Utrecht University
Faculty of Science
Artificial Intelligence

Master's Thesis

# Further Integration Of Emotions Into An Agent's Decision Making Process

A Proposal For A Modular Emotion Integration Architecture For The Purpose Of Model Comparison And Flexibility

Author: Katinka Kalkman
Student Number: 3734102

First Supervisors:

prof. dr. John-Jules C. Meyer
dr. ir. Jan Broersen

Second Supervisor:
dr. Mehdi M. Dastani

22nd November 2017

# Contents

# 1 Introduction

Many new models for emotional agents have been proposed in the last few years, as emotions and their role in cognition have become more of a talking point in the scientific community. Additionally, there is more of a need for realistic seeming (i.e. believable) simulated beings in video games, both for entertainment games (so that players stay immersed in the story) and for serious games (so that the teaching or research value of the game is higher). These emotional agent models tend to be based on a particular theory from neuropsychology or a related field, and often include a list of emotions it considers the basic emotions, with causes and effects. These effects usually restrain themselves to maintaining desirable emotions and avoiding undesirable ones. However, there is no agreement on which emotions are the basic ones, and there are many signs that emotions have more influence on what a person does and thinks.

In this thesis, I want to to examine how to integrate emotion even further into an agent's decision process than most other emotion models, in order to create more believable agents. This emotion integration would not just be considering emotion as a numerical value to be kept in a certain range, or as a direct factor in the next action to take, but it will specifically be looking at how particular emotional states can influence judgement, attention, memory recall, and long-term effects. When I attempted to find theories on this, I came to the conclusion that theories on emotions are rather incomplete: they either consist of research on one or two emotions only, or they only focus on positive versus negative emotions, or even seem to contradict one another. This is not a critique of the psychological studies I considered: in fact, I will be referring to them later on. It was simply an indication for me that the field is still very young, and if I wanted to make a model based on these studies, I should focus on keeping the model easily adjustable and as generally applicable as possible.

This, as well as noticing that there is no consensus on what constitutes a basic emotion, and in fact reading papers that raised the possibility that emotions are culturally constructed interpretations of much simpler biological factors (Barrett, 2006 [4], Russel, 1991 [36]), inspired me to create a modular architecture, to be able to create many types of emotion models. This Modular Emotion Integration Architecture (MEIA) will make very few assumptions when it comes to emotions but can be implemented in a versatile way. It will provide more of a starting point, with a list of instructions to arrive at a model that will be suited for the specific implementation needed: it allows for very simple agents with just 3 emotions, or very complex agents that have emotional reactions tailored specifically to them, allowing for different personalities.

To summarize, the main questions I am trying to answer in this thesis are as follows:

- In what way does emotion influence the decision making progress, according to the current psychological theories and studies (beyond being a value to minimize or maximize)?
- How do I model these influences in an emotional computational agent for the purpose of creating more believable computational agents?

To answer the second question, some further questions arose, and the desire for a modular architecture arose from attempting to answer those questions:

- What is it that makes behaviour seem more believable, according to current literature?
- How do I model an emotional computational agent (so that I can integrate the emotion influences into it)?
  - How are computational agents in general modelled?
  - What general emotion theories are suited for emotional computational agent modelling?
  - How have others modelled emotional computational agents?
- How do I account for all the different views in psychology within one emotional computational agent model?

To answer my research questions, I will be referring to many papers about emotions, both on extensive emotion models and on small-scale research into possible effects, while claiming that there are no definitive answers yet on which, if any, theory is correct. Therefore, I will first expand on the purpose of MEIA, and the reasoning upon which it was based, to elucidate why I nevertheless refer to these (sometimes contradictory) papers. I will also give my reasoning for using non-scientific sources (like the dictionary) occasionally. After that, I will answer the rest of my research questions by outlining all possible modular parts of MEIA, and then giving some more concrete examples. I will conclude with examining how the research in this thesis could be expanded upon, which will include some details on the program I had originally started writing as part of implementing MEIA, and why I decided to provide a description of an implementation instead.

# 2 Striving For Realistic Seeming Emotional Agents

The primary purpose of this thesis is to examine how emotions influence the decision making process in humans (and potentially other beings), and to propose ways to model these influences in an agent, in order to create believable agents. To arrive at such a model, I needed to make several decisions. I will explain my reasoning here.

Firstly, I needed to find out how emotion in general is modelled: How are emotions in general defined, how are individual emotions defined and which emotions are required in a model? It turns out there is no consensus yet on the answer to these questions. For instance, one popular theory of emotion is appraisal theory. As described by a review of the various appraisal theories by Moors et al. (2013 [28]), appraisal theory considers an emotional episode to consist of an appraisal of the environment and the person-environment interaction, and the following changes in motivation, physiological responses, motor responses and subjective feeling. These changes can themselves also adjust the appraisal the person is doing. This appraisal is often automatic and efficient, based on simple rules, a learned associative (or schematic) mechanism, and, in some theories, on unlearned associations (usually evolved ones).

On the other hand, there is also the evolutionary view that specific emotions are the evolved reactions to certain stimuli, which people are born with, first raised by Darwin (1872 [11]). And there is also the constructionist view as offered by, among others, Barrett and Russel (1999 [37]): emotions are socially and culturally constructed, so initially there is only a positive or negative feeling with a certain intensity, and any label we give these feelings is taught by our environment. The fact that different languages do not seem to share the same concepts regarding specific emotions and emotions in general, seems to lend some truth to this (Russel, 1991 [36]). It should be noted that appraisal theory is not necessarily incompatible with the constructionist view, but the different appraisal theories do introduce appraisals that rely on the concept of emotion as found in English.

Rather than choose just one theory (appraisal or otherwise) and a specific list of individual emotions, I have decided to create an architecture that will accommodate multiple theories, and that will only provide the requirements for constructing an individual emotion. This way, a model designer can implement MEIA using the emotion model they find most desirable, and can define individual emotions for themselves. This can be useful for three different reasons.

The first is research: if an architecture is modular, then the same implementation can be re-used to compare one view on emotion to another, as only the emotion parts of the implementation need to be replaced. Another example of an architecture which allows for research in this way is FAtiMA (Fearnot AffecTIve Mind Architecture) Modular (2014 [13]), which allows different views on appraisal theory to be compared to one another.

The second reason is practicality. For entertainment or educative purposes, it might be enough to implement only those emotions that are actually relevant to the implementation. For a game, agents that are only interacted with in a limited capacity only need to be able to show those emotions that might occur in that capacity.

The third reason is creative freedom, and ties directly into the fact that in games, all that matters is what the player sees. If a game writer wants to write one character whose "anger" reaction is to become distant and dismissive, and another whose reaction is to shout and hit things, it would be far easier to have the two agents have two different implementations of "anger" than to attempt to model the range of factors which lead to such differences in humans.

