A Concern-Centric Society-Of-Mind Approach To Mind Design

Steve Allen[1]

Multi-Agent Systems Group,
German Research Centre for Artificial Intelligence,
Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
Email: Steve.Allen@dfki.de

Abstract

In this poster, we argue that mental concern-processing mechanisms are amenable to a society-of-mind approach to mind design. We illustrate our case with an information-level analysis of the emotion process, relating the different classes of emotional state to the different layers of our motivated agent framework. We describe how a society-of-mind design-based implementation strategy allows us to add depth to our agent architecture, and incrementally account for more and more of the phenomena of interest. Finally, we report on the results of recent research into the design of cognitively-inspired emotional agent architectures.

1 The Emotion Process

1.1 An Architectural Framework for Elucidating Emotions

Emotions form a powerful, but ill-defined class of motivational control states that have spawned a wealth of competing definitions and theories. We can take some tentative steps towards untangling this web of conflicting ideas by mapping the *Emotion Process* on to a generic architectural framework - see Figure 1.1.

By referring the different definitions and theories of emotion to the different layers of the architectural framework, we can identify three main classes of emotional state – *primary*, *secondary*, and *tertiary*. Those theorists who: (a) stress emotions based on the limbic system are primarily studying effects of the reactive layer; (b) stress emotions such as apprehension, disappointment and relief, related to phases in the execution of plans, are studying effects of the attentive/deliberative layer; and (c) stress emotions involving loss of control of thought processes are studying processes involving the self-reflective, or meta-management layer.

[1] In collaboration with the Cognition and Affect project at Birmingham University.

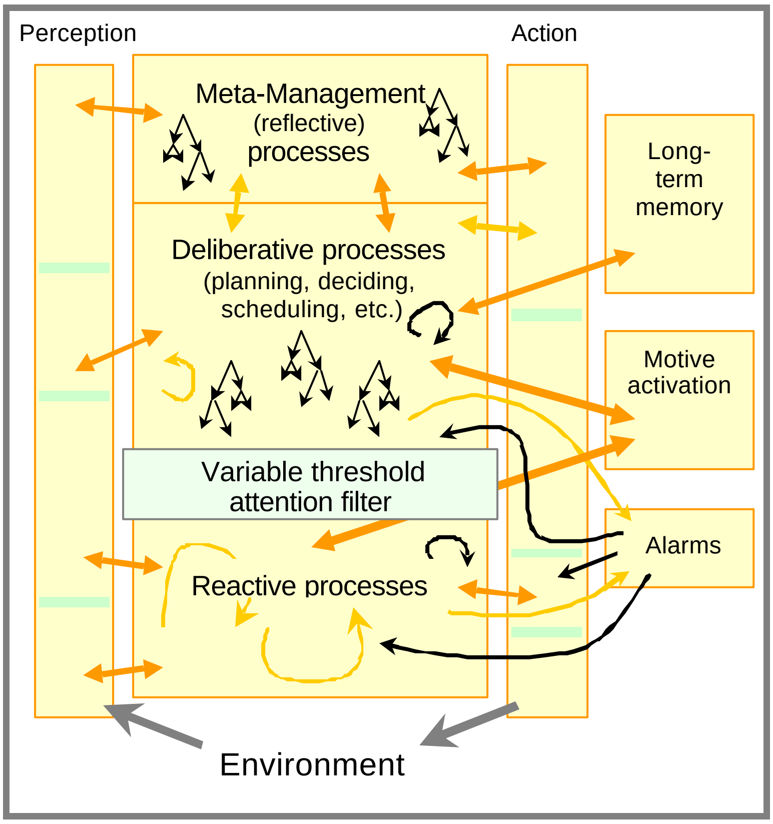


Figure 1.1 A Motivated Agent Framework [Sloman 99]

Pre-Attentive (Reactive) management processes use dedicated circuits to respond automatically to triggering conditions in the environment. There is no considered construction of new plans or explicit evaluation of alternative options. New behaviours and concepts may form through modification of association strengths or relative weights in automated processes such as reinforcement learning.

Attentive management processes use general purpose resources to focus and address the current primary concerns of the agent. As reusable mechanisms and space are dynamically allocated, many of the processes are inherently serial and resource limited. Access to concurrent long-term memory may also be inherently serial due to problems of cross-talk. *Deliberation* (a sub-class of attentive processing) is the process whereby possible world models are constructed and used for the evaluation of plans and goals before actions are selected. Deliberation requires working memory to facilitate the comparison of options, and long-term memory to store the individual steps used in the construction of the plan.

The Attention Filter is proposed as a mechanism to protect the resource limited attentive processes from excessive interruption by reactive motivators. The filter threshold is set by meta-management processes, and reflects the perceived importance/urgency/difficulty of the current attentive task.

Meta-management processes are responsible for monitoring and controlling motivator management mechanisms. It is likely that approaches that work well in an agent's early development may become less-than-optimum as the agent's environment (including its internal environment) change.

1.2 Different Stages of the Emotion Process

Emotions are neither discrete events, nor linear processes. Most of the time information flow is not only from the top down, but bi-directional – stimuli are often actively acquired and context evaluations made prior to information coding. Different stages of the emotion process can be skipped, and the process as a whole interrupted – leading to the myriad of variants of emotional phenomena. However, an emotion in its typical form embodies the process in its entirety (see Figure 1.2), subject to numerous regulation processes. The different stages of the emotion process are:

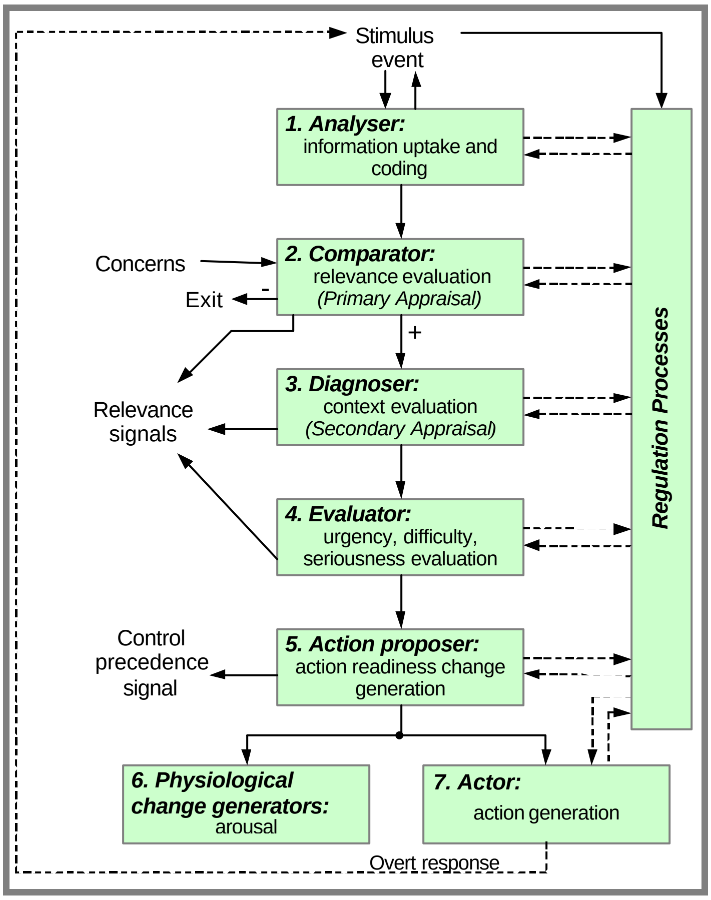


Figure 1.2 The Emotion Process [Frijda 86, pages 454-456]

1. Analyser. The analyser codes the event in terms of known event types and what they might imply with respect to cause or consequence.

2. Comparator: The stimulus event is appraised as to its relevance for one or more of the subject's concerns: relevance evaluation, or *primary appraisal.* Relevance evaluation results from comparing the event with the satisfaction conditions or sensitivities of the various concerns, for many, or all, concerns in parallel. Outputs of the Comparator are the four relevance signals: pleasure, pain, curiosity, and desire; or, by default, irrelevance, whereupon the process exits.

3. Diagnoser: The stimulus situation as a whole is appraised in terms of what the subject can or cannot do about it. Context evaluation or *secondary appraisal* diagnoses possibilities or impossibility for coping, and can be regarded as a series of diagnostic tests. Output is a patterned diagnosis or *situational meaning structure* – how the situation appears to the subject, i.e. the valence of the situation as a whole. The *situation meaning structure* comprises of three kinds of elements: (i) cognitions of what the situation does or offers the subject; (ii) cognitions of what the situation allows the subject to do; and (iii) evaluations as to whether the various outcomes are desirable or not. In addition, the Diagnoser, together with the Comparator, provides the Evaluator with information as to how difficult, urgent, or serious events really are.

4. Evaluator: Urgency, difficulty, and seriousness are computed, and combined in a signal of control precedence for dealing with the current event. They thus cause action interruption if need be, or else they cause distraction when previous action happens to continue.

5. Action Proposer: Action readiness change is generated and presses for, or occupies, control precedence. Action readiness change consists of a plan for action – action tendency (tendencies to execute expressive behaviour, which are present prior to, and independent of actual execution) – and/or for mode of activation. In the case where action programs are fixed and rigid, the concept of action readiness change loses much of its meaning – it only exists to the extent that inhibition can be used to block actions. However where action programs are flexible and alternative courses of action are possible, intentions and goals become independent of the particular actions and the term action readiness becomes more relevant

6. Physiological Change Generator: Physiological change is effected, in accordance with the action readiness mode generated by the action proposer.

7. Actor: Action – overt or cognitive – is selected, as determined by the action readiness mode and by other aspects of the situation.

1.3 Three Classes of Emotional State - Primary, Secondary, and Tertiary

***Primary* emotional states**: such as being startled, terrified, or sexually stimulated, are typically triggered by patterns in the early sensory input (sensory thalamus) and detected by a dedicated global alarm system (centred on the limbic system). These emotional states are sometimes called primes or primary emotions [Buck 85; Damasio 94; Picard 97].

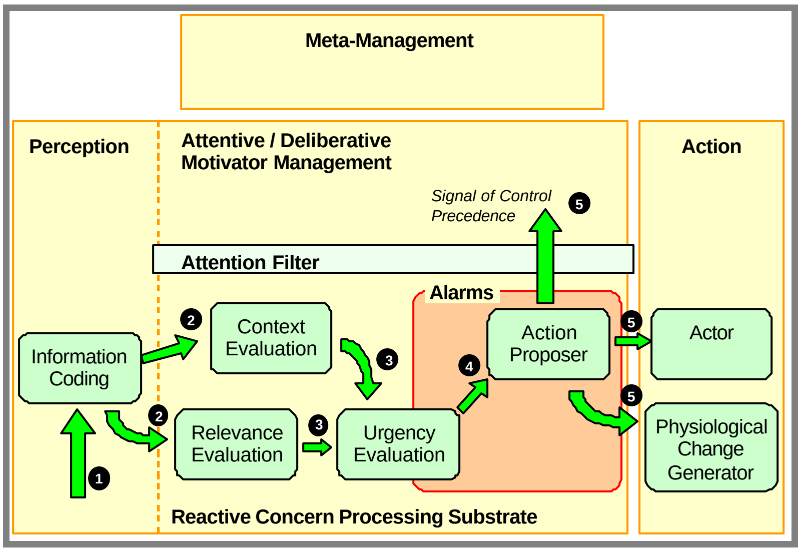


Figure 1.3 Information Flow leading to a Primary Emotion

Figure 1.3 shows a simplified graphical representation of the information flow that leads to a primary emotion state within our three-layered model. In Figure 1.1 we kept the global alarm system as a separate entity, here we attempt to place it within the confines of the three pillars (*perception*, *cognition*, and *action*). Although we would like to depict a clean boundary between perception and cognition (i.e. both deliberation and the reactive *concern-processing* substrate), the border between the two is in reality very fuzzy (this is not all that surprising when you acknowledge that both perception and cognition simply refer to labels that help us carve up the functionality of the brain). It can be argued that relevance and context evaluation belong to both cognition and perception – as physically the sensory thalamus is also responsible for the early sensory processing of such information in our brain (see *'Amygdala Pathways and Fear Conditioning'* in Figure 2.2).

