconscious processing of sensory input, attentional mechanisms, and cognitive planning, together with speech-based communication, is a masterstroke of competence for survival. However, it has one crippling disadvantage: it is too slow to react to immediate and sudden attack.

Rapid reaction to possible threat without the time for much in the way of cognitive processing is clearly of huge value. In this framework emotion can be seen as kind of labelling process for sensory input. Fear in particular fits this pattern and is a label that causes selected sensory input to *scream for attention*. In order for this process to work rapidly it needs to be hard-wired in a way that higher-level cognitive processes are not. There is strong evidence that the amygdala is intimately involved in the perception of threat, and is then able to trigger the familiar sensation of fear (e.g. Whalen, 1999). If this organ of the brain is damaged, individuals may find everyday events terrifying whilst not perceiving any need for alarm in life threatening situations.

This rapid, emotive response to sensory data is inevitably relatively crude and prone to false alarms. Reactive behaviour is triggered that may be involuntary, for example, the startle reaction and physiological changes due to the release of noradrenalin. After the reaction response it takes time for cognitive processes to catch up and make a more informed assessment of the situation and actual threat. If this emotive, reactive stimulation is excited in a chronic manner then susceptible individuals may become less effective, with impaired ability to think and plan clearly. Any kind of anxiety is a form of stress. Because individuals have a finite capacity for absorbing it, excessive stress results in fatigue.

#### 2.3 Unified theory of personality and individual differences

Including behavioural moderators may be necessary for modelling non-doctrinal performance such as insubordination, fatigue, errors, and mistakes. Many authors have particularly noted the role of emotion in fast, reactive systems (Picard, 1997, provides a useful overview). Individual differences may be related to personality and problem solving. That is, the range of emotions may be best explained as an interaction that arises between task performance and situation assessment, the environment, and an agent's likes, desires, and personal cognitive style.

It would be useful to identify features that lead to modelling personality, problem-solving styles, and operator traits. This is a known weakness of current models of behaviour, where agents tend to act alike. Personality will be an important aspect of variation in behaviour between agents. While models that choose between strategies have been created, there are few models that exhibit personality by choosing between similar strategies (although see Nerb, Spada, & Ernst, 1997, for an example used to put subjects in a veridical but artificial social environment).

Including personality requires a task (and the model) to include multiple approaches and multiple successful styles. It is choices between these multiple strategies that can thus appear as a personality. If the task requires a specific, single strategy, or if the model only knows a single strategy, it is not possible to express personality through choosing a strategy. Psychology, or at least cognitive psychology, has typically not studied tasks that allow or particularly highlight multiple strategies. Looking for multiple strategies has also been difficult because it requires additional subjects and additional data analysis that in task analysis terms has not represented real differences. Differences in strategies, however, do appear to lead to variance in behaviour (Delaney, Reder, Staszewski, & Ritter, 1998; Siegler, 1987).

There appear to be at least the following ways to realise variance in behaviour that might appear like personality: learning, differences in knowledge, differences in utility theory and initial weighting, and differences in emotional effects. Such a model would fulfil a need for a source of regular, repeatable differences between agents in a situation.

All of the current cognitive architectures could support models of personality. In addition to differences in task knowledge, in Soar, personality can be expressed as differences in knowledge about strategy preferences either absolutely or based on a different sets of state and strategy features. Implementing these types of changes should be straightforward, as long as there are multiple strategies. ACT-R appears to learn better and faster which strategy to use compared with a simple Soar model, but requires additional state (Ritter & Wallach, 1998). Both models can modify their choice of strategies. The role of (multiple) strategies has been investigated within the EPAM architecture in several tasks, such as concept formation (Gobet, Richman, Staszewski, & Simon, 1997), and expert memory (Gobet & Simon, in press; Richman, Gobet, Staszewski, & Simon, 1996; Richman, Staszewski, & Simon, 1995).

These models could also be crossed with emotional and other non-cognitive effects to see how personality types respond differently in different circumstances (broadly defined). This could even be extended to look at how teams with different mixes of personalities work together under stress (Carley, 1996).

The amount of work to realise a model in this area will depend on the number of factors taken account of in the model. Providing a full model of personality and how it interacts with tasks and with other models is a fantasy at this point. However, a minimal piece of work would take an existing model and give it more of a personality by adding one or two of the moderators noted in Section 4. A more extensive project over a year or two would apply several of these techniques and confirm that it starts to better match human data.

# 3. Review of existing models of moderators: Primarily emotions

This section reviews how moderators in general have been incorporated in models, particularly cognitive models. This review is restricted, however, because most types of moderators have not been included in cognitive models with the exception of emotions. Before reviewing emotions, there are a few exceptions where the inclusion of behavioural moderators can be noted.

The ACT-R architecture has been used to model individual differences in capacity (Lovett, Reder, & Lebiere, 1999) and fatigue in general (Jongman, 1998). Miwa (1993) has explicitly modelled individual differences in knowledge. Otherwise, it is unusual for cognitive models to include such effects as fatigue, individual differences, or the effect of noise that are not effects that are seen or represented as emotional. In the area of modelling emotions, much more work has been done.

There are three related aspects of behaviour that are related to affective or emotional changes in behaviour that have been neglected in information processing models of human behaviour and in AI-based planners in synthetic forces (SF). The first aspect, extrinsic moderators, is how the physical environment influences problem solving and behaviour. We know that temperature, humidity, and noise, for example, all influence attention and problem solving (Boff & Lincoln, 1986). These aspects vary widely in the environment that real military forces inhabit. While change in the physical environment have sometimes been modelled in synthetic environments (SEs), their effects on agents in those environments have not been modelled.

The second aspect, intrinsic moderators, is how the individual problem internal state and resources moderate and support problem solving. These aspects include working memory and its differences, processing speed and its differences, and the mood and style of the problem solver.

The third aspect, task-based moderators, is how the problem solver's history influences their problem solving. For example, we know that previous successes and previous failures, both on a long-term and short-term basis influence behaviour in a variety of ways (e.g. Lindsley, Brass, & Thomas, 1995). These, too, have rarely been included in agent models in SEs.

We review existing models and existing known changes to behaviour caused by emotions. This review can then be used to suggest several emotions or emotional effects that would be the most useful examples of how these types of factors effect behaviour in a SE, and that give rise to significant individual differences in behaviour.

This review will result in a list of important effects along with the types of behaviour they influence and estimates of the difficulty of implementing and measuring these effects. This summary can be used in the next step of designing a set of emotional effects to include in the models of behaviours.

#### 3.1 PSI

PSI is a relatively new cognitive architecture designed to integrate cognitive processes, emotion, and motivation (Bartl, 2000; Bartl & Dörner, 1998; Döerner, 2000). The architecture includes six motives (needs for energy, water, pain avoidance, affiliation, certainty, and competence). Cognition is modulated by these motive/emotional states and their processes. In general, PSI organises its activities according to the Rasmussen's (1983) hierarchy: first, it tries highly automatic skills if possible, then it skips to *knowledge-based* behaviour, and as its *ultima ratio* approach it uses trial-and-error procedures. It is the only cognitive architecture that takes modelling emotion and motivation as one of its core tasks.

This architecture is currently incomplete. It raises interesting questions about how to judge a nascent architecture. PSI does not have a large enough user community and has not been developed long enough to have a body of regularities it has been compared with let alone adjusted to fit. How can it be compared with the larger architectures with existing tutorials, user manuals, libraries of models and example applications?

A model in the PSI architecture has been tested against a set of data taken from a dynamic control task (Detje, 2000). The model's number of control actions was within the range of human behaviour and its predictions of summary scores were outside the range of human behaviour (the model was less competent). This model needs to be improved before it matches human emotional data as well as cognitive models match non-emotional data. It is, however, one of the few models of emotion compared with data.