Building on that, I wanted to incorporate the emotion's influence on the decision making process. While there has been research done on how the emotional states of humans influence their risk-taking (Lerner and Keltner, 2000 [21]), plan making (Laird, Wagener, Halal, and Szegda, 1982[20]) and creativity (Isen, 2001 [17]), among other things, this research tends to be limited to examining just a few emotions. This means that, for instance, there is no researched answer to how risk-taking is affected by each of the 22 emotions proposed by the appraisal model by Ortony, Clore and Collins (1990 [33]), known as the OCC model. Any OCC agent model equipped with an emotion effect on risk-taking would have to be based on speculation or incomplete information. Here, especially, a general architecture seems useful (and it was in fact one of the main motivators for me to propose a general architecture). If research shows emotion can have certain effects, then I can describe in this architecture how to implement such effects, without going into the specifics of how each individual emotion would cause (or not cause) that effect, and to what extent.

Aside from scientific theories and research of emotions, I will also be referring to cultural ideas regarding emotion, like folk theories, "common knowledge", and dictionary definitions, at least in regard to suggested causes and effects of emotions. The justification for doing this ties into the fact that MEIA was made with the purpose of creating believable agents in mind. Therefore, it's pertinent to discuss what makes a media portrayal seem realistic and therefore believable.

According to a model of narrative engagement by Busselle and Bilandzic (2008, [9]), people are engaged with a narrative work as long as there is nothing to disrupt the "flow" of the narrative. To paraphrase their explanation, engaging in a narrative is like building a train track out of the given information. This information can consist of the information and expectancies a person has learned in life and from other narrative works (schemata and stereotypes), as well as the information provided in the narrative. As long as the next part of the train track can be built, the person is engaged, as people are predisposed towards creating the belief required. As long as the flow is not interrupted by an inconsistency, the work is considered realistic.

Considering this, one might figure that the best way to assure an agent does not display behaviour that would contradict any schemata or stereotypes would be to mimic real human behaviour as closely as possible. However, Busselle and Bilandzic also note that expectancies formed by other narrative works can disrupt flow. Using Todorov (1977 [49]) as an example, in a murder mystery it is considered more real if the murderer turns out to be the least likely suspect, despite this not being the case in real life (unless police procedure is fundamentally flawed). In fact, what is considered to be realistic in real life might not always align with reality: judges have a tendency to not believe rape victims that did not attempt to fight or flee from their attacker (Ellison and Munro, 2009 [14]), despite freezing in reaction to threat being a normal human response (Schimdt et al., 2008 [39]). In this case, reality is deemed unrealistic.

When we combine this knowledge of perceptions of realism with the aforementioned paper by Russel [36], which indicated different cultures perceive and conceptualize emotions differently, a case can be made that when designing for a certain group of people (in this case those raised alongside or in Western culture), a person from that same group's judgement on what could be considered realistic in regard to emotional behaviour is relevant. Therefore, suggestions on different ways of implementing emotions using dictionary definitions, examples from popular culture and my own judgement are useful if the goal is to achieve realistic-seeming agents.

For the same reason, drawing upon research papers whose findings have not been validated by other research yet and the hypotheses made by their writers, is still meaningful when aiming for believable agents. The writers of these papers have spent considerable time studying emotions and the effects of it on others: their conceptualization of it, even if it turns out to be biased (by the scientific world's preference for English, for instance), would give a good idea of how emotions are conceptualized by the scientific community. And that would give a good idea of what they would find believable. In fact, when it comes to agents in general, this is an idea most architectures rely on. The agent architectures I will describe in the next chapter are not necessarily reflective of how humans have been scientifically observed to act, but were based on philosophical ideas of how agents could reason.

# 3 General Agent Architecture

## 3.1 Overview Of Agent Architectures

For different purposes, different types of autonomous agents and agent architectures are most appropriate. As such, it might not be accurate to say that there are some universal components every agent should have, but in order to discuss emotion integration, I will have to establish some core components of an overall agent architecture to apply the integration to. In an implementation, these might be more simplified or more complex than I will describe them here, but it is their purpose in the agent that is important.

As described in Michael Woolridge's book, "Intelligent agents" (2013 [51]), there are several types of agent architectures, which I will summarize here. The traditional type is a logic-based agent architecture, which relies on logical rules to arrive at an optimal plan to achieve some optimal situation. Logical agents have a disadvantage when it comes to time-restraints and dynamic environments and as such they are not very suitable for games, or simulations involving multiple, not entirely predictable agents. Therefore, I will not further discuss them.

Conversely, a reactive agent, rather than applying logic to all the information it has, reacts to specific events in the environment with specific, predefined behaviours. There are many reactive agent architectures, but what is important is that they rely on the agent being present in an environment (in other words, that it is embodied) and that intelligent behaviour is a result of the simpler behaviours prompted by the environment. As the focus of my thesis is ultimately on the influence of emotions on the decision making process, and a reactive agent usually has a decision making process that is relatively less complex, I will restrict myself to referencing how the components compare to those in the FLAME (Fuzzy Logic Adaptive Model of Emotions) architecture proposed by Seif El-Nasir et al. (2000[44]), which uses a reactive way to choose emotional behaviours.

An architecture that is a bit more complex and thus allows for more adjustments in regard to emotion integration is the Belief-Desire-Intention (BDI) architecture. To paraphrase Woolridge [51], a BDI architecture has seven core components: belief (the information about the current environment the agent considers true), a belief revision function (a function that adjusts the beliefs based on perceptual input and current beliefs), an option generation function (a function that generates the desires, or options the agent would want to take based on the environment), the set of currently available options, a filter function (a function that determines what the agent will intend to do), intentions (the desires the agent has committed itself to fulfilling), and lastly an action selection function (a function that decides which specific action to take based on the agent's intentions). For each component in my architecture, I'll indicate how it relates to these seven components.

Lastly, there are also layered agent architectures. Each layer in such an architecture is a subsystem using its own way to arrive at the next action to take. If two layers are horizontal in relation to one another, they both receive the same input, and deliver an action output to a function that decides which action to take. If the layers are vertical, one of the layers gets the input and one gives the output, while layers can also give input and output to one another. Layered agent architectures allow both a reactive and pro-active layer to calculate a solution, as well as any other processes that could be useful (like predicting what other agents might do based on the input received). In a way, layered architectures allow for the previous architectures to be implemented simultaneously. Dividing behaviour generation into quick, simple layers and more complex, slower layers is especially useful when time is of the essence, such as in robotics. Use of layered agent architectures for that purpose is researched by, among others, Rodney Brooks (1986 [7]). For each component in my architecture where I feel layers might impact implementation, I will examine how the architecture being layered might influence things.

I will refer to two types of components: Informational Components and Processes. The Informational Components keep track of knowledge or a state of the agent, while the Processes are in charge of adjusting the Informational Components based on external inputs, and affecting the external world in turn. These Processes are not performed in a particular order. Rather, they will activate one another when they have relevant information, with only the Process in charge of receiving direct external input not requiring any activation from other Processes. While optimization for parallel use is beyond the scope of this thesis, I assume the components are all programmed in such a way that they can all run in parallel. I will first describe the Informational Components, and then the processes in rough chronological order of activation. In chapter 5, I give concrete implementations of the Processes, as well as UML diagrams for these concrete implementations.