The information flow that leads to a primary emotion can be summarised as: (**1**) external percepts are detected and encoded into known event types; (**2**) in parallel, the event is evaluated relative to the agent's concerns and the context of the current situation (i.e. the agent's coping strategies) – relevance and context evaluation must rely of simple heuristics such as speed, intonation, size, habituation or familiarity; (**3**) the urgency of the event is evaluated – as a simple function of the current level of arousal, context and relevance evaluations; (**4**) action readiness change is generated and presses for control precedence; and finally (**5**) attentive cognition is interrupted as the motivator gains control precedence, an involuntary action *might be* performed, and some form of physiological change *might be* instigated according to the action readiness mode generated by the action proposer.

***Secondary* emotional states**: such as being anxious, apprehensive, or relieved, depend on the existence of a deliberative layer in which plans (for future states) can be created and executed with relevant risks noticed, progress assessed, and success detected. An alarm system capable of detecting features in theses cognitively generated patterns is still able to produce global reactions to significant events in the thought process that impinge on the concerns of the agent (person). Damasio [94] terms cognitively generated emotional states – secondary emotions.

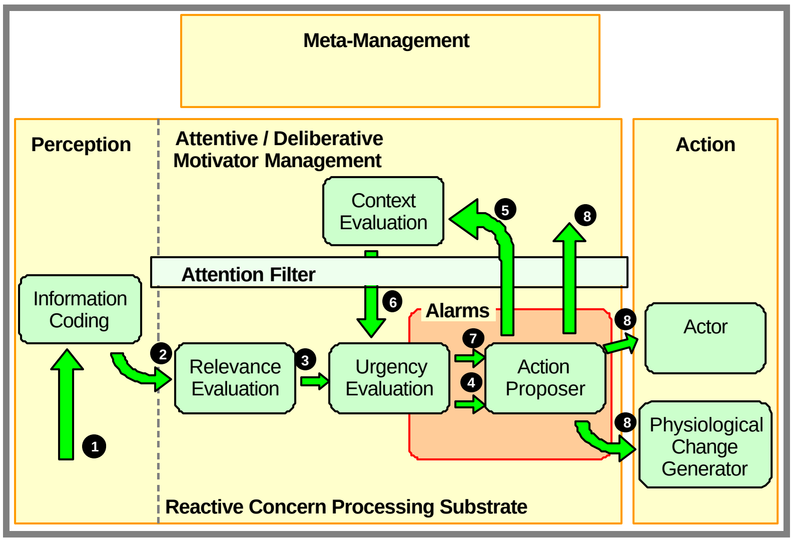


Figure 1.4 Secondary Emotion featuring Deliberative Context Evaluation

Agents that have a deliberative layer (or at least an active attention mechanism and working memory) are in theory better able to evaluate the true seriousness of a situation – and therefore produce a more measured emotional response. Figure 1.4 shows the flow of information for a *Secondary Emotion* that relies on a deliberative evaluation of coping strategies.

In humans the difference between the generation of a primary or a secondary emotion can simply be a question of the initial urgency attached to the stimulus. In Figure 1.4 the emotion process proceeds from (**1**) through (**4**) as per a typical primary emotion. However, instead of triggering a full emotional response, only attention is captured before the context is deliberately evaluated at (**5**). At this point we could simply be reacting to a loud noise (a startle response), without actually assessing the context of the situation (if we were alone in a dark house a reactive context evaluation in the form of heightened arousal could already be enough to trigger a physiological emotional response). Having evaluated the context as serious (**6**), our alarm system kicks in and generates an emotion proper (**7**) and (**8**).

Although secondary emotions might not actually generate physiological change (which varies from individual to individual), they still utilise much of the machinery of primary emotions (global alarm system) when capturing and diverting attention. Secondary emotions can also be triggered by deliberative thought processes as in Figure 1.5.

***Tertiary* emotional states**: such as feeling humiliated, ashamed, or guilty, can be further characterised by a difficulty to focus attention on urgent or important tasks. These emotions cannot occur unless there is a meta-management layer to which the concept of "losing control" becomes relevant. Without meta-management, which provides some sort of evaluation and control of thought processes, there cannot be any loss of control: you can not lose what you do not have [Sloman 99]. Tertiary emotions correspond to secondary emotions which reduce self-control.

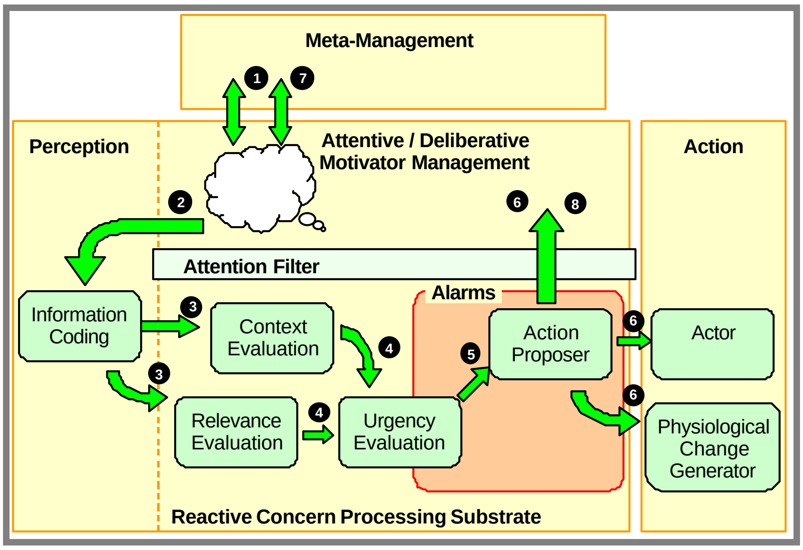


Figure 1.5 Tertiary Emotion featuring a loss of control of Deliberative Management

Finally, we can have the special case of secondary emotions which reduce control of the attention mechanism – the machinery of *Tertiary Emotions*. Figure 1.5 shows the information flow within a tertiary emotion.

Tertiary emotions – such as grief and longing – are characterised by a difficulty to focus attention on urgent or important tasks. The architecture must therefore support a meta-management layer that attempts to manage the deliberative thought process (**1**).

Normal deliberative thought processes trigger the mechanisms of primary emotions resulting in a signal of control precedence which presses for and temporarily gains control of the attention mechanism (**2**) through (**6**). Meta-management processes are still able to detect this change and re-evaluate the situation as less important than the current task and so regain control (**7**). However, thoughts keep returning to the object of concern (as the reactive context evaluation has yet to adjust to the new situation), and ongoing deliberative processes trigger further interruptions (**8**) – resulting in an emergent perturbant state.

1.4 Learning by Building Emotional Agents

Design-Based Approach to Mind Design

There are a number of ways to elucidate the emotion process, but we feel that complex systems can often best be understood through a *succession* of designs, in the downhill mode of invention. The design-based approach [Sloman 93; Beaudoin 94; Wright 97] takes the stance of an engineer attempting to build a system to exhibit the phenomena/behaviour of interest. Here, each design gradually increases our explanatory power and allows us to account for more and more of the phenomena of interest.

Society of Mind Architecture

As a starting point for our work, we have adopted Cañamero's [97] Abbott architecture. Abbott2 is an extension of Cañamero's original design, implemented in Pop11 (see section on the Gridland Environment) as a *Society of Mind* (SoM) with a single parent, and a number of child agents all of whom share a global blackboard. The parent agent is identified in the Gridland world as "Abbott", and has a physical presence that can be detected by the other agents that inhabit Gridland. The child agents form the Abbott architecture shown in Figure 1.6.

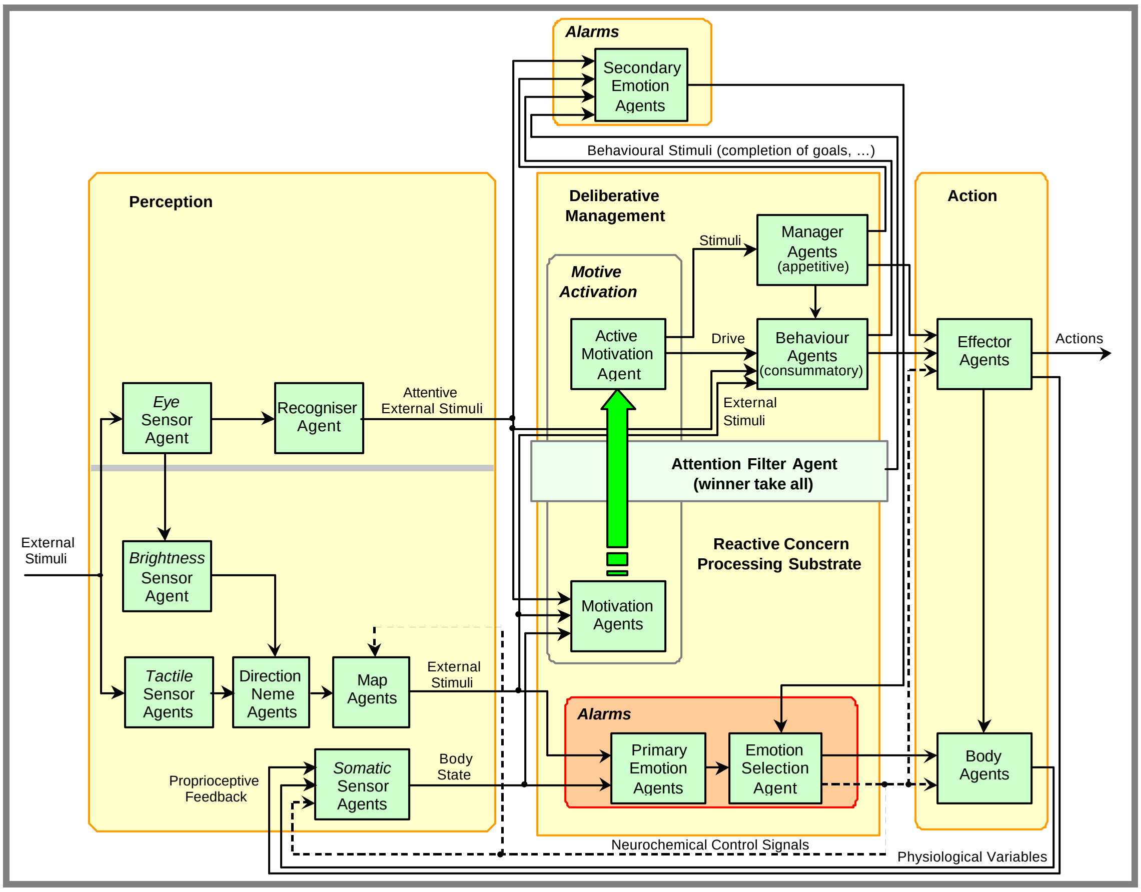


Figure 1.6 The Abbott2 Architecture - based on [Cañamero 97]

Multiple Concerns

Abbott is equipped with eight *action tendencies* with which to maintain its body state: Aggression; Cold; Warmth; Curiosity; Fatigue; Hunger; Thirst; and Self-protection. Each *action tendency* is represented in the architecture by a *motivation* agent which: (i) monitors the status of a single controlled variable (energy, temperature, …); and (ii) selects different behaviours to bring the variable back into range. Life is complicated by the fact that not all the controlled variables can be maintained within the desired range at the same time – in order to increase blood sugar Abbott needs to walk to find food, thus decreasing energy and increasing temperature. Abbott must therefore balance many competing *concerns*, eventually learning to select actions that address not one but multiple sources of motivation. But before Abbott starts to run, it must first learn to walk and attend to his most urgent needs one at a time.