#### 3.2 Architectural ideas behind the Sim\_Agent Toolkit<sup>1</sup>

Sloman's ideas about cognitive architectures and the agent architecture toolkit (Sim\_Agent) were not reviewed in Pew and Mavor (1998). These architectural tools do, however, provide some useful and general lessons about architectural toolkits and about process models of emotions. Further information is available at the CogAff web site <www.cs.bham.ac.uk/~axs/cogaff.html> and from Ritter et al. (1999).

#### 3.2.1 Cognition and Affect

Sloman's toolkit emphasises reactive and multilevel components, perceptual input, and the interaction between reactive and deliberative layers and the role of an even higher level for meta-management. These mechanisms are used to support cognition with affect.

For instance, whereas many people have distinguished primary and secondary emotions (e.g. Damasio, 1994), Sloman and his colleagues have proposed a third type, tertiary emotions, also sometimes referred to as *perturbances* (Sloman, 1998a; Sloman & Logan, 1999). Primary

<sup>&</sup>lt;sup>1</sup> This section is taken and modified from Ritter et al. (1999) and was originally drafted by Sloman.

emotions rely only on the reactive levels in the architecture. Secondary emotions require deliberative mechanisms. Tertiary emotions are grounded in the activities of meta-management, including unsuccessful meta-management. There are other affective states concerned with global control, such as moods, which also have different relationships to the different layers of processing. Many specific states that are often discussed but very unclearly defined, such as arousal, can be given much clearer definitions within the framework of an architecture that supports them.

It looks as if various subsets of the capabilities described here arising out of the three layers and their interactions can be modelled in the architectures developed so far, for example, Soar, ACT-R/PM, the Moffatt and Frijda architecture, and the various logic-based architecture that dominate the ATAL series of workshops and books like Wooldridge and Rao (1999).

However, only small subsets of these capabilities can be modelled at present. That is fine as far as ongoing scientific research is concerned. It is not fine when such models are offered as solutions to hard practical problems requiring modelling of complete human beings rather than some restricted human cognitive capability. Any realistic model of human processing needs to be able to cope with contexts including (a) rich bombardment with multi-modal sensory and linguistic information, (b) in which complex goals and standards of evaluation are constantly interacting, (c) in which things often happen too fast for fully rational deliberation to be used, (d) in which not everything that occurs falls into a previously learnt category for which a standard appropriate response is already known, (e) where decisions have to be taken on the basis of incomplete or uncertain information, and (e) in which the activity of solving one problem or carrying out one intricate task can be subverted by the arrival of new factual information, new orders, or new goals generated internally as a side effect of other processes.

Where the individual is also driving a fast moving vehicle or in terrain that is under fire or possibly contains hidden snipers and other dangers (e.g. mines) then it is very likely that a huge amount of the processing going on will involve the older reactive mechanisms, including many concerned with bodily control and visual attention. It may be some time before we fully understand and can model the implications of total physical immersion in stressful situations, including the effects on deliberative and meta-management processes. (For example, fixing attention on a hard planning problem may be difficult if bombs are exploding all around you.)

#### 3.2.2 Sim\_Agent and CogAff

At present Sloman does not propose a specific overarching architecture as a rival to systems like Soar or ACT-R. He feels that not enough is yet known about how human minds work, and consequently any theory proposing *the* architecture is premature. Instead, he and his group have been exploring and continually refining a collection of ideas about possibly relevant architectures and mechanisms. Although the ideas have been steadily developing they do not believe that they are near the end of this process. So although one could use a label like *CogAff* to refer to the general sort of architecture they are currently talk about, it is not a label for a fixed design. Rather CogAff should be taken to refer to a high-level overview of a class of architectures in which many details still remain unclear.

The CogAff ideas are likely to change in dramatic ways as more is learned about how brains work, about ways in which they can go wrong (e.g. as a result of disease, ageing, brain damage, addictions, stress, abuse in childhood, etc.), and how they differ from one species to another, or one person to another, or even within one person over a lifetime. They also wanted a toolkit that

supports exploration of a number of interacting agents (and physical objects, etc.) where within each agent a variety of very different mechanisms might be running concurrently and asynchronously yet influencing one another, and where they could very easily change the architecture within an agent, change the degree and kind of interaction between components of an agent, could speed up or slow down the processing of one or more submechanisms relative to others (Sloman, 1998b).

Because it did not seem that other toolkits had the required flexibility, as they tended to be committed to a particular type of architecture, they built their own, which has been used for some time at the University of Birmingham and DERA Malvern. It is reported briefly in Sloman and Logan (1999) and in more detail in the online documentation at the Birmingham Poplog FTP site (<ftp.cs.bham.ac.uk/pub/dist/poplog/>). The code and documentation are freely available online. It runs in Pop-11 in the Poplog system (which is inherently a multi-AI language system, so that code in Prolog, Lisp, or ML can also be included in the same process). Poplog has become freely available (<ftp.cs.bham.ac.uk/pub/dist/poplog/new>).

The toolkit is still being enhanced. In the short term they expect to make it easier to explore architectures including meta-management. Later work will include better support for subsymbolic spreading activation mechanisms, and the development of more reusable libraries, preferably in a language-independent form.

#### 3.2.3 Summary of the Sim\_Agent Toolkit

The Sim\_Agent toolkit has commonalties with other approaches. The need for a library of components is acknowledged. They emphasise that reactive behaviours are necessary and desirable, and that the emotional aspects arise out of the reactive mechanisms. It provides a broad range of support for testing and creating architectures. The toolkit provides support for reflection as a type of meta-learning. Other architectures will need to support this as well, particularly where the world is too fast paced for learning to occur during the task (John, Vera, & Newell, 1994; Nielsen & Kirsner, 1994).

The features that the toolkit supports helps define a description of architectural types. The capabilities that can be provided, from perception through to action and from knowledge to emotions, provide a way of describing architectures.

The major drawback is that none of the models or libraries created in Sim\_Agent has been compared with human data directly. In defence of this, Sloman claims that, the more complex and realistic an architecture becomes, the less sense it makes to test it directly. Instead he claims that the architecture has to be tested by the depth and variety of the phenomena it can explain, like advanced theories in physics, which also cannot be tested directly.

#### 3.3 Miscellaneous examples

Several other models of emotions and architectures that use emotions have been created. Reviews of emotional models (Belyavin, Sheppard, & Russell, 2000; Hudlicka & Fellous, 1996; Pew & Mavor, 1998; Picard, 1997) typically present models and architectures that have not been compared and validated against human data, although there have been models created to match human data (e.g. Allander's cited in Pew & Mavor, and Gillis, 1998).

Attempts have been made to add several simple emotions to ACT-R (Belavkin et al., 1999) and validate the model by comparing the revised model with an existing model and comparable data

(Jones, Ritter, & Wood, 2000). These types of changes are being applied to an existing model, which matches adult behaviour well, to improve the match to children's more emotional behaviour (Belavkin et al., 1999). These emotional effects should improve the match to the children's performance by (a) slowing down performance in general, (b) slowing down initial performance as the child explores the puzzle driven by curiosity, and (c) abandoning the task if performance is not successful. This work should be extended and applied more widely. The other model of emotions that has been compared to data is an a unpublished PhD thesis by Araujo at Sussex (1994, cited in Picard, 1997).

Ortony, Clore, and Collins (1988) designed an often cited model that determines emotions based on features of a specific situation. The model does not attempt to simulate how emotions affect cognition. Instead, it focuses on how external stimuli (agents, events, and objects) cause emotions and affective reactions. The model of Ortony, Clore, and Collins has been implemented in a project that attempts to build autonomous agents in virtual reality microworlds ("the Oz Project" http://www.cs.cmu.edu/afs/cs/project/oz/web/oz.html).

The Affective Reasoner (Elliott & Siegle, 1993) is a set of Common Lisp AI programs that are embodied in multimedia computer agents and which can reason about emotion. These agents can have 24 different categories of emotion (i.e. fear, joy, hope etc). For each category of emotion there are 22 intensity variables (i.e. importance, suprisingness etc.) that can affect the intensity of the experienced emotion. In addition, agents have temperaments that allow them to express their emotions once these have arisen (i.e. turn red when they become angry).