## 3.2 Informational Components

### 3.2.1 Long Term Memory

In order for an agent to be able to reason, it needs information to reason with. In the reactive FLAME [44] agent, it has information like which actions it can perform on which objects (if any), and for its learning component it has the learned consequences and probabilities of future events associated with all the events it has observed. It also has a user model, so it knows what actions to expect from a user, and it has learned what its user considers positive and negative behaviour. In a BDI agent, the memory is analogous to the belief component.

I've chosen to differentiate between a Long Term Memory and a Short Term Memory. I will detail the use of a Short Term Memory in the next section, but the important differences are that the amount of information in the Long Term Memory is theoretically infinite and can be assumed to remain the agent's belief unless there is some reason to update it. Conversely, information in the Short Term Memory will be discarded when enough time has passed. Of course, information in the Short Term Memory can be added to the Long Term Memory before it is discarded (otherwise learning anything new would be impossible).

Also, when it comes to memory recall (which will be quite relevant later on), Short Term Memory information is accessed before Long Term Memory information, which is useful because Short Term Memory information is mostly about what can be perceived, or has very recently been perceived by the agent, which is likely to be relevant to the current situation. To put it simply, Long Term Memory holds all knowledge about agents (including possibly the agent itself), objects and actions that might be relevant in the long term.

Beyond differentiating between Short and Long Term Memory, one might want to implement a Memory in layers. I will make further suggestions regarding this when describing the Memory Manager later on.

### 3.2.2 Short Term Memory

Considering the Short Term Memory different from the Long Term Memory is useful when it comes to learning, as it allows us to consider all information recently observed apart from other information. For instance, the FLAME [44] agent learns what actions the user is likely to do next based on their last action by observing the last three actions performed by the user: sequences of three events that occur more often are considered more likely to be performed.

Additionally, while it may be useful to retain the knowledge that, for instance, an obstacle is somewhere between the agent and its goal, once the obstacle has been passed, the knowledge can be discarded. For this reason, it is also useful to store Short Term Memory information separately from other information, and possibly in a different manner. The reason to consider them separately in regard to emotion integration has to do with methods of memory recall and retention being easier to explain if the memory types are considered separately.

### 3.2.3 Goals

Whether an agent is purely reactive or (also) proactive, there needs to be some component that drives them to act, which I will refer to as an agent's Goals. Note that while other components like the Motivational State I'll describe later could also be thought of as drives, for the purposes of abstraction, I consider them more like informing a drive. To give an example, hunger would be part of the Motivational State, but the drive to eat that results from it would be a Goal.

In a reactive agent, Goals are analogous to the functions that determine which behaviours to enact based on the current stimuli (in this case, the drive behind the behaviour is somewhat implicit). In a BDI agent, they are analogous to the intentions. In a layered agent architecture, there is some mediator in place between the different layers, deciding which implicit and explicit Goals are more important.

### 3.2.4 Plan

Once Goals have been determined, the action that needs to be taken next needs to be determined, in the form of a Plan. In a reactive agent, the Plan consists of the behaviours resultant from the stimuli. In a BDI agent, a Plan is the result of the action selection function. When implementing Plans in a layered fashion, the Plan layers would probably be coupled with their related Goal layers. Additionally, one might make multiple ways to arrive at a Plan, meaning multiple Plans might get generated for the same Goal. Depending on the implementation, a Plan could just be the action the agent decides upon as output, or it could be a sequence of actions, to be stored by the agent.

### 3.2.5 Motivational State

The Motivational State consists of a list of numerical values that correspond to states the "body" of the agent is in, corresponding to levels of things like hunger, pain and fatigue. In a reactive agent, this takes the form of stimuli leading to behaviours (in the FLAME [44] agent, the stimuli was the level of hunger, fatigue, etc.), while in BDI agents, the Motivational State informs the generation of its desires (namely, to keep the values in a certain range). Depending on the implementation, you could also think of hormonal levels or even available nutrition inside the body, similar to the neural influences on emotion described in the Cathexis model (Velsquez, 1997 [50]), which I will talk about in more detail later.

However, when it comes to the state of the body in relation to emotions, the line between what would be part of a Motivational State and what is an emotion gets blurry. For instance, some hormones like dopamine and endorphins are directly linked to experiencing pleasant emotions. They are automatically released during exercise (Boecker et al., 2008 [5]), or can theoretically even just be injected. Dopamine is also strongly related to receiving a reward, and even the amount of dopamine released depends on whether that reward was expected (Schultz, 2015 [40]). That seems to indicate dopamine release is also dependent on cognition processes, like emotions are often considered. To what extent emotion can be considered separate from hormone levels and other neurological factors is debatable, so for modelling purposes, consider the Motivational State the bodily reality, and the Emotional State the cognitive interpretation of it.

On the practical side, the difference between them is that the Motivational State values do not have influence beyond consequences for getting too high and/or low, and providing information about the agent's body. An argument can certainly be made that values in the Motivational State could also be considered information to be perceived and then stored in the Short Term Memory, and does not need to be its own, separate component. However, one might want to be able to have the agent not be aware of certain Motivational Values, and information in the Short Term Memory is information the agent is aware of. According to research by Izard (1993 [18]), such bodily states factor into emotion elicitation, so even if the agent is not aware, the process in charge of emotion elicitation should still be able to access the information.

### 3.2.6 Emotional State

The Emotional State consists of a list of all the possible Emotions an agent can have, as predefined in the particular model. It is part of the emotion integration, so it has no analogue to other agent architecture components beyond its similarity to the Motivational State. To use programming terms as an analogy, a particular

Emotion can be considered an instance of the Emotion class. An Emotion has a numerical value to denote its intensity, as well as several other properties I will expand on when discussing how emotion can influence decision making. In order for the Inferencer to read the Emotional State, some of these properties could be thresholds for the intensity, to denote whether they are in the high or low range, or not currently experienced (additional ranges can be implemented). Thresholds do not have to be numerical values, they can take the form of fuzzy sets, for instance.

### 3.2.7 Standards

Models like FAtiMA [13] and FLAME [44] implement cultural or moral values, which I would like to generalize to Standards. Simply put, Standards are all the rules that can lead to a value judgement of something, that do not arise from inferring whether that something was beneficial or detrimental to an agent's Goals. Standards are analogous to the function that informs the behaviour in reactive agents, and to the function that informs desires in BDI agents.

These Standards could include cultural values, which are, put in simplified terms, rules on how to behave, what to expect, and what to value in particular circumstances. But Standards can also include things like aesthetic, taste, or romantic preferences: rules about certain properties of objects, actions or other agents eliciting a certain reaction (emotional or otherwise informational, such as evaluations of agents). For instance, an intrinsic property of sugar is that it lessens hunger (an agent's Motivational Value), but a Standard of the agent might be that to it, sweet things taste bad.