Proto-Specialists

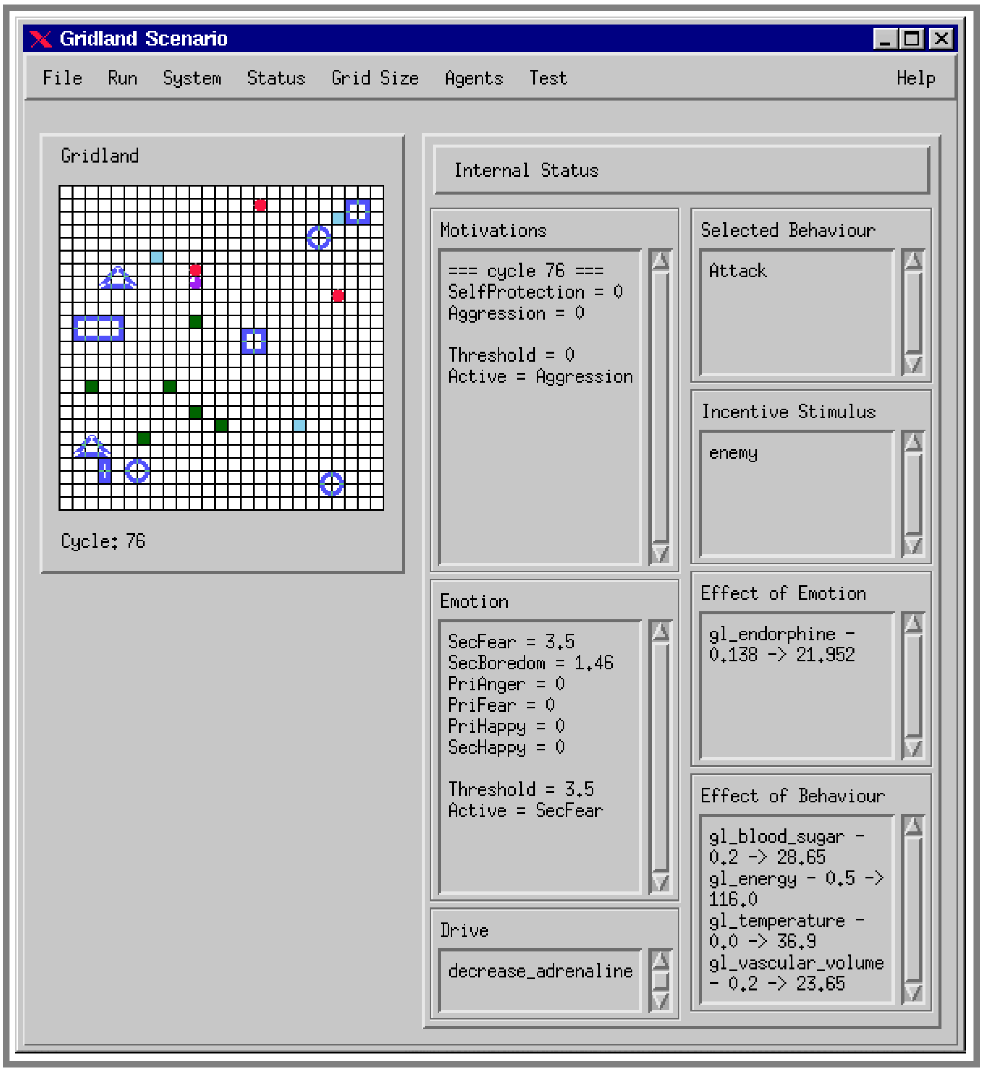
Minsky uses the term proto-specialist to refer to a “separate agency for each of several basic needs” [Minsky 87, page 165]. The Abbott architecture uses two classes of proto-specialists: *motivation* and *emotion* agents. *Motivation* agents are responsible for monitoring Abbott’s internal body state and responding to specific body needs. However, simply responding to the most urgent motivation at any moment in time inevitably leads to dithering as motivations with similar activation levels compete for control precedence. In Abbott this problem is partially addressed by including an artificial sharpening mechanism that operates on the current active motivation. *Emotion* agents provide this general mechanism to sharpen and enhance Abbott's sources of motivation: (i) they modify the activation level of the current motivation (through amplification or dampening); and (ii) they change the perception of certain internal variables. In a sense, *emotion* agents take on a motivator management role, identifying urgent and important situations (fear and anger), marking the successful completion of goals (happiness), or detecting situations in which the current strategy is failing and a new approach needs to be adopted (sadness).

The chemical control signals released by Abbott's *emotion* agents form part of Abbott's internal body state, and can trigger *motivation* agents directly (aggression and curiosity agents monitor the levels of adrenaline and endorphine respectively) – an angry Abbott can therefore strike out aggressively, or a curious Abbott start searching for novel stimuli. This is obviously an oversimplification of the link between neurochemicals and motivations, but provides an interesting starting point. By relaxing the constraint of a single chemical for a single motivation, it should prove possible for the presence of an enemy to elicit an emotion of fear or anger and, in both cases, lead to motivations of aggression or self-protection –depending on Abbott's previous history and current state. The background chemical mood can therefore represent a snap-shot of Abbott's internal state, and so act as an automatic context evaluation mechanism – if Abbott has been successful in achieving goals it will be in a happy state and therefore more likely to persist with new motivations.

Attention Mechanism

Abbott uses a simple attention mechanism to direct resources to the most urgent source of motivation. This attention mechanism is implemented as an *attention filter* agent within the *Society of Mind* model. Although the attention filter has a nominal threshold, there is a need to distinguish between the attention filter setting and modifying the activation level of the current motivation. Repetitive activity leading to boredom should ideally only affect the motivation that generated the repetitive activity (or better still the current *manager* agent). Whereas attempting to solve a difficult problem or responding to an urgent motivator should ideally raise the filter threshold to prevent interruption of the current motivation. The attention filter therefore plays a part in two distinct processes: (i) motivation selection; and (ii) motivation management – separating managed and pre-management motivators.

In Abbott's winner-takes-all attention filter strategy, it becomes meaningless to talk about raising or lowering the filter threshold setting – this is done implicitly when the *motivation* agent is selected. Without the ability to distinguish between actively managed and pre-management motivators, all transactions must remain in the common currency of activation energy. The *attention filter* agent can however give a boost to the activation energy of the selected motivation and thus provide Abbott with a motivator persistence mechanism.



Cycle: 76

```
Motivations
=== cycle 76 ===
SelfProtection = 0
Aggression = 0

Threshold = 0
Active = Aggression
```

```
Selected Behaviour
Attack
```

```
Incentive Stimulus
enemy
```

```
Emotion
SecFear = 3.5
SecBoredom = 1.46
PriAnger = 0
PriFear = 0
PriHappy = 0
SecHappy = 0

Threshold = 3.5
Active = SecFear
```

```
Effect of Emotion
gl_endorphine -
0.138 -> 21.952
```

```
Drive
decrease_adrenaline
```

```
Effect of Behaviour
gl_blood_sugar -
0.2 -> 28.65
gl_energy - 0.5 ->
116.0
gl_temperature -
0.0 -> 36.9
gl_vascular_volume
- 0.2 -> 23.65
```

Figure 1.7 Gridland Scenario - based on [Cañamero 97]

Blackboard Architecture

Abbott is implemented using a global blackboard architecture, although each agent is allowed its own private work area on the blackboard. Agents communicate by posting messages on the blackboard, and can respond to messages directly addressed to them or by eaves-dropping on messages posted between other agents. This communication transparency allows *emotion* agents to easily monitor the progress of the current motivation, detect *manager* agent failures, identify repetitive activity, etc.

Abbott SoM agents run asynchronously to each other and to their parent “Abbott” agent. However, as agents in Gridland can only move at the end of a World time-step (currently five clock cycles), external stimuli and actions are synchronised to the environment and World time. The SoM model also allows us to run different child agents at different rates: (i) by duplicating them in the processing order; and (ii) by specifying a cycle limit for each agent.

2 A Quick Tour of Brain Anatomy

2.1 The Somatic Marker Hypothesis

We are born with certain innate neural machinery capable of generating somatic states (both visceral and non-visceral) in relation to certain classes of stimuli – Damasio's [94] machinery of *primary emotions*. In addition to these innate capabilities, we also possess the ability to form systematic connections between categories of objects and situations on the one hand, and primary emotions, on the other. These learned associations and feelings – which Damasio has termed the mechanisms of *secondary emotions* [Damasio 96, page134] – are the somatic markers of the *somatic marker hypothesis*. These two emotion pathways are shown in Figure 2.1

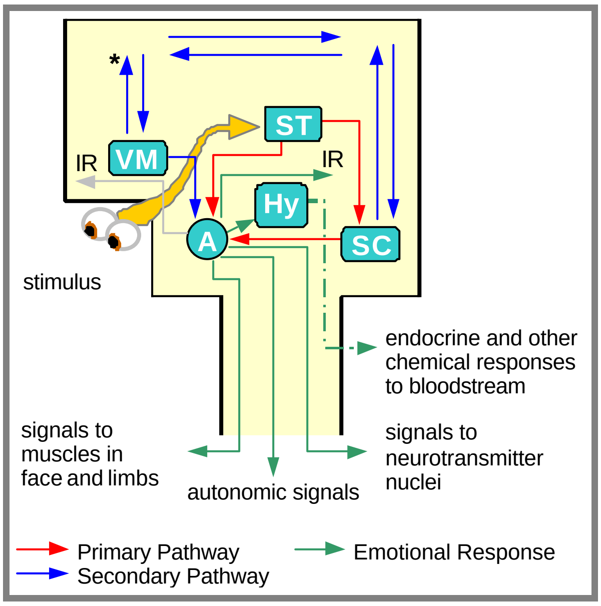

*
IR
VM
ST
IR
Hy
A
SC
stimulus
endocrine and other chemical responses to bloodstream
signals to muscles in face and limbs
autonomic signals
signals to neurotransmitter nuclei
Primary Pathway
Secondary Pathway
Emotional Response

Figure 2.1 Emotion Mechanisms [modified Damasio 96, page 137]

Primary Emotions. The black perimeter stands for the brain and brain stem. After an appropriate stimulus activates the amygdala (A) – via sensory thalamus (ST) and sensory cortex (SC) –, a number of responses ensue: internal responses (marked IR); muscular responses; visceral responses (autonomic signals); and responses to neurotransmitter nuclei and hypothalamus (Hy). The hypothalamus gives rise to endocrine and other chemical responses which use the blood stream route.