There are also further models to be considered. They sometimes appear in the Computer Generated Forces conference and in the International Conference on Cognitive Modelling.

#### 3.4 Further use of the moderators

These models show that including moderators can lead to changes in behaviour including new types of behaviours. Where they have been tested they are starting to lead to a more accurate match to human behaviour. As a set, these models show that it is possible to include a wide range of behaviour moderators and still have successful problem solving.

Including the moderators of behaviour within an agent model allows other types of moderators and effects to be included and studied. These are not suggested for testing in this project, but illustrate the power of the representation available. The most important effect is perhaps the effect of being a novice. With the moderators currently included one can simulate novices, in that they appear to have a slower processing speed, slower movement speed, and make more errors. While this is a simplification, it may be a very useful one for trainers and simulations.

While these models include example moderators, they do not provide a rich enough set for our purposes. A review of the experimental literature will be necessary to create a wide enough set of moderators.

#### 4. Internal and External Moderators of Behaviour

In this section we expand on, and clarify, the relationship between cognition and behavioural moderators (extrinsic, intrinsic, and task based). We will argue that these effects have primarily been viewed as emotions or in emotional terms. While this is appropriate, there are other moderators of behaviour that are also important. These moderators have probably been viewed as

part of emotions because in nearly all cases they are related to and interact with emotions. To our knowledge these effects and interactions are not often implemented and tested, which adds to the difficulty of discussing and developing them.

The design will be done in two parts over this section and following sections. The first aspect will be provide a list of moderators or adjustable features to be included in the model. The external moderators will be available to models for modifying their information processing and behavioural architecture. The internal moderators will be available for modification at start-up to represent differences between problem solvers, and can be changed by external, task-based, and other internal moderators to represent how problem solving changes due to these effects. Task-based moderators will be recorded and fed back to the information processing and behavioural architecture in order to modify their performance accordingly. The second aspect is a review of various ways to implement the design as components in a synthetic environment.

These variables serve as an initial design to augment models of behaviour as a list of features to include as building blocks, that is, as an initial list of variables to include in such a model. The design will accommodate both a simple direct implementation (e.g. wearing gloves leads to less success), and an implementation with a cognitive architecture and mechanisms in mind (e.g. wearing gloves lead to slower reaction times, which leads to less success). We have chosen to include more variables than are necessary in order to provide a support for additional work and because while anxiety and fatigue are clearly important concepts, the hierarchy of importance after them is not completely clear. We have also only chosen variables where we can see could be operationalised. Some proposed moderators, even some here to a certain extant, while important are vague and difficult to implement. We have tried to avoid these.

How these variables are used to moderate behaviour is not explored here. This is a preliminary design of what factors to include. Their combination and the testing of these combinations is another project. Inclusion and extension of models to combine these moderators, such as the BCE model by Gillis (1998) that predicts the effect of sleep, fatigue, circadian rhythm, stress, and experience on decision making performance.

The variables are not designed to fit into equations such as 'success = motivation \* wisdom + strength', but instead to be moderate behaviour in a model that generates the actual behaviour. That is, they are designed to influence how a behaviour generating architecture such as Soar generates behaviour. Thus, in some cases, it will be more appropriate not to increment a variable called stress for its own sake, but to spend more time on a task called 'reflect'. Spending the agent's time on this task will operationalise stress in that less time is available for the main task and energy is expended on reflection.

In most cases, the data supporting the inclusion of a moderate is available from Boff and Lincoln (1986). Support for the remaining moderators either comes from references cited in Ritter et al (1999), or is generally accepted knowledge from psychology. Suggestions were also taken from other summaries (Banks & Stytz, 1998; Hudlicka & Billingsley, 1998).

#### 4.1 Types of moderators

Variables that moderate behaviour can be classified into three categories, external, internal, and task-based moderators. External moderators are effects that originate outside the problem solver, for example, the temperature of the environment. Internal moderators are those that originate inside the problem solver, for example, basic processing speed. Task-based moderators are those that arise from the problem solver performing their task, for example, recent success and failure at

the task. These definitions are related to but are different from those used by Pew and Mavor (1998, Ch. 9).

These moderators will be entwined and interact. They have been sorted out into tables, but the tables overlap. Where interactions are particularly strong the moderators and their effects are put where they appear to fit most naturally. Whilst temperature originates as an external moderator, processing speed will be affected by temperature if it leads to fatigue. The tables following note which moderators should be inputs to the behaviour architecture, which are used to modify the architecture, and which are outputs from the architecture.

While not emotions themselves, these resources and mechanisms can be used to implement emotions and the effects of external moderators. For example, anger might lead to a reduced fovea and less working memory, while sleep depravation might lead to a loss of memory and slower processing speed (please note, these examples are simplistic).

#### 4.2 External moderators

External moderators are by definition inputs from the environment that influence how the problem solver performs their task. Aspects related to the external world, such as temperature, will be taken from the existing synthetic environment where possible, and an external model of important variables to be modelled in an environmental simulator will be suggested.

Table 1 notes several measures to include from the external environment. These are fairly strong moderators, which themselves will be moderated by time. That is, the effect of temperature will vary with respect to how long the agent has been hot.

The external moderators in Table 1 should be passed to the behavioural model. These variables come from the external simulation and can modify how behaviour is generated. The behavioural model can then use these inputs to moderate behaviour.

Some of them can vary behaviour immediately, such as weather, and others the behavioural model will have to aggregate over time to compute how behaviour is modified. Temperature is an example. A high temperature does not provide a very drastic immediate effect, but over time it leads to fatigue.

#### 4.3 Internal moderators

Internal moderators need to be split into two types. First, they can provide a set of initial attributes that lead to individual differences or personality traits. In this regard, examples include differences in initial knowledge, processing speed, and settings of their internal decision process. As an agent starts up, it should be possible to provide them with a personality by reading in a list of values for these variables. The agent can then use these initial traits to moderate how these variables are influenced by and influence other variables (is he mad or is he always like this?). This set of variables is a fairly complete as a set of variables with which to experiment.

Second, the internal moderators also represent variables of what can be changed. This is a slightly radical concept. In Newell's (1990) arguments for cognitive architectures, architectures were noted as the aspects of cognition that were fixed. Allowing the variables associated with internal

**Table 1.** Aspects of the external environment that act as moderators of individual's behaviour.

	SOURCE OF	EFFECT(S)	EXAMPLE
	MODERATOR		
Temperature and Exposure time	Heat: external Exposure time: task	Impair mental tasks requiring memory and/or speeded decision making. Temperature alone impairs vigilance and tracking.	A team that is exposed to excessive heat for a long period of time will not respond as fast to sudden threat such as an ambush.
Humidity	External	Interacts with temperature to lead to fatigue.	High temperature and high levels of humidity will increase the fatigue produced by a strenuous physical task (i.e. a 20 mile walking through a jungle).
Additional clothing to protect from cold	External	The amount of additional insulation to hands decreases manual dexterity.	Soldiers wearing uniforms and gloves will be constrained in terms of producing fast movements (uniform) and manual tasks (gloves).
Vibration and noise	External	Impairs the perception of visual stimuli (through vibration of the eye or vibration of both the eye and the stimulus) and influences limb control movements.	Soldiers on board of a moving tank will perform worse in locating accurately enemy targets and throw accurate shots. The impairment will increase with an increase of the speed of the tank (a faster moving vehicle produces more severe vibration).
Time of day	External/Internal	Impairs and moderators cognition and quality of sleep	Attention and processing speed are moderated by the time of day. Quality of sleep is moderated by when it occurs.
Visibility	External	Moderates vision.	In low visability due to fog accuracy of recognition is impaired.
Percipitation	External	Moderates multiple factors.	In addition to visibility, humidity, temperature, this will moderate visibility and heat loss.
Fired upon	External	Can be used to derive threats to self and loss of comrades.	Being fired upon increased stress and related factors. Indirect fire would be a weaker version of this.

moderators to be varied by external moderators and task-based moderators runs counter to the letter of this approach. It does not, however, run counter to the spirit of this approach. In any architecture there are resources that get depleted and replenished, and different modes that an architecture can be set to. Allowing working memory to vary due to stress or fatigue is consistent with this view.