## 3.3 Processes

### 3.3.1 Sensor

This process is can be constantly active, receiving external input (which also includes the state of an agent's body, the Motivational State). It will observe any changes to the previous input, and will be in charge of adjusting the Motivational State (due to its non-cognitive nature). A cycle of cognition is initiated by the Sensor sending out information. I've separated it from the next component because unlike the next component, the two processes performed by the Sensor are not influenced by any other component. Their BDI and reactive analogues, however, are the same: they are analogous to (part of) the belief revision function in a BDI agent, and to some of the processing of perceptions in the FLAME architecture.

Dividing the Sensor into different layers might allow an agent to emulate the different senses animals have (like seeing, hearing, tasting, etc.), and how these might differ in speed or accuracy. It would also make it easy to temporarily disable one of an agent's senses.

### 3.3.2 Sensory Filter

After the Sensor receives new information, it sends it to the Sensory Filter, who decides which information is relevant, and then sends it on to the right processes. Whether the Filter actually removes information that is deemed irrelevant, or simply sorts the information so that the Inferencer and Planner (processes that are beholden to time limits) will try using relevant information first, would depend on the implementation.

Something to note here is that filtering of information might also be done by other Processes, including the next one. However, the Sensory Filter only runs when prompted by the Sensor, and might perform filtering that is more simple and quick than other Processes (as it is supposed to simulate the way humans notice things in a different order). In order to make an explanation of an emotion influence later on smoother, I've decided to refer to the Sensory Filter as a separate component.

If the Sensor was divided into layers, it might make sense to divide the Sensory Filter into layers as well, to give each Sensor layer its own Filter layer. Additionally, if the Memory works with multiple sets/layers of information, having different layers for different ways of filtering might also be useful.

### 3.3.3 Inferencer

The Inferencer uses any new information it receives, as well as potentially all Informational Components, to infer new conclusions, presumably through some logic system. It is analogous to the belief revision function in a BDI agent, but without a clear analogue in a purely reactive agent. As the possible new information to infer might be infinite, exactly how long the Inferencer is allowed to run is determined by time pressure, which can for instance be calculated by looking at the Emotional or Motivational State or the current Plan. Basically, references to this Component, reference all means of deduction an agent has that is not directly tied to the workings of another Process.

The Inferencer especially could benefit from being split into layers: one layer could produce quick, shallow inferences, while others focus on a certain area and are allowed to go deeper.

### 3.3.4 Memory Manager

The Memory Manager is activated as new information is added to the Short Term Memory, so that it can see which information can be removed from the Short Term Memory, and which information can be moved over to Long Term Memory. The Inferencer might also learn new information that can be added to the Long Term Memory, which the Memory Manager has the methods for. Part of the belief revision function in a BDI agent. In the FLAME [44] agent, analogous to its learning component.

If the Memory is implemented in such a way that the stored information is divided in some way (for instance via layering), then having a different Memory Manager layer dedicated to each divide might be useful. For instance, if information on previous events for learning purposes and information on other agents is stored differently, it may be useful to have a dedicated layer for each information storage.

### 3.3.5 Goal Manager

The Goal Manager decides, based on new information, whether the current Goals and Plan being pursued should continue to be pursued. It considers whether any new Goals need to be added, whether any Goals are no longer feasible and whether the Goals the current Plan was made for are still the most important Goals to attend to at the moment. If it turns out no Plan can be made for the current Goals or a new Goal was chosen, the Goal Manager will prompt the creation of a new Plan. If a new Goal should be chosen to be fulfilled/maintained, the Goal Manager should be able to rank Goals somehow.

In a reactive agent, the Goal Manager is the function which decides which behaviour takes precedence, in the case that behaviours cannot be performed at the same time. In a BDI agent, the Goal Manager is the filter function which uses the set of currently available options.

### 3.3.6 Planner

When a new Plan needs to be made, it is the Planner which forms it. In a reactive agent, the Planner might not exist, if the reactive behaviour is entirely preprogrammed, but if the behaviour has some level of abstraction, then the process which translates the behaviour to actions would be the Planner. In BDI agents, the Planner is analogous with the action selection function. There are multiple planning architectures available, so an agent could have different Planner layers for the different architectures. Relevant to emotion causes later on, some implementations of a Planner can also calculate predictions regarding the future if they have time to do so (in the form of Plan outcomes), which can be stored in the Short Term Memory.

### 3.3.7 Emotion Monitor

This process is activated whenever information related to emotion triggers is sent to it, and is in charge of adjusting the Emotional State. Depending on the complexity of the emotion triggers you decide on, this process can be activated by almost every other process. The Emotion Monitor adjusts the intensities for each emotion. It also determines which emotions are currently "dominant" (basically, which emotions are currently being experienced by the agent), if knowing that is required in the implementation. Multiple emotions can be dominant at the same time in an implementation, and the easiest way to determine that would be to simply have a threshold intensity at which they become active. Such details are highly dependent on which theory of emotion you're implementing. In regard to analogous components in BDI and reactive agents, it has some elements of BDI's belief revision function and the process in charge of calculating emotion intensity in the FLAME [44] agent. However, it is more than that, as a "new" emotion integration-related component. In regard to layering, each emotion could get its own layer, as was done in the Cathexis [50] model, which I will describe in more detail later. Naming this process a Monitor, as opposed to a Manager, is a nod to emotions potentially being how people experience and register far more basic processes in different contexts, as I'll detail later. Under that framework, emotions are not internally adjusted, but inferred from context. Outside of that framework, this Process can also be called an Emotion Manager.

### 3.3.8 Output Generator

This is the process that is ultimately in charge of sending information of the agent to the external world. It communicates the action that should be performed by the agent to whatever external process is in charge of that (like a graphics engine, for instance), but it's also in charge of deciding which types of animation, sounds, emotional signalling, and so on, the agent is going to use. This is where the Emotional State is used to make sure the player is shown the emotions of the agent through body posture, facial expression and tone of voice, if different ones are available for different emotions and surrounding circumstance. Reactive and BDI agents usually do not differentiate between the action decided on by the Planner and the actual action the agent will take.

The reason I refer to it as a Generator and found it useful to refer to as a separate component, is due to the following consideration. If an agent has anger as its highest intensity emotion, the Output Generator could just send the order to use the "Angry" animations. But what if we want our agent to (successfully or unsuccessfully) hide that they are angry, maybe because one of their Standards is that showing anger is unseemly? In that case, the Output Generator needs to decide which animation would be best. Of course, you could go even further and consider emotion signalling an action, and make it part of the planning process itself. In this case, the Output Generator would still be in charge of regulating how successful the agent is in changing their emotion signalling, which would require some base value for agent acting ability combined with adjustments to this ability based on emotion intensity (at the very least).