Secondary Emotions. The stimulus may still be processed directly via the amygdala but is now also analysed in the thought process, and may activate frontal cortices and the ventromedial prefrontal cortex (VM). VM acts via the amygdala (A). In other words, secondary emotions utilise the machinery of Primary Emotions.

2.2 The Emotional Brain

Emotions do not refer to a nice self-contained system of the brain, with a well-defined functional role, and a clear physical boundary. Emotions emerge from the interaction of many different systems, performing many different roles, and operating at many different levels within a biological agent architecture. LeDoux [94] believes that the best way to unravel the underpinnings of emotional life is to systematically study the neural pathways of the individual emotion systems – starting with the system that mediates fear.

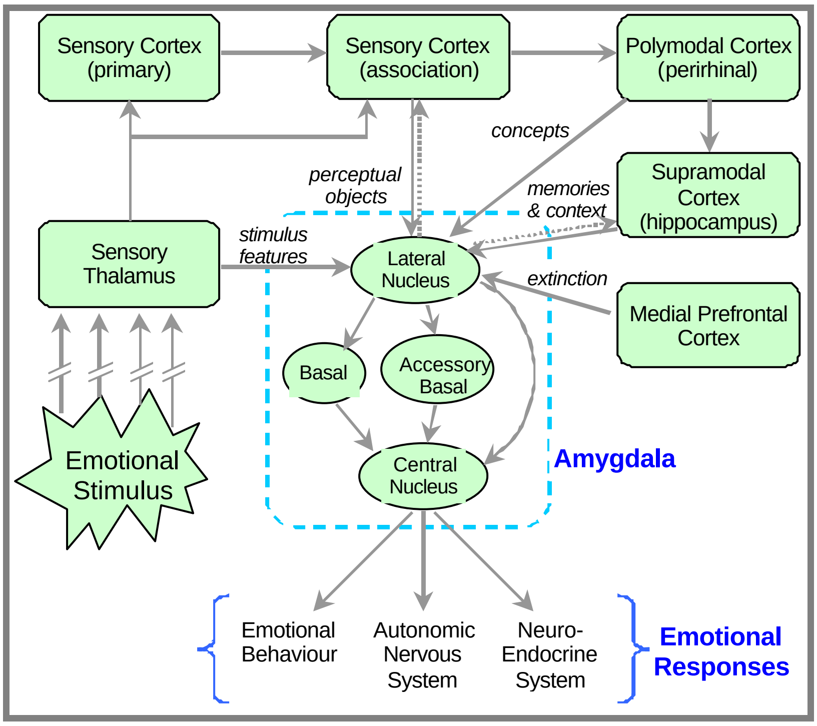


Figure 2.2 Amygdala Pathways in Fear Conditioning [modified LeDoux 95, 96]

The amygdala (Figure 2.2) lies squarely at the centre of the fear emotion complex: (i) the sensory thalamus provides a fast pathway to the amygdala, by responding to low-level features of the stimulus; (ii) the sensory cortex provides a path for more complex aspects of the stimulus (event/object) to reach the amygdala; (iii) the polymodal cortex creates concepts/associations between the different sensory modes (visual, auditory, and somatic); which then feed into (iv) the supramodal cortex (hippocampus) to allow explicit past memories of similar situations to affect the emotion process; and finally (v) the medial prefrontal cortex allows extinction of previously conditioned responses through habituation.

Figure 2.3a shows the high and low information-processing pathways of the brain that lead from emotional stimulus to emotional response. The low road provides the quick and dirty pathway for our immediate reactions. The high road leads through the sensory cortex and provides a more accurate representation of the stimulus, but takes a little longer to reach the amygdala. Whereas the sensory thalamus is biased towards evoking a response, the role of the sensory cortex can be viewed as that of preventing an inappropriate response (rather than producing an appropriate one). In the terminology of the *Emotion Process*, the sensory thalamus performs the initial relevance evaluation of the stimulus, and the sensory cortex performs part of the context evaluation process (relevance and context evaluation are labels used to describe operations that occur within the emotion process, that may or may not map on to discrete physical structures of the brain – see Figure 2.4 below).

2.3 Regions of the Brain Involved in Reasoning and Emotion

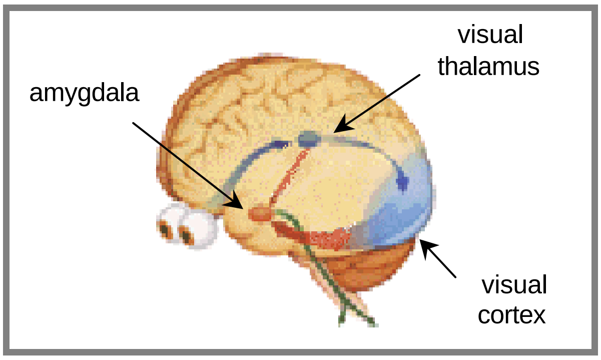


Figure 2.3a Low and High Roads to the Amygdala

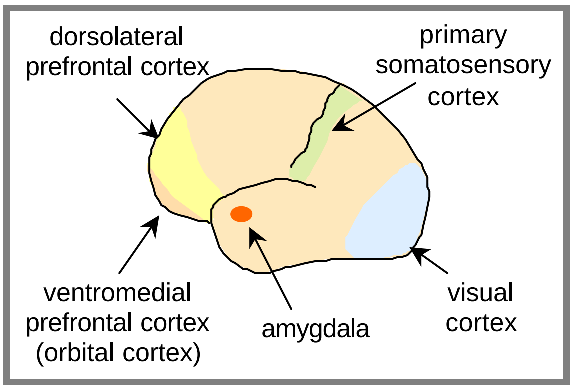

dorsolateral prefrontal cortex
primary somatosensory cortex
ventromedial prefrontal cortex (orbital cortex)
amygdala
visual cortex

Figure 2.3b External (Lateral) View of Left Hemisphere

We are now in a position to identify some of the key players in the emotion process. Figure 2.3a shows the high and low roads to the Amygdala for a typical fear response reaction (see LeDoux [96]). Figure 2.3b shows the lateral (or external) view of the left brain hemisphere – with the ventromedial prefrontal cortex, the dorsolateral prefrontal cortex, and the primary somatosensory cortex highlighted. Damage to these structures has a number of serious effects on cognition and emotion:

1) Damage to the ventromedial prefrontal cortices results in both compromised reasoning/decision making and emotion/feeling (especially in the personal and social domains).

2) Damage to the dorsolateral region compromises decision making but "Either the defect is far more sweeping, compromising intellectual operations over all domains, or the defect is more selective, compromising operations on words, numbers, objects or space, more so than operations in the personal and social domain." [Damasio 96, page 70]

3) Damage to the right primary somatosensory cortex disrupts the process of basic body signalling. Such damage also compromises reasoning/decision making and emotion/feeling, providing further evidence for the influence of *somatic markers* in decision making.

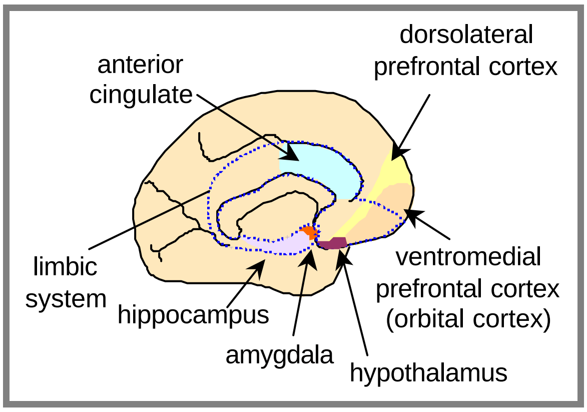


Figure 2.3c Internal (Medial) View of Left Hemisphere

Figure 2.3c shows the medial (or internal) view of the left brain hemisphere, and the region normally associated with the limbic system. The term "limbic system" does not actually refer to an anatomically unique region of the brain, but it is generally defined to include the hippocampus, the cingulate, the amygdala, and parts of the ventromedial prefrontal/orbital cortex. These areas have specific tasks in reasoning and emotion/feeling:

1) The hippocampus creates a representation of the context that contains not individual stimuli but relations between stimuli. In bringing together all the actors, the hippocampal system is able to generate explicit memories about emotional situations.

2) The anterior cingulate (along with the lateral prefrontal cortex to which it is interconnected) forms part of the frontal lobe attention network, "a cognitive system involved in selective attention, mental resource allocation, decision making processes, and voluntary movement control." [LeDoux 96, page 277]

3) The amygdala has been identified as being central to the fear emotion system, but its role in other emotion systems is still unclear. The amygdala does *not* appear to play a significant role in positive emotions [Damasio 99, pages 62-65]

2.4 Extending the Motivated Agent Framework

Mapping Damasio's emotion mechanisms on to our motivated agent framework (see Figure 2.4) allows us to clarify a little better what exactly is meant when talking about "secondary emotions utilising the machinery of primary emotions". Things get a little complicated in biological systems as attention and interruption are in part mediated by arousal (a physiological change) – we will circumscribe this problem by defining physiological change to exclude effects on the attention system. We should also point out that somatic markers offer just one mechanism through which the machinery of secondary emotions can trigger the machinery of primary emotions – there are many other pathways from the high-level cortex to the *anterior cingulate* and *limbic* system without first going though the *somatosensory cortex*.