As we review internal moderators, keep in mind that they can be used two ways, defining an initial agent characteristic like a trait, character, or raw intelligence, and as a way the architecture is modified by the other moderators. These two systems of analysis have been used together (and confused together) in the past and keeping the distinction in mind will make the following discussion clearer.

A list of internal moderators of human behaviour and internal state variables can be proposed for use as implementation building blocks of emotional and other influences on behaviour. These should include, for example, fovea size (how much can the eye see), processing speed (how fast

can the agent think), working memory size (how much can the agent keep in mind), how memory is lost (from the goal stack or from long-term memory), and how strategies are chosen (e.g. best first, randomly, or based on previous performance). As an initial strategy, including more than necessary is probably a safer and more robust approach.

These internal moderators acting as changeable but sometimes stable aspects of behaviour must also represent will and tenacity. Without a representation of will and tenacity it will be impossible for people training against these models to break the model's will (which is often the current situation).

The factors are organised into perception, cognitive, action, personality, and constructed factors to assist in generating and understanding this list. How these factors are treated and combined by the behaviourally model is likely to vary, but no distinction is implied yet. Some of these measures can be directly operationalised in most architectures (e.g. working memory) and others are likely to be constructed or implicit measures (e.g. anxiety).

#### 4.3.1 Perception

Perceptual processes allow another area for individual differences and for short term effects. These areas are summarised in Table 2.

**Table 2.** Aspects of perception that can be modified to represent individual differences and that can be modified to implement the effects of other moderators such as emotions or fatigue.

	SOURCE OF	EFFECT(S)	EXAMPLE
	MODERATOR		
Attention to objects moving in the periphery	May be intrinsic, related to attention	Can notice objects faster.	Under stress this may be reduced to help focus on task at hand, but will also reduce ability to find items in the world.
Size of visual field	May be intrinsic, modifiable by task and fear	Can see more and/or be distracted more.	Same as above, but for stationary items.
Perceptual accuracy	May be intrinsic, related to multiple factors	More likely to find searched for items.	When accuracy changes, will spend more time on false alarms and on searching for targets.
Scan speed	May be intrinsic, related to multiple factors	Able to find an item more quickly.	A high scan speed will allow targets to be found more quickly. Moderated by knowledge and other factors.
Interpretation of ambiguous stimuli	May be intrinsic, related to multiple factors.	Ambiguous stimuli may be classified as hostile.	Multiple factors influence how ambiguous stimuli are classified. When anxious, ambiguous stimuli may be more likely classified as hostile.
Threats to self	External perception and knowledge	Will focus attention on the threat stimuli.	Knowledge about implication of sensors that suggest that an attack is coming will focus attention on those stimuli.

#### 4.3.2 Decision process

The exact decision process will vary based on the architecture used. There are several factors that are likely to exist in most architectures, and that one could argue should exist in every architecture. These factors are listed in Table 3.

The decision process for choosing internal operators and external behaviour appears to be moderated by multiple factors including emotions. A useful adjustment would be to provide multiple algorithms to make choices. Some emotional states will lead to choosing conservatively, others will choose hastily. The choice of multiple decision algorithms (e.g. risky, how noise is used, how affect about objects are incorporated) are influenced by and influence these factors. Alternatively, the model can choose based on affect towards the objects involved.

**Table 3.** Aspects of problem solving that can be modified to represent individual differences and that can be modified to implement the effects of other moderators such as emotions or fatigue.

	SOURCE OF	EFFECT(S)	EXAMPLE
	MODERATOR		
Working memory	Internal, may be influenced by expertise	Affects the ability to hold context	With more working memory more complex thoughts, processing, and multiple actions in parallel.
Knowledge	Internal, will be moderated with practice by strategy (knowledge) about how to learn	Affects capabilities generally, strategies, and task time.	It may make some tasks feasible due solely to speed.
Noise	May be intrinsic, moderated by mood and other factors	Noise in the decision process will mean that the best action will not always be performed.	With a high noise setting, perhaps as a result of poor performance, better strategies will not be selected, and less useful strategies will be tried.
Cognitive Processing speed	May be related to IQ, maybe result of practice.	Tasks are completed faster, more tasks may be completed.	Agents with higher processing speed can work inside an opponent's decision cycle, not only performing their tasks better but also anticipating others. See Laird and Duchi for a nice example.
Decision Process	Internal, choice may be influenced by task and other factors.	Multiple decision algorithms may be available. Some risky, some risk aversive, how emotions are incorporated.	When payoffs are large, a more adventuresome decision process might be preferred.
Level of training	Stored internally, but due to task performance	Perhaps operationalised as knowledge and strength of knowledge)	Additional training moderates many of these constructed variables, and will also influence behaviour through knowledge.

#### 4.3.3 Multitasking variables

Several moderators are related to working with multiple tasks. These areas are summarised in Table 4. These variables will tend to be affected rather than to affect others. They will be important when the task load is high and the ability to switch tasks is most important.

	SOURCE OF	EFFECT(S)	EXAMPLE
	MODERATOR		
Ease of engagement /disengage ment	Internal	Allows changing between tasks faster.	Those with high ease of engagement and disengagement can switch tasks easily. Those who cannot may become fixated on a single task to the detriment of other tasks.
Number of tasks allowed in parallel	Internal	Allows multiple tasks to pursed either in parallel or seemingly in parallel.	If the number of tasks is greater than what the system can support, tasks will be dropped.

**Table 4.** Moderators related to multitasking.

#### 4.3.4 Action

Action and other output processes allow another area for individual differences and for short term effects. These areas are summarised in Table 5.

	SOURCE OF MODERATOR	EFFECT(S)	EXAMPLE
Movement speed	May be intrinsic Moderated by fatigue, attention	Can move faster.	Stronger, faster soldiers will be able to march faster; faster hand movements will help those using instruments. This is one of the places fatigue will moderate performance.
Movement accuracy	May be intrinsic, modifiable by task, anxiety, and fear	Can move more accurately.	Attention and intrinsic settings will interact to moderate accuracy of movement.

#### Table 5. Moderators related to action.

#### 4.3.5 Personality variables

Personality variables are hard to operationalise directly, and the first implementations may be somewhat non-behaviour. They remain important, however, as a major source of individual differences. We note in Table 6 some general personality traits that may have some bearing on military simulations, as a subset taken primarily from Banks and Stytz (1998). This set is not

complete -- and other sets are likely to plausible as well -- but this set serve as a good initial place to explore how to implement personality factors and judge which are of value.

	SOURCE OF	EFFECT(S)	EXAMPLE
	MODERATOR		
Stability	Internal	Moderators other personality variables.	More stable individuals experience smaller changes in the other personality factors.
Humour	Internal	Accounts for emotional 'bounce-back'.	Individuals with more 'humour' are able to absorb and dissipate own and others losses and shocks.
Acquiescence	Internal	Willingness to follow orders.	This varies between individuals it also leads groups to perform differently.
Eagerness	Internal	Willingness to take up tasks.	Too eager may lead to problems in this domain, and be related to a type of false courage.

**Table 6.** Moderators related to personality.

#### 4.4 Task-based moderators

The final type of moderator is task-based moderators. These are sometimes grouped with the internal moderators, because they arise from the actions the problem solver performs, and sometimes are associated with the external moderators because they are fed back from the environment. They are related to both, and may arise out of multiple aspects, but appear to form a natural kind.

The task-based moderators to include are shown in Table 7. These variables arise out of the tasks that the agent performs, often based on knowledge to interpret stimuli. These should be recorded and passed to the agent so that it can moderate its behaviour accordingly. A confident agent may not be influenced by failure as much as a less confident agent.