Integrating emotion signalling into the planning process could be done by simply considering performing a certain action with undefined emotion signalling different from doing the action with a specific emotion signalling (although that would multiply your agent's possible individual actions by the number of possible signals), or you would need to use planning that can deal with parallel actions and possibly multiple goals being fulfilled at the same time, such as the proposed approach by Chen et. al (2013 [10]).

# 4 Proposed General Emotion Architecture

## 4.1 Overview Of Emotion Models

When talking about emotion, it's useful to define what is meant by emotion. However, due to the nature of my architecture, I will define it as little as possible, to allow the broadest variety of interpretations to be implemented. Therefore, my definition is based on what an emotion's function is in the architecture: an emotion is a cognitively arrived at state of the agent, potentially both influenced by and able to influence any other process within the agent. An agent usually has more than one potential emotion, and its total state can consist of multiple emotions at a time, although one or more may be considered dominant. To give a somewhat clearer look at emotion, I will describe two other models of emotional agents, and then describe how their model is reflected in my emotion model architecture. These other models would be able to fit into my architecture. I will also describe the constructionist model of emotion in more detail, and how it influenced MEIA.

The Cathexis model (Velsquez, 1997 [50]) uses a theory of basic emotions: Anger, Fear, Distress, Happiness, Disgust and Surprise. These emotions each get a specialist process that keeps track of their intensities, activated by sensors detecting relevant stimuli to the emotion. The specialists that detect an intensity above zero send information to the other specialists and a Behaviour System. It considers emotions that are not among the basic emotions as blends of the basic emotions. It does not further define which stimuli would cause which emotions, but it does define which sensors can send stimuli, influenced by, among others, Izard's systems of emotion activation (1993 [18]): neural (neurotransmitters, brain temperature, influences of hormones, sleep, diet, etc.), sensorimotor (facial expression, posture, muscle action potential, etc.), motivational (drives like hunger, emotions, pain, odours), and cognitive (any cognitive process).

For MEIA, I do not define any basic emotions, but the sensors mentioned can be used as an emotion's cause. Neural and motivational sensors would be part of the Motivational State, sensorimotor senses would be decided by the Output

Generator, and cognitive sensors would be part of the Inferencer. The Emotion Monitor handles all emotions, but there is no reason the Emotion Monitor could not be realized as multiple processes, one for each emotion. Unlike the Cathexis model, I don't directly define blends, as that would require assumptions of how emotions function: rather, it depends on how the interpretation of the Emotional State is defined in the model, which in turn determines which emotion effects are chosen. I also allow for defining sub-emotions (detailed later).

The Cathexis model also differentiates between moods and emotions: moods persist over time, and are low in intensity (low arousal), while emotions are high in intensity and are quick to fade. Emotions can inhibit other emotions, but a low intensity emotion (mood) does not inhibit other emotions and multiple emotions can keep existing in the background. Emotions in Cathexis also decay slowly until they become inactive if there is nothing to increase their intensity. MEIA also has the option of emotion decay, and allows an Emotion's intensity calculations to involve properties of other Emotions. The Emotion Monitor can also determine the dominant emotion if needed. The Cathexis model uses a behaviour system rather than step-by-step planning system, but its behaviour system consists of an expressive component (analogous to the Output Generator), an experiential component (which is reflected in MEIA in the ways emotion can influence the Inferencer and Goal Manager) and of course, behaviour selection (analogous to the Goal Manager and Planner).

Another model I've been inspired by, based on appraisal theory, would be EMA (EMotion and Adaption) (Marsella and Gratch, 2009 [24]). Appraisal theory, as applied in Marsella and Gratch's paper, says that emotions result from (or in some cases, are) inferences (or appraisals) made by a person of the events that are happening (including internal events). Following ideas by Smith and Lazarus (1991 [45]), EMA appraises and re-appraises its emotions throughout the cognitive process. It also considers this appraisal to be fairly shallow and quick. In other words, they arise from cognition, but the emotion processes themselves don't perform deep inferences. The appraisals are made based on a couple of inferences: the impact of an event on the agent (how positive or negative it is), future implications, blame and responsibility, power and coping potential (how much control the agent has to change events, others or themselves), and coping strategies. The last inference is about how emotion is being coped with: which coping strategies have been performed?

Aside from that last one, these inferences are all suggested as causes in MEIA in the this chapter. The coping strategies are implicitly included, in the sense that if a cause for an emotion no longer holds, the Emotion Monitor will adjust the Emotional State. The coping strategies suggested by EMA are about trying to retain or attain information (which can be implemented by the Memory Manager when deciding which information to add or keep in the Long Term Memory and the Sensory Filter when deciding which information is most relevant respectively), which beliefs and judgements to choose in order to have a world-view that is least distressing (which can be implemented in the Inferencer), changing utility attribution of a plan or action based on perceived emotional consequence (which can be done by the Goal Manager), and which Plan or action to perform next. Implementing such coping strategies depends on which actions an agent is capable of performing, and is therefore somewhat more implementation specific. However, they are very interesting to consider when defining actions and causes of emotions. EMA does not give an exhaustive list of emotions, but the list they do give, defines each emotion by their unique, singular trigger. MEIA allows a list of triggers, but the triggers provided in EMA can all be implemented.

I could very well have chosen to extend one of these models, and I'm sure it will be quite clear while reading the rest of this thesis that I've taken some inspiration from them, and appraisal theory in general. However, what ultimately made me decide against doing so was reading papers by Lisa F. Barrett and James A. Russel, which introduced me to a constructionist view on emotions. In Russel's paper, "Culture and the Categorization of Emotions" (1991, [36]), he talks about how other languages, and therefore other cultures, do not always share the emotion-related words that are considered basic in English. For instance, Luganda has one word for what we would consider both anger and sadness; at the other end, Utku has two words for what we consider different types of fear, namely fear of social injury and physical injury, and no word that groups them. How people evaluate and talk about emotions seems to depend on culture, Russel claims.

Additionally, words that we would translate as "emotion" sometimes include feelings or even events we would not consider emotion: for instance, the most direct translation of emotion in Japanese also encompasses "being lucky". So, how English speakers conceptualize emotion might not be applicable to how people with other mother tongues conceptualize it. This suggests that emotion, at least partially, is socially constructed. Further evidence for this was that Barrett, in "Solving the Emotion Paradox: Categorization and the Experience of Emotion" (2006 [4]), found that there were no clear or consistent criteria to see whether someone is experiencing a particular emotion or not, and that not everyone is "skilled" at determining which emotion they are feeling beyond good and bad. Barrett offers the theory that emotions are simply labels we give to more simple feelings, depending on the broader context and, subsequently, the behaviour they result in. Just as we label colour, despite light waves being a continuous spectrum, we label some sequence of actions by others as a discrete behaviour, despite the fact that people are continuously doing something (not to mention the fact that behaviour labels can also be different for the same sequence of actions in different contexts). Therefore, any ability to recognize emotions in oneself and others is learned, just as one learns what blue and green are. If we take this view of emotion, then any proposed model of emotions that defines emotions beforehand enforces a particular construction of emotions. I would prefer not to do so.