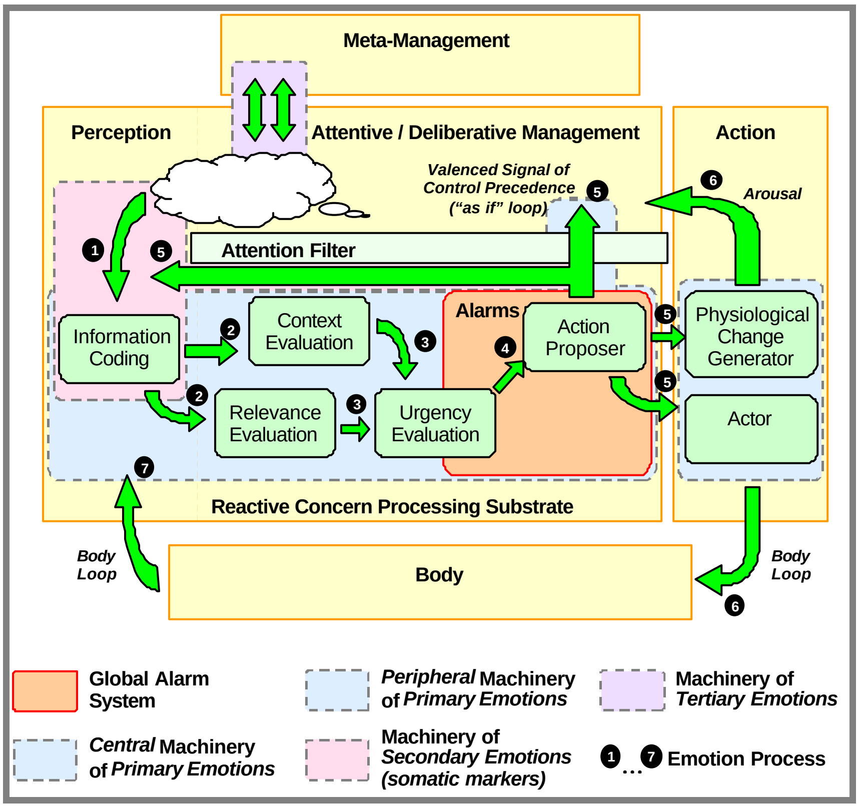

Meta-Management
Perception
Attentive / Deliberative Management
Action
Valenced Signal of Control Precedence ("as if" loop)
Arousal
Attention Filter
Information Coding
Context Evaluation
Alarms
Action Proposer
Physiological Change Generator
Relevance Evaluation
Urgency Evaluation
Actor
Reactive Concern Processing Substrate
Body Loop
Body
Body Loop
Global Alarm System
Peripheral Machinery of *Primary Emotions*
Machinery of *Tertiary Emotions*
Central Machinery of *Primary Emotions*
Machinery of *Secondary Emotions (somatic markers)*
1 ... 7 Emotion Process

Figure 2.4 Emotion Mechanisms and the Body Loop

Conceptually, there are two different parts to the machinery of primary emotions: (i) the *central* machinery is responsible for detecting the relevance of the event and interrupting attentive processing; whereas (ii) the *peripheral* machinery generates overt action and physiological change. An emotion can utilise the *central* machinery of primary emotions, without triggering the *peripheral* machinery (this is especially true of the more cognitive *tertiary* emotions such as guilt). If we are generous and allow the *central* machinery of primary emotions to include most of the limbic system (i.e. the *anterior cingulate* and *amygdala*), then most theorists are likely to agree that "secondary emotions utilise the machinery of primary emotions" – especially as the *anterior cingulate* plays an important role in attention. However, not all parts of the machinery of primary emotions are utilised by all secondary emotions (not all parts of the machinery are even utilised by all primary emotions), and parts of the machinery play a critical role in non-emotional processes as well

3 Gridland Environment

3.1 SIM_AGENT Toolkit

The SIM_AGENT toolkit is a general purpose toolkit for investigating different types of agent architectures. The toolkit supports both rule-based and sub-symbolic (i.e. neural) mechanisms, and comes with extensive on-line help and teach files. A description of the toolkit can be found in Sloman and Poli [96] – the main features are summarised below:

- Minimal ontological commitment supporting many different kinds of objects with very different architectures.
- External behaviour which can be detected by or affect other objects or agents.
- Internal behaviour involving mechanisms for changing internal control states (percepts, beliefs, maps, goals, …) that are not directly detectable by others.
- Rich internal architecture within agents allowing several rule-sets and rule-families to run in simulated parallelism. An architecture can therefore support several levels of sensory perception, reactive and deliberative processes, neural nets and other trainable sub-mechanisms.
- Use of classes and inheritance to allow generic behaviour.
- Control of relative speed allowing both agents and sub- mechanisms within agents to run at different relative speeds.
- Rapid prototyping through the incremental compilation of the Pop11 environment.

The Gridland extension provides the SIM_AGENT toolkit [Sloman and Poli96] with a graphical interface and simulated environment in which to explore the design-space of autonomous agent architectures. The toolkit has been heavily influenced by Cañamero's work on the Gridland World [Cañamero97], growing out of the design requirements for the initial implementation of the Abbott architecture. The key benefits offered by the toolkit are:

- A mouse driven interface
- Run, pause, and single step a simulation.
- Load, save, and reset a simulation.
- Set trace and debug options for any agent.
- Multiple windows to display the agent's internal status.
- Controllable scheduler loop for real-time interactions.
- Interactive control and display of agent status.
- Capture trace/debug windows to disk
- Uses the standard SIM_AGENT toolkit.
- Easily expandable.

3.2 Gridland Virtual Machine

The various files that form the Gridland environment are best described within the context of a stack of “virtual machines” – machines with no definable physical form – with the operating system at the bottom and the target autonomous agent architecture at the top. This whole structure is shown graphically in Figure 3.1 below.

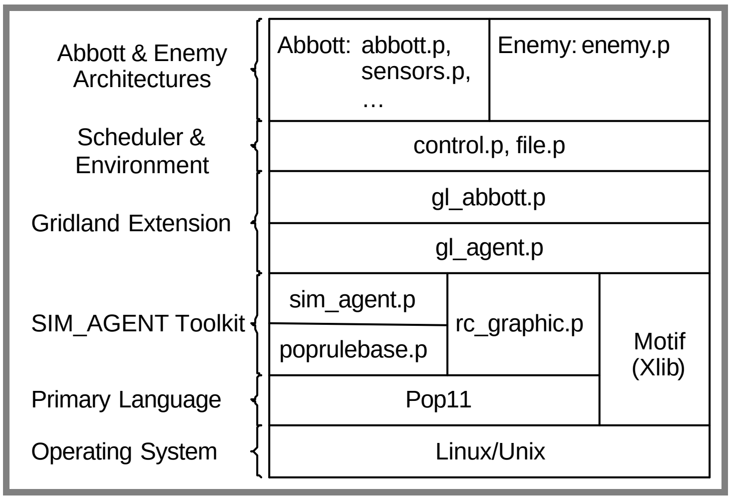


Figure 3.1 Gridland Virtual Machine

The Gridland environment makes extensive use of the Motif widget set and **rc_graphic.p** (Relative Co-ordinate Graphic) library. A number of high-level routines are included to aid the creation of popup and cascading pulldown menus, as well as access to the trace and debug scrollable windows. Object classes, mixins, and methods, for the physical attributes of the Gridland environment as well as the Motif widget set are stored in the **gl_agent.p** library. Classes, mixins, and methods, specific to the Abbott architecture (the *Society of Mind* model as well as support for physiological variables such as blood_pressure, heart_rate) are stored in the **gl_abbott.p** library.

3.3 Scheduler

The scheduler and rules for the scenario form the final part of the Gridland environment. The Gridland scheduler hooks into the standard SIM_AGENT toolkit (see Figure 3.2) to provide the toolkit with a powerful mouse-driven graphical interface. A simple command queue is used to synchronise mouse and menu events within the scheduler.

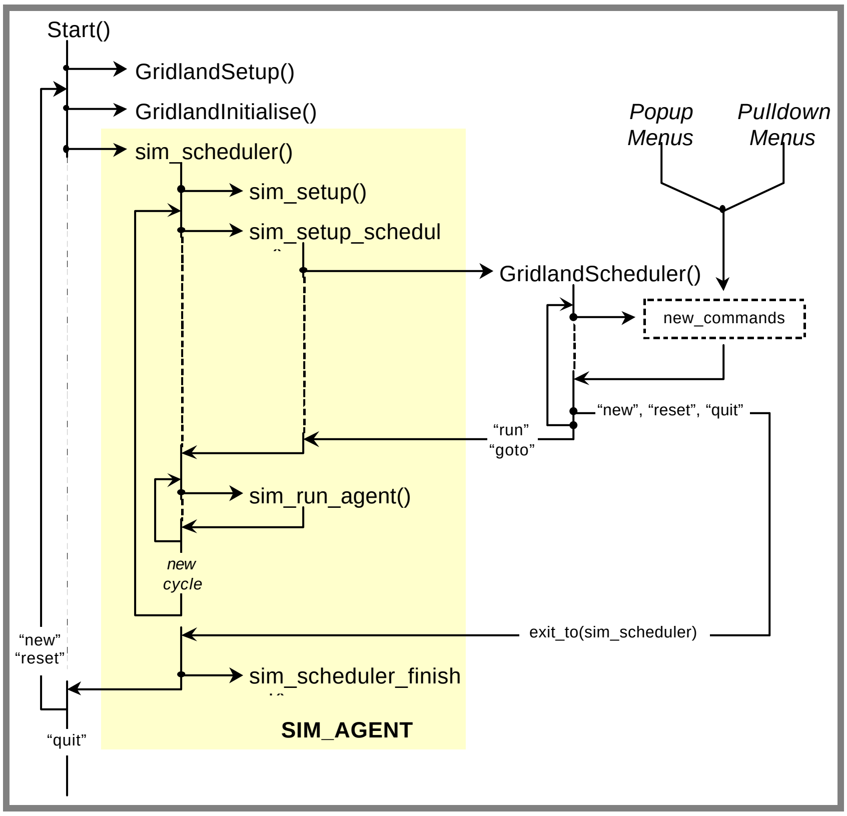
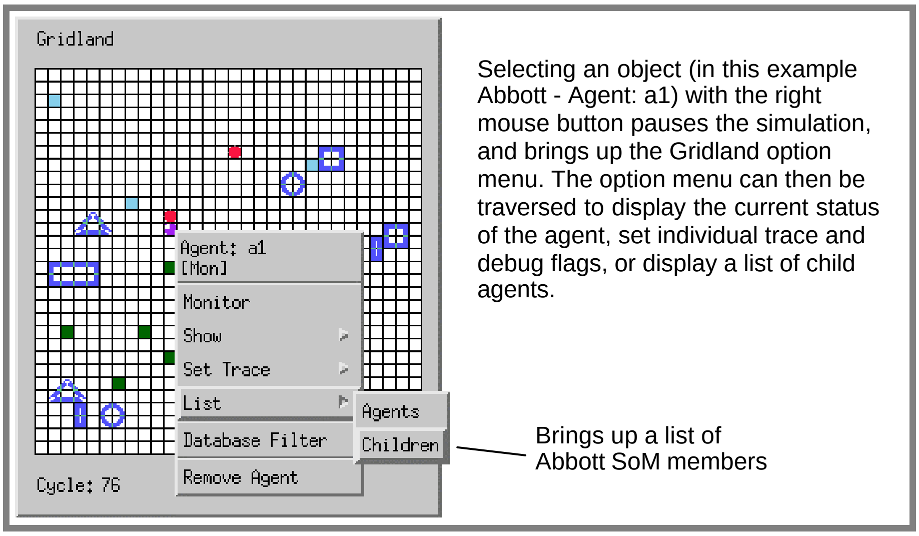