Some external moderators that require reasoning and knowledge to judge are included here. Their precursors, such as perceptual input about attacking force size, are provided by the external simulation. However, the interpretation and influence are done by the behaviour model.

#### 4.5 Other moderators and interactions

There are numerous moderators of human behaviour that we have not included here. This would include such aspects as social behaviour, what you will do for your buddy, and what affect you have for objects and locations.

The exclusion of these moderators in this report and design is not necessarily because they are less important, but often simply because they are less well understood or the path to operationalising them is less clear. With time, they should be and will be included as well. For example, the ability to choose between reasoning methods or mental representations makes assumptions that most implementations will not be able to implement for quite a while.

We note a few interactions between moderators in Table 8.

	SOURCE OF	EFFECT(S)	EXAMPLE
	MODERATOR		
Task History	Task	Success and failure and their ratio will moderate the mood, motivation, and the decision process of the problem solver.	After failure, the problem solver will be more conservative and less likely to try risky actions.
Location of engagement	External, but related to task scheduling, measurement from internal	Affects motivation.	A soldier will be more passionate and more willing to risk his/her life when the task is to protect his/her own country than to defend others (i.e. fighting in Falkland Islands vs. fighting in Kosovo).
Boredom	Task	Affects detection efficiency.	Prolonged exposure to a task that contains no frequent changes (e.g. a radar during peace time) will increase reaction time in the occurrence of an event (e.g. hostile aircraft in the monitored airspace).
Visual Fatigue	Task	Produces anomalous patterns of eye movement, resulting in low velocity, long duration saccades.	A visually fatigued soldier will demonstrate impaired vigilance and ability to shift attention when a sudden threat (e.g. an ambush) occurs.
Level of training	Task and previous task	Practice at a task or related task decreases task performance time.	Practice with a task will allow it to be performed more quickly. This will make dual tasking more easy, and will also increase confidence.
Size of attacking force	External/ perception	Provides information to constructed moderators to generate fear or used as knowledge in planning.	Fear may be caused when attacked by an overwhelming force. Lack of perception may hide this.
Threats to self	External/ perception & knowledge	Can cause panic but requires knowledge to be caused.	Attacks directly against units, leaders, or individuals will lead to different responses than a general attack.
Task changes and importance	Task	Multiple tasks lead to multitasking and problems, a single task performed for a long time leads to problems with boredom.	Having more tasks to do than the agent can perform may lead to task shedding. Single long, low payoff tasks lead to boredom and problems with vigilance.
Type of task	Task	Need to implement types of fatigue.	Tasks that involved physical motion, working memory, visual scanning, can with time lead to fatigue in these systems.
Local casualties	Task/External & perception	Casualties to own unit and nearby units influence fear and anxiety (at least).	Local casualties can be seen, heard, and inferred. Their effect is likely to increase anxiety and fear, which would moderate other variables.

**Table 7.** Aspects of behaviour that act as inputs to implement changes.

	SOURCE OF	EFFECT(S)	EXAMPLE
	MODERATOR		
Physical and Mental fatigue	External/Task	Produces physical and mental fatigue	A sleep-deprived soldier will be less able to operate high-tech equipment
Fear/terror	Internal/Task	Affects decision making. Puts noise in the decision process and this may lead to taking wrong decisions.	A terrorised soldier may decide to surrender even when there's still good chance of winning.
Anger and Rage	Internal/Task	Affects decision making.	An enraged leader may attack earlier than it should. Overwhelmed by the desire to finish the job, he may engage in activity when the odds are against him rather than avoid engagement (i.e. when the enemy has more personnel and arms)
Fatigue	Task	Affects skilled performance	Larger deviations of instrument reading are tolerated before corrective action is taken. Lapses in attention are more frequent. Also, fatigue individuals are more likely to choose low effort/low probability of success strategies over high effort/high probability options. Tracking of targets and driving is also affected by fatigue.
Grief	History	Affects working memory and speed of processing, leads to fewer errors, produces changes in strategies.	A well-trained soldier will react faster in the presence of sudden danger and more likely to make the right decision since s/he is more knowledgeable. S/he may adopt decisions that were successful in similar situations in the past rather than trying to come up with solutions at that moment.
Grief	Internal/Task	Distracts concentration on the goal. May produce terror and fear and affect the decision-making process.	Soldiers experiencing casualties in their team will find it hard to concentrate on their task. Grief may lead to terror and fear and push the soldiers to give up fighting.
Confidence	History, but large individual differences	Influences motivation, respect and obedience, may lead to overestimation or underestimation of the situation.	A leader with many and recent successes will be respected and obeyed more by his soldiers, and will be more able to boost the morale of his team. Obedience to a successful leader will prevent errors resulting from individual actions. However, leaders and soldiers with high successes/failures ratios will be more likely to underestimate a situation and this can lead to a failure on the given task.

**Table 8.** A small sample of how moderators can interact.

	SOURCE OF	EFFECT(S)	EXAMPLE
	MODERATOR		
Anxiety and Noise	Noise: External (environment) Trait Anxiety: Internal State Anxiety: from Task	Increase selectivity of attention especially in dual tasks. Anxiety produces no improvement in main-task performance while noise does. Both noise and anxiety reduce secondary-task performance. Anxiety may also lead to overestimating potential danger.	A soldier has two tasks; to track the enemy position and monitor the layout of his teammates. A highly anxious leader (anxiety defined as either a trait or state anxiety) will perform better in the main task (i.e. monitor enemy) but worse on monitoring his fellow soldiers. In the presence of noise in the environment his performance on both the primary and the secondary task will decrease. An anxious leader may also overestimate potential danger and commit to a task an unnecessary number of troops and resources
Stress	Internal	Influence the ability to replenish other strengths.	Stress appears to be a long term effect of anxiety. A unit that has had a long period of activity will not recover from a shock as well as a unit that is unstressed.
Support	Internal/group	Group members may provide social support to each other.	This concept may be a general concept, or it might be operationalised through actions that make the team perform better and thus lightening the workload of each member.
Morale	Internal	Moderates confidence and ability to work.	A group with high moral will recover faster from set backs and perform tasks faster. Low moral may cause groups to give up.

**Table 8.** A small sample of how moderators can interact (continued).

#### 5. Possible implementation testbeds

We examine here the preferred options (best informed guesses) for implementing an extended human behaviour representation. The 'extended' is intended to cover any (and all) appropriate extensions that would make the behaviour more representative of the behaviour population modelled as referenced to the proceeding sections.

There are several options which are taken up in turn. The reader should note that there may be other concerns and benefits to these approaches. Our experience with these software systems limits us.

Implementing one or several of these moderators in any of these approaches would improve the existing models and provide value and benefits. For different sets of resources and timetables the preferred choice may vary, but all should be considered feasible.

#### 5.1 A human behaviour server

A human behaviour server, as shown in Figure 1, could be created that sits outside ModSAF and modifies the behaviour ModSAF generates on the fly. This design could have access to previous behaviour, the simulated environment, and perhaps some aspect of the internal state of the ModSAF simulation. It could modify the overt behaviour, for example, by dropping actions or modifying the perceptual input, such as by removing perceptual input.

The advantages of this approach is that it would not require modifying ModSAF itself, and could definitely be an easy way to show that modifications to behaviour are possible and would improved the simulation.

There are serious problems, however, with this approach. While it works, and in many cases will be quite useful, it fails to be as general as necessary and to fully support complex behaviour. It may be difficult to properly simulate what and how behaviour needs to be changed without including in the behaviour server everything that is already in ModSAF that generates behaviour. A more fine grained level more accurately matches how the behaviour is generated, allowing for higher level effects where heat and humidity and general fatigue could interact. For example, computing and implementing the implications of reduced working memory on the behaviours generated by ModSAF will pretty much require duplicating ModSAF. A reasonable implementation appears to be limited to a narrow range of effects, simple delays, dropped actions, and so on. There are rules that could be provided at the domain task level, for example, 'in a high temperature environment slow down each task proportional to an included table'.