Whether a constructionist view is held of emotions or a view that certain base emotions are innate, emotions are differentiated by how they are triggered, and how they cause one to act. Under the constructionist view, this is due to the sociocultural labelling being based on context, and in innate views this is due to the brain reacting to certain stimuli by activating the relevant emotional circuits (simply put). So, in order to integrate emotions into an agent's decision making process, we have to decide which emotions to implement, and how we want them to be caused and what we want their effects to be. To arrive at this list, there are some questions to ask. How detailed are your characters? Are they constant companions to the player character, and do you want them to be able to express a complex emotion like schadenfreude? Or is the character a pet the player character might come across sometimes, and will just joy, anger and fear suffice?

For the first type of character, they may have many emotions with shorter (but more specific) lists of triggers, that require a lot of information (for instance, schadenfreude requires the character who feels it to have an opinion towards a character they observe and judge an event as unfortunate for that other character). On the other hand, the pet's emotions do not require much information, only knowing in general whether a character in the immediate vicinity is having a positive or negative effect on it (and in the case of anger vs. fear, maybe just whether the other character is more or less powerful than it). And on the effects hand, we might want the constant companion to show a hidden sadness when presented with an object that reminds the companion of a late parent, and to make quick, impulsive decisions when scared. But for the pet, it might be sufficient if it just growls when angry and wags its tail when happy. I will first describe the meta-structure of emotions (emotions vs. moods, sub-emotions, emotion decay, etc.), and then I'll provide the causes and effects I have come across.

## 4.2 Meta-structure Of Emotional State

While each individual emotion can be defined by its causes and effects, the Emotional State is not necessarily only defined as the set of these emotions and their respective intensities. A decision that needs to be made is how the Emotional State should be read by the other Processes. It could just read the list of emotions with their intensities, and decide what to do based on the specific combinations of values. Cathexis, being a reactive agent model, does something like this, in the sense that the intensity values are used in calculating the preferableness of the individual behaviours. Also, as mentioned, Cathexis differentiates between "moods" and "emotions". The Cathexis-emotions are those emotions that are high in intensity, while the Cathexis-moods are low in intensity. So the Emotional State has some extra information on classifying the emotions as "emotion", part of the "mood" and inactive (no intensity). In an agent that uses logical reasoning, it might help to have a smaller amount of different possible Emotional States. For instance, one could assign a valuation to each emotion's intensity and pass that on, like "high" and "low". That would make the total amount of different Emotional States two to the number of emotions. In EMA, the Emotional State as interpreted by the process in charge of deciding how to cope with the emotion, is the emotion with the highest intensity, meaning the Emotional States possible are the number of emotions.

Another thing I mentioned before is that emotions can have some relationship with other emotions. For instance, "joy" and "gratitude" seem closely related, and in fact if we define the cause of gratitude as "a positive event happens for the agent due to another agent" and joy as "a positive event happens for the agent", then an agent feeling gratitude will always feel joy. There are several ways to deal with this. If there is no need for joy and gratitude to be linked somehow (and this link would mostly be relevant when it comes to deciding on which effects of emotions on decisions to apply, so less complex agents might not need it), you could simply make sure joy only triggers when "a positive event happens for the agent, not due to another agent". For more complex agents, you could define gratitude as a sub-Emotion of joy. A sub-Emotion is an emotion that has triggers that are the same as its parent Emotion, but with added conditions. In that case, when it comes to the agent's expressed emotion, it should always prioritize expressing the sub-Emotion over the parent Emotion. But when making other decisions, it might only look at the parent.

When implementing a sub-Emotion, the Emotion Monitor should first consider the triggering of the sub-Emotion, and then adjust its parent with it. The intensity of a parent Emotion is at least the total intensity of its children, as any increase in the child's intensity would also increase the parent by the same amount. When it comes to causes for an Emotion stored in the Short Term Memory, all the causes of its children should be the assumed causes of the parent Emotion, for consistency.

Aside from matters of interpretation by other Processes, there is also the decision on how to implement emotion decay over time (if at all). In EMA, the current intensity of an emotion is a function of all intensity changes of events that are still considered relevant (to paraphrase a bit). Basically, the decay is implicit in that as the events that caused the emotion intensity to increase are dismissed, the intensity becomes lower. Cathexis, on the other hand, applies some decay function to the stored intensity value that makes it decrease more as the intensity is higher. Some other options are to apply no decay at all, which might work as long as the interpretation of the Emotional State relies on intensities only in relation to one another, to decrease the intensity by a constant value until it reaches zero, or to apply some function that tries to normalize the intensity to some default value (which is not necessarily zero). That last one would allow for an agent whose is happy or sad by default (for instance), as the default value would not have to be the same for each emotion. However, the decay of a parent emotion should not be greater than that of a child emotion, or the child emotion might be more intense than its parent, which would not make sense.

## 4.3 Causes of Emotions

One of the most common models for emotion generation (often used in appraisal implementations), the OCC model by the eponymous Ortony, Clore, and Colllins (1990 [33]), defines 22 emotions based on a valenced reaction of the agent to either: an action of an agent, the consequence of an event or an aspect of an object. This valence can be positive or negative, and is further defined by Fortune-of-others, Prospect, Well-Being, Attribution, the compound of those last two, and Attraction. For instance, if an action is considered positive by some standards and attributed to the self, it generates pride. In a way, MEIA takes the opposite approach to arriving at a list of emotions. Rather than figuring out how emotions are caused, and labelling all the combinations of causes, it asks the modeller to decide on emotions they want their agent to have, and for each emotion, what should be the combination of factors to generate it.

Of course, this has the practical downside that if a way an emotion might be caused is not thought of, it won't occur in the implementation. Any list I give of possible causes will always be expandable and therefore seem incomplete, as it's not necessarily based on a logical structure like the OCC model. On the other hand, the list I will provide will include all factors used in the OCC model, as well as the appraisal components described by Smith and Lazarus (1991 [45]), Roseman (1996 [35]), Scherer (2001 [38]), and Izard (1993 [18])(as used in Cathexis). Additionally, it includes considerations I thought of when considering how to implement some more complex emotions like schadenfreude and home-sickness, using their dictionary definitions. So considering current emotion elicitation models can all be implemented with MEIA, the list of causes that can be included in MEIA can be as long as, or longer than, any other list provided by current emotion elicitation models.

The causes an Emotion may have take the form of a list of trigger conditions belonging to that emotion. One Emotion may trigger under multiple conditions. For instance, shame might trigger from the agent itself causing a goal to no longer being achievable, or from breaking a cultural rule. The following list consists of the factors that would form the conditions' antecedents (related factors might be under one header). The Processes of the agent that would perceive these emotion condition factors send this information to the Emotion Monitor, so that it can check if any emotions are triggered, and change the Emotion State accordingly. Some of these condition factors not only help trigger an emotion, but can also influence how much that emotion's intensity was impacted, if the condition factor is a numerical value or can be ranked. Some might even impact an emotion's intensity without being part of its trigger condition. Where applicable, I will also refer to such influences on emotion intensity.