Figure 3.2 Scheduler

3.4 Commands and Menus

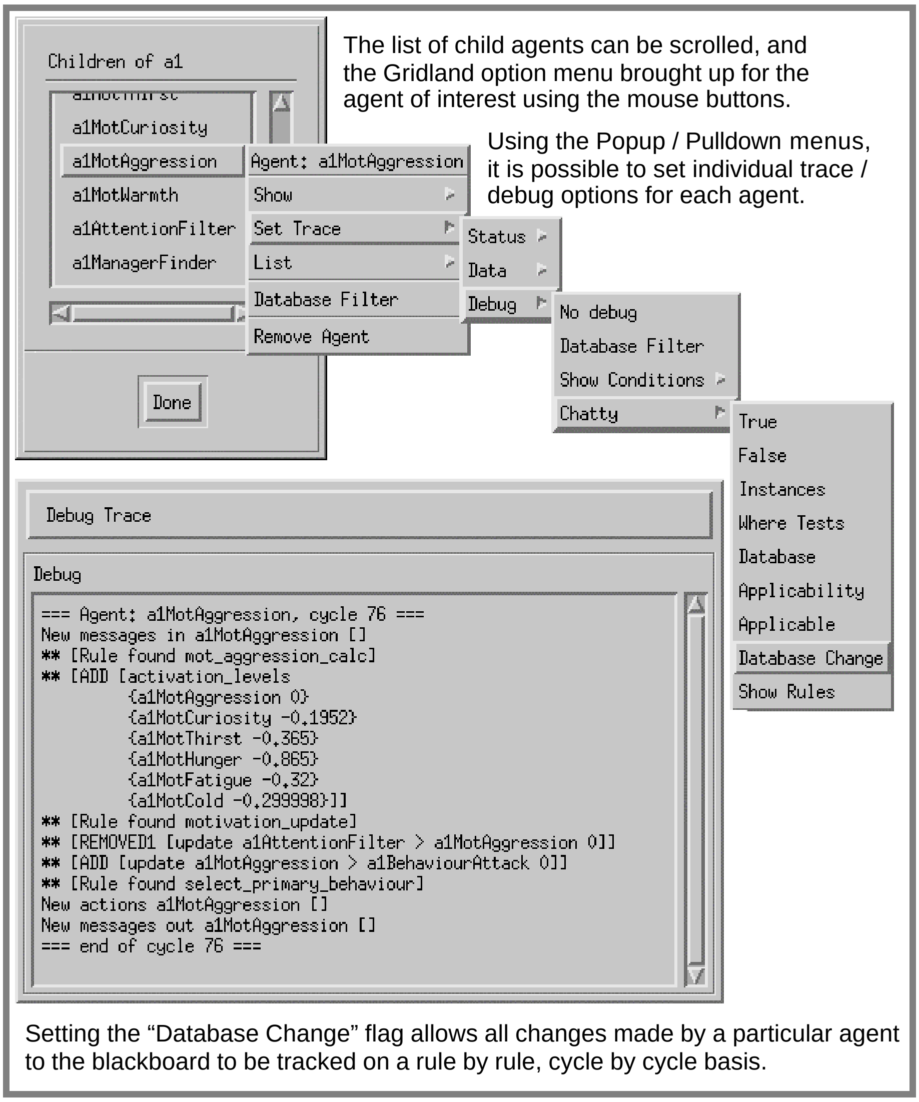
The Gridland commands fall into three basic categories: (i) file commands associated with resetting, loading and saving experiments; (ii) run commands concerned with pausing, single stepping and running the simulation; and (iii) system commands concerned with setting cycle times and saving trace and debug results to disk. In addition to these basic commands, other menu options can be used to select between different status windows, display a list of all agents, or provide simple help facilities. The graphical interface also gives us the chance to select and interactively set debug and trace flags for individual agents (see Figure 3.3).



Selecting an object (in this example Abbott - Agent: a1) with the right mouse button pauses the simulation, and brings up the Gridland option menu. The option menu can then be traversed to display the current status of the agent, set individual trace and debug flags, or display a list of child agents.

Brings up a list of Abbott SoM members

Figure 3.3a Interactive Debug Facilities



The list of child agents can be scrolled, and the Gridland option menu brought up for the agent of interest using the mouse buttons.

Using the Popup / Pulldown menus, it is possible to set individual trace / debug options for each agent.

```
=== Agent: a1MotAggression, cycle 76 ===
New messages in a1MotAggression []
** [Rule found mot_aggression_calc]
** [ADD [activation_levels
        {a1MotAggression 0}
        {a1MotCuriosity -0.1952}
        {a1MotThirst -0.365}
        {a1MotHunger -0.865}
        {a1MotFatigue -0.32}
        {a1MotCold -0.299998}]]
** [Rule found motivation_update]
** [REMOVED1 [update a1AttentionFilter > a1MotAggression 0]]
** [ADD [update a1MotAggression > a1BehaviourAttack 0]]
** [Rule found select_primary_behaviour]
New actions a1MotAggression []
New messages out a1MotAggression []
=== end of cycle 76 ===
```

Setting the “Database Change” flag allows all changes made by a particular agent to the blackboard to be tracked on a rule by rule, cycle by cycle basis.

Figure 3.3b Interactive Debug Facilities

4 Towards an Infant-Like "Emotional" Agent

4.1 Concern-Centric Design

Abbott3 (see Figure 4.1) represents the latest in our series of broad-but-shallow agent designs that attempt to address the requirements of virtual information-processing architectures for human-like minds [Beaudoin 94; Wright 97; Complin 97; Sloman 99].

Our latest incarnation of Abbott marks quite a big departure from Cañamero's [97] original design. However, we feel that it is still appropriate to maintain the convention of using the label “Abbott” to refer to the collective *Society of Mind*. In cases where ambiguity is likely to exist, we will explicitly refer to the two designs as *Abbott3* and *Cañamero's original design*. Finally, it is worth emphasising that although we describe our architecture as being broad-but-shallow, it is nevertheless based on the deep(er) theory of concern-processing requirements for intelligent autonomous agency.

In this section we will describe how the competence layers in Abbott's concern-centric architecture co-evolve, and identify the mechanisms that lead to the emergence of “emotional” states.

Co-evolution In Abbott

One of the deficiencies we identified in the subsumption-style architecture [Brooks 86] was the problem of command fusion and the potential for a reversal of concern-processing priorities. A subsumption-style architecture should be capable of being partitioned at any level, with the layers below forming a complete control system. This places an implicit requirement on the designer to capture the primary concerns of the agent in the behaviours of the base layer of the architecture. In Brooks’ original proposal, these low-level concern-processing mechanisms are subsequently subsumed by the more specific behaviours represented in the higher-level competence layers. This means that unless the primary concerns are then replicated at each level, there exists the very real possibility of overriding them with lower priority concerns as the architecture evolves.

We address the problem of command fusion by adopting a concern-centric design stance that recognises the need to allow the different levels of competence (coping strategies) to co-evolve. In the following discussion we will describe this evolution process as we grow our agent architecture from the base level Abbott3a (Figure 4.2) into the fully fledged Abbott3 (Figure 4.1).

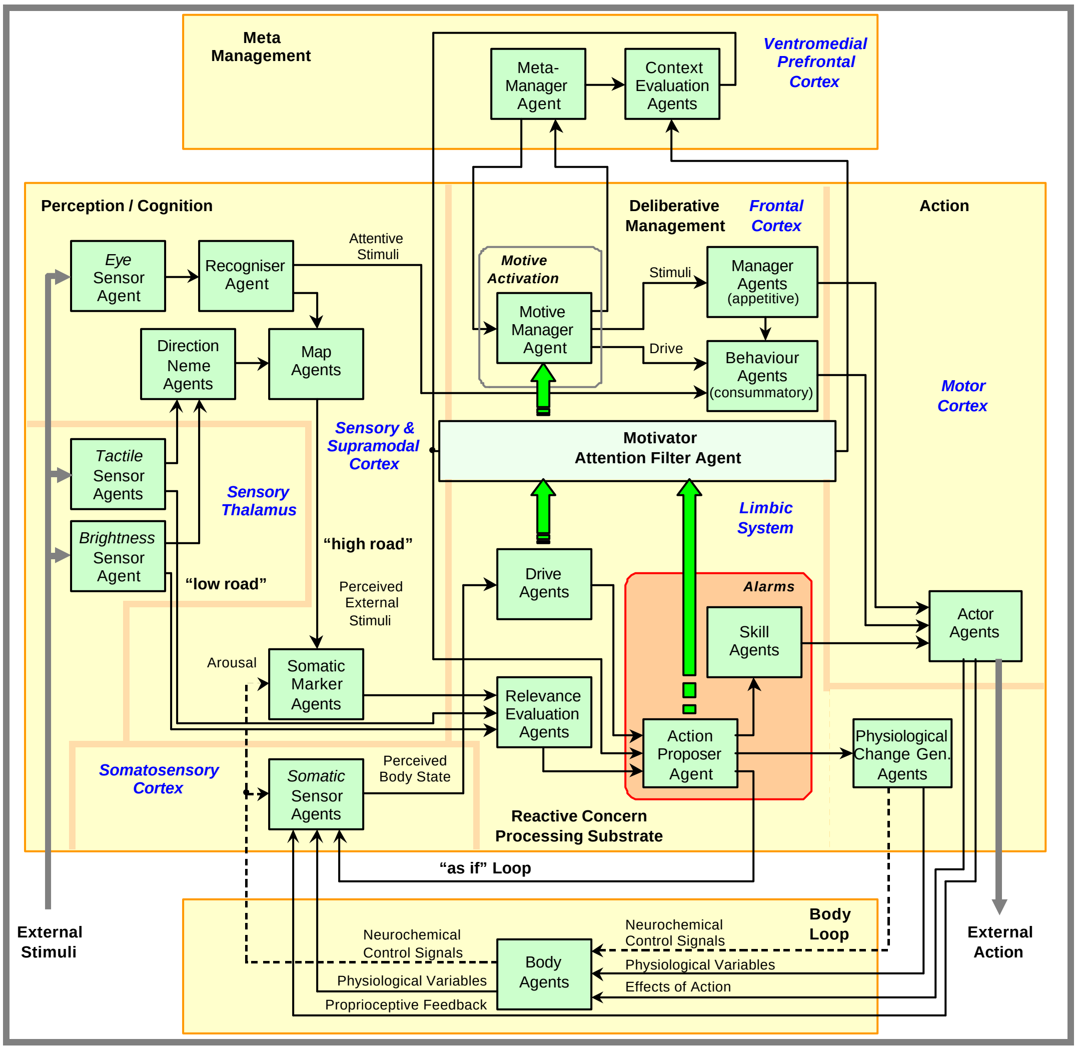


Figure 4.1 Abbott3 Architecture

Emergent Emotional States

The main difference between the design of Abbott3 and its predecessors (Abbott and Abbott2), is the absence of a clearly demarcated emotion system (Cañamero's [97] original design called for *emotion* agents to act as proto-specialists in a similar style to *motivation* agents). We believe that "emotions" are emergent mental states *caused* by the interaction of a variable number of intricately connected cognitive systems (i.e. systems that mediate arousal, attention, perception, concepts, memories, and physiological change) operating at different information-processing levels of brain. Our approach towards elucidating emotions in Abbott3, is to replicate some of these systems at the information-level, and then explore the possible pathways through which emotional states can emerge.