If the domain tasks can only be done (or dropped) as in this approach, it will be difficult to incorporate changes most directly observed and described at the human information processing level such as sensory/perception; cognition (situation awareness, planning, decision making, multitasking, learning etc.); working memory; and motor behaviour levels. If an EHBR architecture can be made to work at a fine grained level, then domain task behaviour could be emergent, interuptable, and incorporate more of the moderators listed here.



Figure 1. HIP based computer generated force. Figure provided by Sheppard.

### 5.2 Modifications to ModSAF

Another alternative is to incorporate the moderators into ModSAF or related software, as shown in Figure 2. In this case, the external moderators are passed into ModSAF to be combined with internal moderators and the task-based measures.

Internal moderators would be loaded as traits into the same building blocks, and be modified as the effects of other moderators. Task-based moderators would be derived from the stimuli coming in, the responses going out, and internal performance metrics.

This approach is more costly in that it requires integrating the moderators with ModSAF, a not insubstantial system, and will require more extensive testing to find, control, and set the moderating variables.

There are numerous advantages, however. The implementation will be cleaner in the end because the knowledge and effects that interact closely will be closer to each other. The effects of multiple moderators can be combined at the point of behaviour generation. The effect of moderators will be more direct as well, moderating behaviour or delaying action as part of the process of behaving rather than as an appendum or afterthought.



Figure 2. State-based task-frame computer generated force. Figure provided by Sheppard.

#### 5.3 Modifications to SAFs that ModSAF uses

Another approach is to include the moderators in an architecture that ModSAF uses. Figure 3 shows a rough approximation where many of the moderators would be placed. The external

moderators would be available as inputs and direct modifications to sensing, cognition, working memory, long-term memory, and motor behaviour as indicated by the sweeping arrow, similar to implementing this in ModSAF.

Such an architecture should support multi-tasking. Many of the effects of the moderators modify how multiple tasks are approached. Without multiple tasks, the role of attention and task switching, for example, are not exercised.

These effects could be implemented in multiple agent architectures, e.g. Jack (Busetta, Rönnquist, Hodgson, & Lucas, 1999), Soar (Laird & Rosenbloom, 1995), ACT-R (Anderson & Lebiere, 1998), and Cognet (Zachary, Ryder, Ross, & Weiland, 1992).

One disadvantage of this approach is that another architecture must be created or learned in addition to ModSAF, and this architecture must be interfaced with ModSAF.

There are several advantages though. This approach has all the advantages of incorporating the changes into ModSAF with one additional advantage. ModSAF does not have a very complete model of several of these functional blocks. Choosing an architecture that provides more complete initial models of these function blocks will greatly assist in creating these moderators and their effects.



Figure 3. Stage model of human information processing. Figure provided by Sheppard.

#### 6. Testing and analysis: An experimental programme

Many of the theories of behaviour moderators added to agent models, particularly models that include emotions, have not been compared with detailed human performance data. Partly this may be because there is not always a lot of data available on how behaviour changes with these moderators (although, see Boff and Lincoln, 1986, for some general regularities at least). These are no doubt difficult factors to manipulate safely and reliably. But the models must not just be based on intuitions, they should be compared with the behaviour they are designed to simulate as a way of validating them (to know that they are worth taking seriously) and as a way for knowing where to improve them (Grant, 1962). The aspects of testing the model and analysing the results are tightly coupled, so they are covered here together.

After first noting the importance of setting experimental goals, noted here are several ways to test, validate, and improve models with experimental data. This area of experimentation with computer models has a small but useful literature. Interested readers and those running these experiments are referred to Cohen (1995) and Kibler and Langley (1988).

#### 6.1 Experimental goals

The goals of the experiment should be clear before the experiment is performed and analysed, even if the goals are simply to explore how to run such experiments. These goals will suggest and sometimes dictate what conditions to run (e.g. the model with and without a bit of knowledge), and which measure to take (e.g. time to perform a task, percent success). The goals can also include describing how the behaviour has changed in qualitative terms (e.g. which strategies used), and that new measures of behaviour are reliable and useful.

#### 6.2 Experiments to demonstrate and test changes

These models will have a large number of interacting and stochastic elements, and their behaviour will depend on and determine the environment. As such, multiple runs of the model will be necessary for understanding and summarising the model's behaviour.

The model could attempt to compute expected values of behaviour. This is useful when the environment is simple and non-interactive. If the model does provide just expected values, then history and sequences and how these interact will be difficult to compute and summarise because the expected value of a sequence is poorly defined.

The question then, of how many times to run the model, must be answered. Currently, there are several answers used in the field. Some researchers run as many runs as there are data points to compare. The use of paired runs arose out of maximising the output when running subjects in two conditions and how to allocate subjects for the maximum statistical power. While this makes sense when running subjects, it is wrong to treat a theory this way. The model is not another subject, it is a theory. As a theory it should make complete predictions, even if they are about means and distributions. This suggests numerous runs as are necessary, ideally, enough runs to make the predictions clear. This will become more important when extreme values are desired from the analysis, for example, the one in a hundred situation.

#### 6.2.1 Measuring behaviour in a known environment

Researchers studying cooperation in zero-sum games (Nerb et al., 1997) have developed a useful methodology to test models and people that interact in multi-agent systems. They place the subject in a situation with two models. That is, the subject interacts with agents with known personality types. The model can then later also be placed in the same situation for comparison.

This same approach can be used to test behavioural models in synthetic environments. In this case, a single subject or subjects are put into a known simulation. Subjects here would interact with a population of known agents and the resulting data could be compared to the revised model's interactions in the same environment. The use of simulators may provide a very clean way to obtain further data, with some validity.

The weaknesses to this approach are that the situations are not likely to be as representative of the real situation as one would like. The stress and reality of the simulated environment may be far less than in real life.

# 6.2.2 Showing that moderators lead to 'looking' like a human

Experiments like those used to test the effect of a moderator or to show that it makes a difference can be based on a modified experimental psychology design. There will be several types of subjects (i.e. models) and at least one type of task (i.e. a scenario). The design would test the model with and without its modifications. For example, three conditions might be used: (a) model with emotion X defending against a normal attacking model, (b) model with emotion Y defending against a normal attacking model.

Multiple measures of the defender and attackers could be gathered from the simulation. These data could be summarised and compared to see how and what type of changes in behaviour occurred. When this has been done in the past (e.g. in IPME of high/low alertness and gloves/no gloves), the model that has been modified is exhibited different behaviour (Belyavin et al., 2000).

These data can be tested quickly using a simplified version of the Turing test. In the most simple form of this test, the analyst looks at the behaviour to see if it looks human.

Another possible test of the model is to show a video tape of its before and after performance either to active officers or officers in training. This test, a type of Turing test, provides a holistic evaluation of performance. This is judged by the field to be a weak test. Many models can pass it if insufficient exposure is given, a small enough range of behaviour is sampled, or some combination. It is useful, however, for when the model fails, the reasons for its failure will help point out directly where to improve the model. Human observers will catch sequences and timing, difficult measures to quantify, but which can be used to validate objective measures and suggest new objective measures

A more advanced methodology is to have several levels and types of modifications. These different modifications are presented to subjects who are asked to rate how human the modified models are. In addition to seeing if the modifications lead to human-like behaviour (as judged by humans), the measurements also suggest the relatively impact and quality of the modifications. Laird and Duchi (2000) present a nice case study of this applied to making a model of playing

Quake more realistic by modifying its reaction times and strategies. Belyavin, Sheppard, and Russell (2000) also provide a nice example of this for the factors of alertness and operators with and without gloves.

#### 6.2.3 Detailed comparisons of models and human behaviour

The gold standard of testing models of behaviour has to be comparing the behaviour of the models with the behaviour of subjects in the same conditions (Ritter & Larkin, 1994). This data can be compared on multiple levels, that of aggregate or average behaviour, of strategy use, and of sequences of behaviour.