### 4.3.1 Attribution

An important condition factor when defining emotions like "pride" and "shame", Attribution has been raised as an emotion cause by many models. For instance, in the OCC model [33], which agent performed an action determines to what extent the emotions like pride and anger are generated. Smith and Lazarus [45] refer to it as accountability, one of the six appraisal components they describe. Scherer [38] also refers to checking causal attribution for emotion appraisal. Roseman [35] differentiates not only between events caused by the agent itself and those caused by other agents, but also those events that are not clearly attributed to any agent, or where the attribution is disregarded in favour of the event itself.

In that light, I will refer to Attribution to Self, Other, and World. Self is the agent itself, Other is another agent considered to have agency (so a human or a dog might trigger Other, but a swarm of bees might not), and World is attributed to any event that was caused by the environment (although as previously mentioned, what the agent considers environment or not might depend on implementation). An event is attributed to at least one of these, possibly more.

For instance, an event can be attributed to both Self and multiple Others in case of a group effort. A stranger bumping into the agent and making it spill its coffee might be attributed to both Other and World, because while the stranger was a person and therefore responsible, an agent might also attribute the event to bad luck/the world being against it.

An example for how conjunction with another emotion condition factor might lead to different emotions: if the agent evaluates another agent as "liked", then an event caused by them which is positive might inspire the emotion "gratitude" (the general example given regarding attribution in the OCC model), while a "disliked" Other might inspire negative feelings (due to the implication the agent needed help). This last example is derived from common knowledge, as can be observed in the video-game "The World Ends With You" [47], where the protagonist saves his enemies and says he's doing so to annoy them (and one of them is indeed disturbed by it). Attribution can be observed by the agent's Inferencer Component.

### 4.3.2 Motivational Values

As described when I defined the Motivational State, the Motivational values describe some bodily reality of the agent, like hunger and pain levels, hormone levels, temperature, and so on. The OCC model [33] only indirectly deals with some of these values, namely the ones an agent is aware of (like hunger), as a factor in determining future consequences. Consider the appraisal "if I don't lower my hunger level soon, I will starve", as a negative consequence prospect for the self.

The model of Smith and Lazarus [45] also only deal with such values in the sense that they impact Goals. Roseman's model [35] also refers to appraisal regarding punishment minimizing motives. "I want to feel as little pain as possible" would be such a motive, although I will consider that a Goal for consistency.

Scherer's model [38] more directly refers to some Motivational values, as it checks for "intrinsic pleasantness", referring to those things intrinsically positive or negative. This has overlap with Motivational values like pain, which is referred to as an intrinsically negative thing.

When it comes to emotion intensity, the quantity a Motivational value has that is used in the trigger condition for an emotion could have a direct effect on the intensity change of that emotion.

Izard's model [18] also describes pain as a non-cognitive emotion activator, and additionally describes the evidence of hormones, brain temperature and neurotransmitters influencing the emotional state. In the latter case, the levels of these Motivational Values might impact intensity changes of emotions that they did not trigger. For instance, Izard refers to studies linking neurotransmitter levels with depression. In the Diagnostic and statistical manual of mental disorders (5th ed., 2013 [3]), major depressive disorder has as one of its criteria "markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day".

To model this, either accurately for research purposes or in a simplified manner for creative purposes, one might want certain neurotransmitter levels to decrease the intensity of certain emotions. Changes in Motivational values are made by the Sensors of the agent, so they are best equipped to send that information to the Emotion Monitor.

### 4.3.3 Likelihood

When it comes to evaluating an event, one factor can be how likely it was that this event happened. Roseman's model [35] differentiates "unlikely" events that have occurred from "certain" events (events that already happened) and "uncertain" events (events that will happen in the future or might have happened) as a factor in emotion elicitation. These "unlikely" events are considered to cause the emotion "surprise" in Roseman's model. Scherer's model [38] also refers to emotion appraisal checks for "novelty" (which includes whether something was predictable or not) and "discrepancy from expectation" (which requires some expected likelihood). The likelihood of something happening might also contribute to how positive something is judged to be, which I will detail in the next factor.

The Likelihood can be observed by the Inference Component, using the information in the Short Term Memory (like the calculated possible outcomes the Planner produced) and the Long Term Memory (the learned expectations based on previous observed events, and any preprogrammed expectations). One could also consider Standards determining how likely an agent judges an event, if the agent's cultural values or learned schemata imply that certain people do not perform certain actions. As an example, consider the cultural value that decent people do not hit children. Say the agent considers person A a decent person (or assumes most people are decent), and then sees person A hit a child. They would judge the event as being unlikely.

### 4.3.4 Predicted Alternative Events

If an event happens, but the agent has information about what also could have happened, that might elicit certain emotions. As described in Loewenstein and Lerner's paper "The Role of Affect in Decision-Making" (2003 [23]), people tend to feel comparatively bad when they've won a silver medal, as they feel they could have gotten gold, but comparatively good if they've won bronze, as they could have won nothing at all. Simplified, the likelier an agent judges it might have won, the more negative it feels to lose. The Predicted Alternative Events factor and its related Likelihood factor combined would lead to such emotion triggers.

The information regarding the Alternative Events would be stored in the Short Term Memory. It could have been predicted by the Planner when it was determining possible outcomes to a Plan, or it could have been received via a Sensor (the agent was told about it), or deduced by the Inferencer. Either way, the Inferencer would be the one to note which line of events has occurred (and which lines, by implication, are no longer possible).

#### 4.3.5 Effect on Goals

After an event occurs, the Goal Manager may find that the event has had a positive or negative effect on one or more of the agent's Goals. Fulfilling a Goal can be easier, harder, the same, successful, or impossible. An event could have a different effect on different Goals. In an implementation, it's necessary to abstract the total effect of an event over all Goals in some way, rather than consider each quantity of Goals affected as a separate case. For instance, you may want one Goal becoming easier and another becoming harder not to be considered a different trigger from two Goals becoming easier and two harder. On the other hand, you may want one Goal becoming slightly harder but another much easier to be considered different from one Goal becoming much harder and another slightly easier. So the effect could be considered as such: total improvement/worsening, based on importance of Goals and severity of easier/harder change, and whether there is a conflict or not. The way to calculate such things depend on how complex of a trigger/ emotion intensity calculation you want to have.

The OCC model [33] considers the consequences to the agent itself during emotion generation, which includes an effect on Goals. Smith and Lazarus' model [45] proposes "motivational congruence" as an appraisal concern, which refers to the extent to which an event is consistent or inconsistent with an agent's Goals (and other desires). It also refers to "motivational relevance", which refers to how relevant an event was to Goals and other desires. "Motivational relevance" is implicitly modelled for Goals considering an Effect on Goals factor would not cause an emotion to trigger unless the event impacted a Goal (i.e., be relevant to the

Goal). Likewise, Scherer's model [38] includes checks for "goal relevance" and "goal conduciveness". Roseman's model [35] considers all emotion elicitation within the frame of whether an event was consistent with the agent's motives (Goals) or not. It also divides the emotion elicitation further by considering whether the Goal was reward maximizing or minimizing. To model that, combine the Effect on Goals factor with the following Expected Consequence factor. While this may first be observed by the Inference Component, it is guaranteed to be observed by the Goal Manager, so it is better to leave the prompting of the Emotion Monitor to the Goal Manager.