In a sense, we are advocating a systems theory of emotion (as part of the more general requirements for intelligent autonomous agency). We must therefore remain vigilant to the accusation that the flexibility of our approach makes it possible to explain almost any data and therefore makes it hard to formulate concrete predictions with which to test its validity. To counter such a claim we must re-emphasise the fact that our approach not only consists of an architectural framework (our three-layered model derived from an information-level analysis of the requirements for intelligent autonomous agency – see Figure 1.1), but also a design-based research methodology. As we add more depth to our agent architectures, we will be forced to make design decisions that will lead to concrete predictions (such as the number and type of reactive concern-processing mechanisms, and their implications on the types of primary, secondary and tertiary emotional states they can support).

We can make some tentative predictions (even if they are hard to verify) such as: (a) we would expect tertiary emotions to be more cognitive in nature and not easily distinguishable by physiological measurement alone – as they are unlikely to map on to unique/distinct reactive concern-processing systems; (b) we would expect secondary and tertiary emotional states to appear later in the development of a human mind, with secondary emotions subject to more cultural variability. Tertiary emotional states (such as those normally associated with grief, infatuation, and anxiety) should exhibit statistically less cultural variability than secondary emotions, as their perturbant nature arises out of a mismatch between cultural conditioning and the more universal mechanisms of primary emotions.

A single emotion type can cover a wide range of forms and intensities (with an even wider range of associated securities, insecurities, dreams, and feelings) – i.e. *being in love* covers: romantic interest; infatuation; longing; intense passion; mature love. To aid the verification of our predictions, one possible avenue for future research would be to develop a robust scheme for mapping sub-classes of common emotion types (possibly by linguistic labels) on to the underlying concern-processing mechanisms active in the emotion process. But first, we need to develop a series of emotional agents within which to elucidate the human emotion process, and clarify what exactly we mean at the information-level when we talk about love, hate, envy, and joy.

Our design is still far too shallow to claim that we can actually support human-like (or even infant-like) emotions, but as the following discussion will show, we can still usefully elucidate the emotion process within such a framework.

4.2 The Emergence of Emotion

Competence Level 0

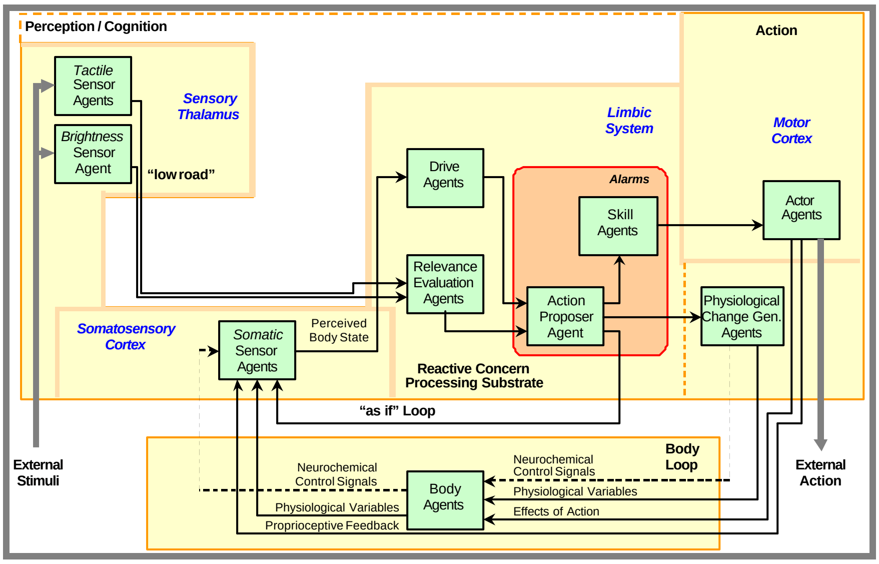


Figure 4.2 Abbott3a Competence Level 0

In keeping with the subsumption philosophy, Abbott3's base competence level (Figure 4.2) can actively sense the environment and respond to its basic needs. Our agent is also able to pursue an agenda (that of ensuring that a number of physiological variables are maintained within a desired range), and can even be said to possess a primitive personality in the form of a motivational profile that can be biased towards self-preservation, eating, or drinking (see section). We will call our base agent Abbott3a.

Abbott3a has two sources of motivational drive: (a) homeostatic drives – represented by *drive* agents; and (b) non-homeostatic drives – represented by the actions of *relevance evaluation* agents. The *drive* agents respond to error-signals in a controlled variable (i.e. blood sugar level or vascular volume), whereas the *relevance evaluation* agents detect significant features of the environment such as the colour/brightness of objects. These motivational drives are then able to trigger action through *skill* agents and/or generate real or vicarious physiological change through the *physiological change generator* agents. This latter *affective* pathway allows sustained activity to be initiated by one-off external events (partially addressing the persistence problem by providing a primitive motivational sharpening mechanism), and enables Abbott to respond to external incentives. Unfortunately, Abbott3a does not yet support learning, and is therefore unable to generate motivational control states from internal incentives.

Change in physiological arousal affects the perception of Abbott's internal somatic state, and thus allows one motivational drive to inhibit (or enhance) another. Inhibition occurs at the level of agent concerns, and not the individual behaviours as in the more common ethologically inspired behaviour-based architectures. We are thus able to implement a primitive attention mechanism to direct behaviour towards alleviating the most pressing concerns – here we are not advocating a system-level “winner-take-all” mechanism, but rather a selective mechanism that biases the type and number of concerns attended to at any one time. Concern inhibition acts as a first stage attention filter, with more complex *active* mechanisms evolving as additional competence levels are added to the architecture.

Competence Level 1

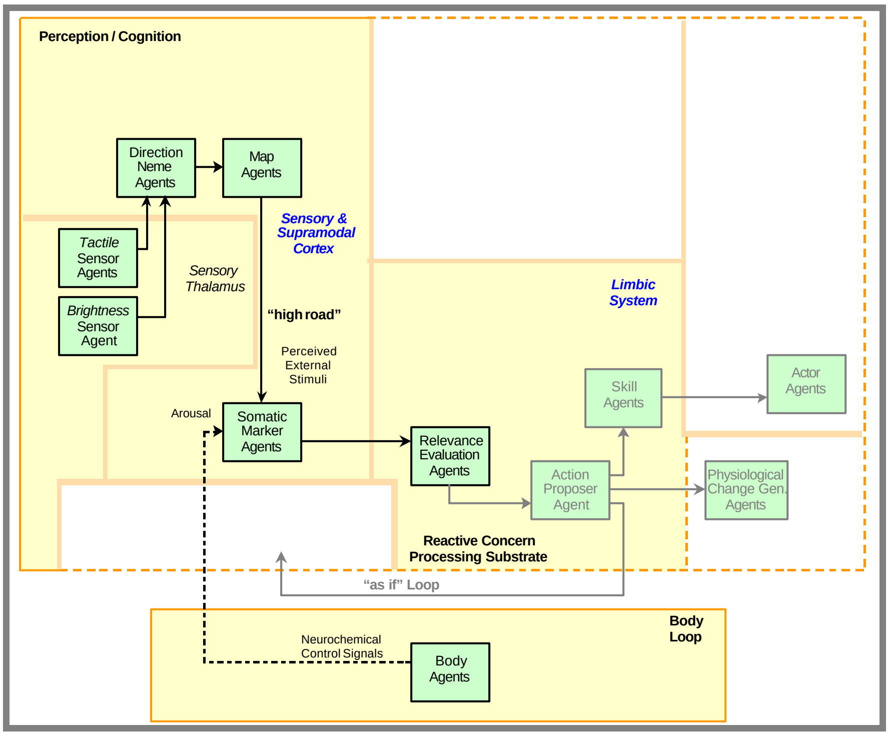


Figure 4.3 Abbott3b Competence Level 1

Figure 4.3 shows competence level 1 of the Abbott3 architecture. These level 1 competences dove-tail into the lower level 0 mechanisms without extensively redesigning the agent – i.e. we comply with the ethos of the subsumption style architecture, but adopt the practical stance of allowing our levels of competence to co-evolve. We will call our new agent Abbott3b.

The *direction neme* and *map* agents integrate Abbott's sight and touch modalities to produce *percepts* – internal representations of distinct objects in Abbott's immediate environment. These *percepts* are grounded in Abbott's level 0 concern-processing mechanism through *somatic marker* agents (somatic markers are used to mark percepts that are coincident with aroused body states). Abbott is thus able to supplement its innate *primary appraisal* mechanism (relevance evaluation [Frijda 86, page 401]) with signals generated with respect to learnt "affective" experiences.

The addition of somatic markers, in combination with a primitive attention mechanism, allows our simple agent to exhibit rudimentary *primary* and *secondary* emotions. Objects or events in the external world reach the *relevance evaluation* agent (through the *sensory* agents in the case of *primary* emotions, and via *somatic marker* agents in the case of *secondary* emotions) and cause the action proposer to switch attention and signal a change in the agent's somatic state (through real or vicarious pathways to the *somatic sensor* agents). Although there is no *self* to feel the emotional episode, we have at least met the main criteria normally associated with emotional states – i.e. that of interruption of attention, valence, and motivational attitude. We however attach the label "rudimentary" as there is still no deliberative functionality within these first two levels of competence (Damasio and Sloman attach the label *secondary* emotions to events triggered by deliberative thought processes).

Competence Level 2

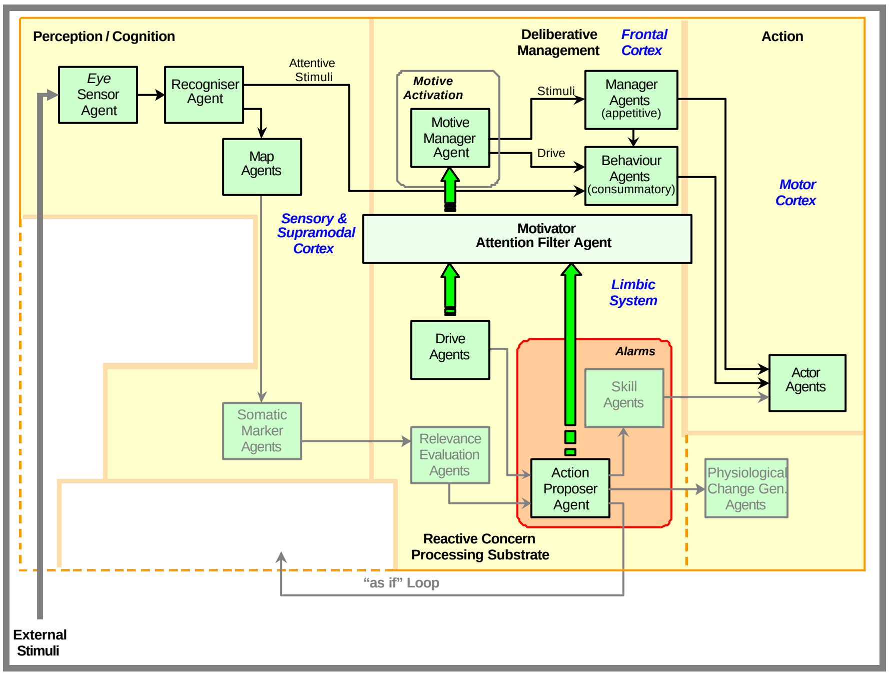


Figure 4.4 Abbott3c Competence Level 2

Figure 4.4 shows competence level 2 of the Abbott3 architecture. Each additional level adds new degrees of competence to the level below, without completely subsuming the original functionality (levels of competence increase the coping strategies available to the agent).