Consider as an example adding emotions to the Tower of Nottingham model (Belavkin et al., 1999). This model solves a simple blocks puzzle. It has had its behaviour compared with adults and then modified to match children's behaviour more closely (Jones et al., 2000).

A simple model of emotions/motivations was added to it. The model modified its problem solving behaviour based on recent successes and failures. After several failures its strategy choice would widen, and after many failures it would actually give up. With success, it would stick at the task through a longer string of failures.

This model, while quite plausible and matched new qualitative effects, was one of many possible. In order to validate it, the model's behaviour was directly compared with the children's behaviour to see if it actually did improve the fit. While the fit was not substantially better, the comparison suggested where to improve the model as this work continues.

The comparable situation with behaviour models in synthetic environments is to gather direct human data about moderators. Better instrumentation of some primary features of physiology as it relates to behaviour (e.g. heart rate, blood pressure) is providing new insights (Picard, 1997) and will also be useful for testing more veridical models.

We will have to choose dependent measures to examine. These are likely to include time to perform the task, percent successful completion, and type of strategy used. If these are not available from log files from synthetic environments like ModSAF, they should be added. Further detailed and tactical measures can be selected with the help of subject matter experts.

#### 6.3 What testing will mean

Testing models will be useful for at least three things. First, they will help users believe and help validate the models. Currently, moderators are included in models, but it is not clear that the changes lead to more accurate behaviour, they may just lead to different behaviour.

Second, they will tell us where to improve the models. The test that shows that the modified model's behaviour is indistinguishable from a human's is useful for selling or validating the model. The tests that show the model's behaviour is different will also show where to improve the model. This is more valuable in some sense, for it leads to progress.

Finally, this will lead to a better science of human behaviour. For example, some argue that emotions are necessary for problem solving. The data and the model may help answer whether this is true. Examples of brain damaged patients are put forward (e.g. Damasio's (1994) Elliot) who have damaged problem solving and damaged emotions. It is not clear that emotions per se are required, or if multiple aspects of behaviour were impaired as well as emotions by the patient's trauma. Clearly, AI models of scheduling do not have the same troubles scheduling an appointment despite their lack of emotion.

#### 7. Summary and future possibilities

Including behavioural moderators in human performance models offers two significant advantages. This work will lead to better models of behaviour in a scientific sense, contributing to the science and understanding of behaviour and what influences it. This work will also lead to more believable and useful agents to train with, to train against, and to use within synthetic environments. Improving models in this way represents a move from multiple simple opponents to a smaller number of more intelligent opponents. As such, the complexity increases.

This design remains a preliminary design. The necessary basic results in experimental, cognitive, individual differences, and affective psychology are available to help create a more complete design.

There are more internal variables in the architecture that can be modified by emotions. These types include knowledge structures, learning and modelling other's emotions, and further and more subtle changes to perception, processing, and output. This report has offered an initial list and focused on those that can be easily operationalised. There are also more task variables to consider including, but this is more subtle. The impact of these emotions will be moderated by the decreasing opportunities or situations where the emotions will obtain. That is, the most productive time and the biggest payoff is the first several emotions because these will represent the largest effects on human behaviour.

It will be these internal moderators acting as changeable but sometimes stable aspects of behaviour that must represent will and tenacity. The point of many current tactics is to break the opponent's will. Without a representation of will and tenacity it will be impossible for people training against these models to break their will (which is often the current situation).

The question of validating behavioural moderators remains an important one. Further work will be needed to understand how much detail is needed for any application, the different needs between training and analysis, and when and how to effectively develop good value models. And, quite importantly, the output of these models needs to be validated by comparing it with human behaviour.

This work matches the call for unified theories of cognition. Newell's (1990) arguments for UTCs, concerns, and potential payoffs and solutions apply to this work as well. Keeping in touch with the cognitive modelling community would be very helpful for both groups.

#### 8. References

- Anderson, J. R., & Lebiere, C. (1998). The atomic components of thought. Mahwah, NJ: Lawrence Erlbaum.
- Anderson, J. R., Matessa, M., & Lebiere, C. (1998). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, *12*, 439-462.
- Banks, S. B., & Stytz, M. R. (1998). An approach to enhance human behavior modeling for computer-generated actors.
- Bartl, C. (2000). Metacognition in complex problem solving: How to implement the effects of metacognition on goal selection in the architecture of PSI. In N. Taatgen & J. Aasman (Eds.), *Proceedings of the Third International Conference on Cognitive Modelling*. 33-41. KS Veenendaal: Universal Press.
- Bartl, C., & Dörner, D. (1998). PSI: A theory of the integration of cognition, emotion and motivation. In F. E. Ritter & R. M. Young (Eds.), *Proceedings of the 2nd European Conference on Cognitive Modelling*. 66-73. Thrumpton, Nottingham, UK: Nottingham University Press.
- Belavkin, R. V., Ritter, F. E., & Elliman, D. G. (1999). Towards including simple emotions in a cognitive architecture in order to fit children's behaviour better. In *Proceedings of the 1999 Conference of the Cognitive Science Society*. 784. Mahwah, NJ: Lawrence Erlbaum.
- Belyavin, A., Sheppard, C., & Russell, S. (2000). Final presentation on the Human Science/ Synthetic Environment Integration Demonstrator programme given to the TG5 customer on 14 March 2000.
- Boff, K. R., & Lincoln, J. E. (1986). *Engineering data compendium: Human perception and performance*. Wright-Patterson Air Force Base, OH: John Wiley & Sons.
- Busetta, P., Rönnquist, R., Hodgson, A., & Lucas, A. (1999). JACK intelligent agents -Components for intelligent agents in JAVA. *AgentLink News Letter*, 2(Jan.), <http://www.agent-software.com/white-paper.pdf>.
- Carley, K. M. (1996). A comparison of artificial and human organizations. *Journal of Economic Behavior & Organization, 31*, 175-191.
- Cohen, P. R. (1995). Empirical methods for artificial intelligence. Cambridge, MA: MIT Press.
- Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York, NY: Gosset/Putnam Press.
- Delaney, P. F., Reder, L. M., Staszewski, J. J., & Ritter, F. E. (1998). The strategy specific nature of improvement: The power law applies by strategy within task. *Psychological Science*, 9(1), 1-8.
- Detje, F. (2000). Comparison of the PSI-theory with human behaviour in a complex task. In N. Taatgen & J. Aasman (Eds.), *Proceedings of the Third International Conference on Cognitive Modelling*. 86-93. KS Veenendaal: Universal Press.

- Döerner, D. (2000). The simulation of extreem forms of behavior. In N. Taatgen & J. Aasman (Eds.), *Proceedings of the Third International Conference on Cognitive Modelling*. 94-99. KS Veenendaal: Universal Press.
- Elliott, C., & Siegle, G. (1993). Variables influencing the intensity of simulated affective states. In AAAI technical report for the Spring Symposium on Reasoning about Mental States: Formal Theories and Applications. 58-67. Menlo Park, CA: American Association for Artificial Intelligence.
- Gillis, P. D. (1998). Realism in computer generated forces command entity behaviors. In Proceedings of the Seventh Conference on Computer Generated Forces and Behavioral Representation. Orlando, FL, 12-14 May 1998.
- Gobet, F., Richman, H., Staszewski, J., & Simon, H. A. (1997). Goals, representations, and strategies in a concept attainment task: The EPAM model. *The Psychology of Learning and Motivation*, 37, 265-290.
- Gobet, F., & Simon, H. A. (in press). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science*.
- Grant, D. A. (1962). Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychological Review*, 69(1), 54-61.

Hudlicka, E., & Billingsley, J. (1998).

- Representing behavior moderators in military human performance models. In *Proceedings of* the Seventh Conference on Computer Generated Forces and Behavioral Representation.
- Hudlicka, E., & Fellous, J.-M. (1996). Review of computational models of emotion (Technical Report No. 9612). Psychometrix. Arlington, MA.
- John, B. E., Vera, A. H., & Newell, A. (1994). Towards real-time GOMS: A model of expert behavior in a highly interactive task. *Behavior and Information Technology*, *13*, 255-267.
- Jones, G., Ritter, F. E., & Wood, D. J. (2000). Using a cognitive architecture to examine what develops. *Psychological Science*, *11*(2), 1-8.
- Jones, R. (1998). Modeling pilot fatigue with a synthetic behavior model. In *Proceedings of the 7th Conference on Computer Generated Forces and Behavioural Representation*. 349-357.
- Jongman, G. M. G. (1998). How to fatigue ACT-R? In *Proceedings of the Second European Conference on Cognitive Modelling*. 52-57. Nottingham: Nottingham University Press.
- Kibler, D., & Langley, P. (1988). Machine learning as an experimental science. In *Proceedings* of the Third European Working Session on Learning. 81-92. Glasgow: Pittman. Reprinted in J. W. Shavlik & T. G. Dietterich (Eds.) (1990), *Readings in machine learning*. San Francisco, CA: Morgan Kaufmann.
- Laird, J., & Duchi, J. C. (2000). The humanity of the soar Quakebot. In *Proceedings of Soar Workshop 20.* 84-89. USC/ISI Soar Group.
- Laird, J. E., & Rosenbloom, P. S. (1995). Evaluation of the Soar cognitive architecture. In D. M. Steier & T. M. Mitchell (Eds.), *Mind matters*. 1-50. Hillsdale, NJ: LEA.

- Lindsley, D. H., Brass, D. J., & Thomas, J. B. (1995). Efficacy-performance spirals: A multilevel perspective. *Academy of Management Review*, 20(3), 645-678.
- Loral (1995). ModSAF software architecture design and overview document (SADOD) No. Report ADST/WDL/TR--95--W003339B). Prepared for U.S. Army Simulation, Training and Instrumention Command (STRICOM).
- Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified architecture. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*. 135-182. New York, NY: Cambridge University Press.
- Miwa, K., & Simon, H. A. (1993). Production system modeling to represent individual differences: Tradeoff between simplicity and accuracy in simulation of behavior. In A. Sloman, D. Hogg, G. Humphreys, A. Ramsay, & D. Partridge (Eds.), *Prospects for artificial intelligence: Proceedings of AISB'93*. 158-167. Amsterdam: ISO Press.
- Nerb, J., Spada, H., & Ernst, A. M. (1997). A cognitive model of agents in a commons dilemma. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society*. 560-565. Mahwah, NJ: Lawrence Erlbaum.
- Newell, A. (1990). Unified theories of cognition. Cambridge, MA: Harvard University Press.
- Nielsen, T. E., & Kirsner, K. (1994). A challenge for Soar: Modeling proactive expertise in a complex dynamic environment. In *Singapore International Conference on Intelligent Systems (SPICIS-94)*. B79-B84.
- Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge, UK: Cambridge University Press.
- Pew, R. W., & Mavor, A. S. (Eds.). (1998). Modeling human and organizational behavior: Application to military simulations. Washington, DC: National Academy Press. http://books.nap.edu/catalog/6173.html.
- Picard, R. W. (1997). Affective computing. Cambridge, MA: The MIT Press.
- Rasmussen, J. (1983). Skills, rules, knowledge: signals, signs and symbols and other distinctions in human performance models. *IEEE Transactions: Systems, Man & Cybernetics, SMC-13*, 257-267.
- Richman, H. B., Gobet, F., Staszewski, J. J., & Simon, H. A. (1996). Perceptual and memory processes in the acquisition of expert performance: The EPAM model. In K. A. Ericsson (Ed.), *The road to excellence*. Mahwah, NJ: Erlbaum.
- Richman, H. B., Staszewski, J. J., & Simon, H. A. (1995). Simulation of expert memory with EPAM IV. *Psychological Review*, *102*, 305-330.
- Ritter, F. E., & Larkin, J. H. (1994). Using process models to summarize sequences of human actions. *Human-Computer Interaction*, *9*(3), 345-383.
- Ritter, F. E., Shadbolt, N. R., Elliman, D., Young, R., Gobet, F., & Baxter, G. D. (1999).Techniques for modelling human performance in synthetic environments: A supplementary review (Tech. Report No. 62). ESRC CREDIT, Dept. of Psychology, U. of Nottingham.

- Ritter, F. E., & Wallach, D. P. (1998). Models of two-person games in ACT-R and Soar. In Proceedings of the Second European Conference on Cognitive Modelling. 202-203. Nottingham: Nottingham University Press.
- Rosenbloom, P. (1998). Emotion in Soar. In *Proceedings of Soar Workshop 18*. 26-28. Vienna, VA: Explore Reasoning Systems.
- Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology*, *115*, 250-264.
- Sloman, A. (1998a). Damasio, Descartes, alarms and meta-management. In *Proceedings of the International Conference on Systems, Man, and Cybernetics (SMC98).* 2652-2657. IEEE.
- Sloman, A. (1998b). What's an AI toolkit for? In B. Logan & J. Baxter (Eds.), Proceedings of the AAAI-98 Workshop on Software Tools for Developing Agents. 1-10. Menlo Park, CA: AAAI Press.
- Sloman, A., & Logan, B. (1999). Building cognitively rich agents using the Sim\_Agent toolkit. *Communications of the Association of Computing Machinery*, 42(3), 71-77.
- Whalen, P. J. (1999). Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. *Current Directions*, 7(6), 177-188.
- Wooldridge, M., & Rao, A. (Eds.). (1999). *Foundations of rational agency*. Dordrecht, The Netherlands: Kluwer Academic.
- Zachary, W., Ryder, J., Ross, L., & Weiland, M. Z. (1992). Intelligent computer-human interaction in real-time, multi-tasking process control and monitoring systems. In M. Helander & M. Nagamachi (Eds.), *Human Factors in Design for Manufacturability*. New York, NY: Taylor and Francis.

#### Appendix 1. Example moderator interaction diagrams

The figures in this appendix are fashioned after Figure 9.4 in Pew and Mavor. They illustrate how example behavioural moderators noted in section 4 interact with the cognitive and physical architecture to give rise to changes in behaviour. These examples are ordered roughly by complexity.

#### Vibration

In addition to the direct effects to cognitive architecture parameters, over time vibration will lead to additional effects of physical fatigue (when considered with respect to other factors including task, clothing, and temperature).



#### Temperature and exposure time

In addition to the direct effects to cognitive architecture parameters, over time heat will lead to additional effects of physical fatigue (when considered with respect to other factors including task, clothing, and temperature).



#### Visual fatigue

In addition to the direct effects of being a difficult task, a visually strenuous task will lead to fatigue effects.



### Level of training



#### Boredom

The task moderator here, a monotonous task, may actually be based on cognitive architecture parameters, such as the value of the task, and the behaviour itself, such as very little movement because the task stimuli do not require it.



### Cold and additional clothing

In addition to the direct effects on the physical architecture, over time the effect of the external temperature will lead to further changes in the physical and cognitive architectures.



#### Humidity, heat, and fatigue

Heat and humidity interact with multiple factors, including task performance, to generate fatigue.



#### Anger and rage

Anger and rage illustrate rather complex interactions that touch nearly all of the architecture. Even this representations is probably simplistic, but would be a useful addition to behavioural models.



#### Anxiety and noise

Anxiety, both as a trait and as a state, and noise, both continuous and sudden, will influence the cognitive architecture both directly and indirectly through fear and terror. If anxiety as an internal moderator is moderated itself by behaviour there is a possibility to set up self-reinforcing feedback loops.



#### Intelligence, fatigue, and anxiety

Intelligence and experience can moderate fatigue and anxiety under heavy task demands. Time of day would shift the ratio between fatigue (increased processing in the morning could decrease fatigue) and anxiety (increased alertness in the afternoon could decrease secondary task decrement).