Our simple skill-based action selection mechanism is enhanced by the addition of a deliberative motive management layer capable of actively managing behaviours to meet Abbott's many competing needs. As deliberative action takes both time and requires access to limited deliberative reasoning resources, the original functionality encapsulated within Abbott's *skill* agents is still utilised by the *action proposer* agent in situations that require immediate action using simple perceptual (i.e. proprioceptive) feedback.

Deliberative and reactive action selection are intimately connected. As well as driving behaviours, the *recogniser* agent also activates the reactive *map* agents – which in turn feeds into the reactive concern-processing substrate. This allows our agent to produce affective reactions to deliberatively triggered images. Furthermore, Abbott's reactive concern-processing mechanism can also interrupt deliberative management and thus influence all levels of the architecture – receiving input from attentive perception, and acting through a global alarm system.

Abbott's primitive inhibition and fatigue motivator attention mechanism is supplemented with an active *attention filter* agent. This gives us the flexibility to produce a single drive-based motivator for urgent (highly insistent) sources of motivation, and still allow multiple non-urgent motivators to surface and be decided simultaneously. Finally, we have added the restriction that a *behaviour* agent's incentive stimuli can only come from the *recogniser* agent (i.e. the attended to object), implementing a form of perceptual attention within Abbott.

Competence Level 3

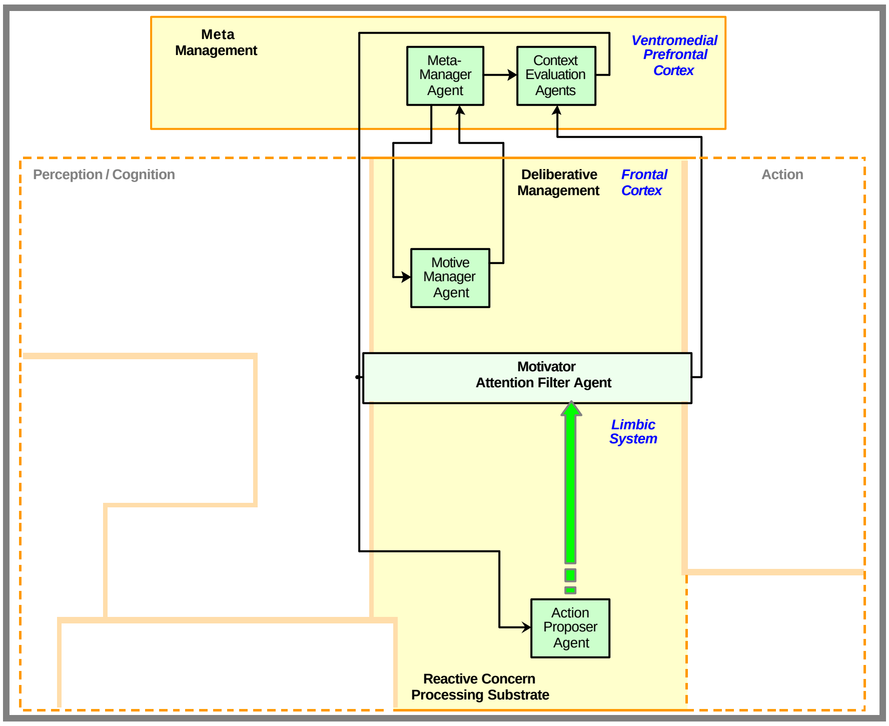


Figure 4.5 Abbott3d Competence Level 3

Figure 4.5 shows competence level 3 of the Abbott3 architecture. With the addition of a meta-management layer, we finally attain the basic three-layered model for intelligent autonomous agency introduced in section. Meta-management has two main responsibilities within the architecture: (i) it actively manages deliberative motivator management – for example, ensuring that the agent does not expend too much energy attaining low importance goals; and (ii) it provides the reactive concern-processing substrate with more accurate input as to the coping ability (context evaluation) of the deliberative management layer.

Although the fine detail still needs to be worked out (i.e. how behaviours, skills and managers are created and adapted), we have demonstrated a plausible pathway for the computational evolution of a cognitively inspired mind – through added levels of competence in a subsumption style *Society of Mind* architecture. By grounding each competence level in the concern-processing mechanisms of level 0, we are also able to offer a plausible pathway for the development of mind during the lifetime of the individual.

Competence Level 3 Emotions

We can complete our Abbott architecture by adding the society members that make up the level 3 competence layer (Figure 4.5). Abbott3d's *meta-manager* and *context evaluation* agents contribute the final cognitive elements to the emotion process picture described in chapter – allowing us to elucidate the *primary*, *secondary* and *tertiary* emotion pathways:

1) *Primary Emotions* utilise two different eliciting pathways: (a) the “low road” from the early *sensory* agents (tactile and brightness); and (b) the “high road” through the *direction neme*, *map*, and *somatic marker* agents. The low road represents the route for Abbott's *innate* emotional responses to external stimuli,

and the high road for stimuli previously *associated* with earlier primary emotional episodes. The relevance of the external stimuli are assessed by a small number of *relevance evaluation* agents – resulting in the information-level equivalent of the relevance signals of pleasure, pain, curiosity and desire (see section 1).

The *action proposer* agent makes a heuristic estimate of the importance/urgency of the situation/event based on the relevance signals – resulting in any of: (a) selection of a *skill* agent in very urgent situations; (b) activation of a *physiological change generator* agent in situations that might require physiological arousal; (c) generation of a motivator in situations that require deliberative attention; (d) modification of the *somatic sensor* agents through the "as if" loop. Motivator generation either occurs directly via the *action proposer* agent, or indirectly through the changed somatic state and *drive* agents.

This control process replicates the information flow for a primary emotional state shown earlier in Figure 1.3. A primary emotional state emerges when the generated motivator attains control precedence and is adopted by the *motive manager* agent – i.e. when its insistence level is higher than the threshold defined by the *attention filter* agent. Valence is either attached to the situation/object through a change in somatic state (real or through the vicarious "as if" loop), or directly associated with the motivator itself.

2) *Secondary Emotions* are emergent emotional states that require deliberation at some point in the emotion process. For example, Damasio [96] concentrates on emotions generated in response to specific situations, events, or objects which have previously been paired with primary emotions, but are now triggered by deliberative thought processes; whereas Sloman [99] also highlights emotions generated with respect to the planning process itself – i.e. when relevant risks are noticed, progress assessed, and success detected. Here we will concentrate on emotions generated in response to inner perception and action within the deliberation process itself.

Abbott's *meta-manager* agent continually monitors the *motive manager* agent, and is thus able to detect if, for example: a relatively low-level motivator is taking too long to satisfy; repeated behaviours are failing; the same low-level motivators are always being attended to; or a behaviour succeeds in satisfying a motivator. This type of information (along with the current threshold of the *attention filter* agent) allows the *context evaluation* agent to assess the effectiveness of the current coping strategy adopted by the deliberative layer. In situations where the current strategy is not working, the *action proposer* agent can use this context information to interrupt the *motive manager* agent with a new motivator, replicating the information flow for a secondary emotional state shown earlier in Figure 1.3/1.4.

3) *Tertiary Emotions* are normally associated with Damasio's class of secondary emotions – described in the context of competence level 2 emotions above. After the adoption of the new motivator (within the secondary emotion process), the *meta-manager* agent evaluates the new motivator as irrelevant and signals both the *context evaluation* agent and the *motive manager* agent. Control of attention is regained through context evaluation of the current situation, allowing the action proposer to evaluate a relevant event as non-urgent. However, repeated triggering of the secondary emotion via subsequent actions

of *recognise* (or *behaviour*) agents leads to a perturbing state. This temporary loss of control of attentive processing replicates the information flow for a tertiary emotional state shown earlier in .

In a sense, it is the difference in the adaptation rates of reactive and attentive meta-management processes to new situations that leads to the emergence of *tertiary* emotional states – the emergent state is terminated when the reactive motivator meta-management mechanisms (in the form of *somatic marker* agents) have had a chance to catch up with the new situation, which in the case of strong attachment concerns may never completely happen (i.e. with the tertiary emotional state of grief).

With the addition of a motivator meta-management layer, Abbott is able to exhibit simple emergent *primary*, *secondary*, and *tertiary* emotional states. We can now ask "What else is required?"

5 References

Allen, S. (2000). Concern Processing in Autonomous Agents. Submitted PhD Thesis, The University of Birmingham.

Beaudoin, L. (1994). *Goal Processing in Autonomous Agents.* PhD Thesis, School of Computer Science, University of Birmingham.

Brooks, R. A. (1986). A Robust Layered Control System for a Mobile Robot. *IEEE Journal of Robotics and Automation,* Vol.RA-2, No.1, pages12-23.

Cañamero, D. (1997). Modeling Motivations and Emotions as a Basis for Intelligent Behavior. In *Proceedings of the First International Symposium on Autonomous Agents, AA'97,* Marina del Rey, CA, February5-8, The ACM Press.

Complin, C. (1997). The Evolutionary Engine and the Mind Machine: A Design-based Study of Adaptive Change. Ph.D. Thesis. The University of Birmingham.

Damasio, A. R. (1994, 96). *Descartes' Error: Emotion, Reason and the Human Brain*. London: Papermac. (first published1994, New York: G. P. Putman's Sons.)

Frijda, N. H. (1986). *The Emotions*. Cambridge: Cambridge University Press.

LeDoux, J. E. (1995). In Search of An Emotional System in the Brain: Leaping from Fear to Emotion and Consciousness. In Gazzaniga, M. S. (Ed.), *The Cognitive Neurosciences*. Cambridge, MA: The MIT Press, pages1049-1062.

LeDoux, J. E. (1996). The Emotional Brain: The Mysterious Underpinnings of Emotional Life. New York: Simon and Schuster.

Minsky, M. (1985, 87). *The Society of Mind*. London: William Heinemann Ltd. (first published1985, New York: Simon & Schuster.)

Sloman, A. (1999). Architectural Requirements for Human-like Agents Both Natural and Artificial. (What sorts of machines can love?). To appear in K. Dautenhahn (Ed.) Human Cognition And Social Agent Technology, John Benjamins Publishing.

Sloman, A. and Logan, B. S. (1998) Architectures and Tools for Human-Like Agents, In F. Ritter and R. M. Young (Eds.), *Proceedings of the2nd European Conference on Cognitive Modelling*. Nottingham: Nottingham University Press, pages58-65.

Sloman, A. and Poli R. (1996). SIM_AGENT: A toolkit for exploring agent designs. In *Proceeding IJCAI workshop on Agents Theories Architectures and Languages ATAL'95*, Springer-Verlag Lecture Notes in Computer Science.

Wright, I. P. (1997). *Emotional Agents.* PhD Thesis, School of Computer Science, University of Birmingham.