#### 4.3.6 Expected Consequence

For an event that just happened (including a decision to perform a certain action), the Expected (but not yet confirmed) Consequence of that event might elicit some emotion. For instance, the OCC model [33] uses prospect-based as a factor in its emotion generation. Smith and Lazarus' model [45] has something similar but less broad, with "future expectancy" as an appraisal component, which refers to any possibility for future change that could impact how the current event impacts the agent's motivations. The aforementioned "uncertain" events of Roseman's model [35] also fall under Expected Consequences. Scherer's model [38] has a check for outcome probability, which is the same thing. An example of an emotion that might rely on Expected Consequence is "hope". Something to note here is that when determining an Expected Consequence, this includes the emotions that might arise in the future. So all the factors in this list might become relevant to how this factor elicits an emotion, in the sense that they are used to predict a future emotion. The Expected Consequence of an event would be thought of by the Inferencer, or by the Planner when it creates a Plan.

#### 4.3.7 Sense of Control

Related to the Expected Consequences, to what extent the agent feels that they might affect situations to be more positive might influence which emotion is caused. For instance, consider the difference between seeing someone get hit by a flying newspaper due to the wind, and seeing them get bitten by your own dog. In the first case, you had no control of the event, so you might feel sad. In the second case, you might feel guilty, because you could have kept the dog on a leash. In all examples I've found, Sense of Control impacts differentiation between negative events only. This does not mean it can't be used for positive emotions, of course.

This factor is used in Smith and Lazarus' [45] emotion appraisal components, specifically "problem focused-coping potential" and "emotion focused-coping potential". The first indicates how the agent judges their ability to affect the world to be more in line with the agent's goals and desires, and the second how the agent judges their ability to adjust their beliefs, goals and desires in order to remain emotionally stable (having no negative emotions with high intensity for long periods of time).

Roseman's model [35] also refers to low and high control potential of regarding an event to determine which emotion is felt. However, unlike in Smith and Lazarus' model, Roseman contends that what differentiates the emotions is not coping potential, but whether the agent feels they can "do something about [the negative circumstance]". In other words, only a specific part of coping potential is relevant. Scherer's model [38] differentiates between checking "control" (whether the event can be changed by anyone), "power" (ability of the agent) and "adjustment" (the emotion focused coping potential).

Whichever model for Sense of Control-related condition factors you want to implement, how to implement them depends on other implementation factors. If you want to apply the idea that emotion appraisal is quick and shallow, like in EMA [24], then these factors should be found among the other Processes. Quick and shallow implies, after all, that no additional computational resources should be devoted to figuring out which emotion is felt.

I will give some examples. When the Planner is creating a Plan, if it has time to calculate the success of a plan, if the success is low, that says something about an agent's ability. Before that, the Goal Manager may make assessments on which Goals can be achieved, which would signal an agent's problem-focused coping potential. When the Sensor receives or the Inferencer infers new information regarding what an agent could have done differently, what its options are now, or how current events might impact the agent in the future, that too can signal a Sense of Control factor.

Depending on how actions or behaviours are implemented, taking time to let the Inferencer derive more information about the agent itself and its relationship with the world in general, would give the agent more information to use (more "awareness") about itself, which might make an agent more believable.

### 4.3.8 Memory Association

If a memory of a similar event or observed object is accessed, the agent "reliving" the memory might induce the related emotion. Psychological research induces certain emotions into test subjects by making them think about a related memory, as reviewed in Brewer, Doughtie, and Lubin (1980 [6]). If, in your model, some memories related to observed objects or events are always accessed, as part of the identification process (or to simulate that an identification process has happened), some objects or events might always influence an agent (which might make the agent feel more consistent). For instance, a gun might have an associated emotion of "fear" with a high intensity, and might be a factor that could trigger "fear" even if the context does not cause the agent to predict the gun is about to be used to shoot people.

A memory being accessed would be observed by the Sensory Filter (in the case of an object), or the Inferencer or Planner as they retrieve information from the memory for other purposes. Additionally, reliving a memory could just be an action an agent is capable of taking.

### 4.3.9 Standards Rule

As earlier defined, Standards are all rules that can lead to a value judgement of something. For Standards about taste, smell, aesthetic preference, sense of humour and the like, the value judgement is about whether something is experienced as positive or negative, so if any object or event follows such a Standards Rule, that might directly cause an emotion. This direct emotion causing would have to be sensed by the Inferencer, as it's applying the Standards Rules while trying to infer new information. This use of Standards falls under the "desire" mentioned of other models when I discussed Effect on Goals. The OCC model [33] uses "aspects of an object" as a determinant in emotion generation, which would also fall under this. Scherer's model [38] uses an "internal standards" check when evaluating an event, which refers to whether an event or action aligns with personal morals and the like, which would also fall under this factor. Standards can also be used when determining how likely an agent judges an event to be, as mentioned before.

### 4.3.10 Confirmation or Disconfirmation of Current Emotion Cause

Some emotions, like relief and disappointment, are often defined by the emotion that preceded them. For instance, relief can be seen as a reaction to previously feeling fear, and then finding the cause of the fear is removed. The OCC model [33] uses this in the prospect-based elicitation. First an emotion like hope or fear is elicited by Expected Consequences, and then a different emotion is elicited if that consequence becomes confirmed or disconfirmed. Similarly, relief is defined as "removal or lightening of something oppressive, painful, or distressing" in the Merriam-Webster dictionary (MWD) [26]. Checking when a distress-related emotion's cause no longer holds is an easy way to model that.

The Emotion Monitor is in charge of making sure the causes listed for the Emotions still apply, so it can immediately use this condition factor to check if the Emotional State needs adjusting. There are other ways to model relief that do not include previous emotions, like how Roseman's model deals with relief. There, it is elicited when circumstance regarding a Goal that was punishment minimizing is positive, which is covered under Effects On Goals.

### 4.3.11 Emotion Intensities

Emotion elicitation can also be caused or influenced by the intensity of other Emotions. This possibility is raised in Izard's model [18], where a reference is made to theories that excessive sadness may become anger (due to how neurons are fired in both emotional state). However, Emotion Intensity combined with other factors could also be a cause for another Emotion.

For instance, an agent might have Standards about how a good or normal person would react emotionally to an event. From personal experience, for instance, I have felt guilt or anxiety for not being as sad as an event was "supposed" to make me feel, which I believe is not uncommon. After all, humans judge other people's moral character by their perceived emotional reaction (Stearns and Parrott, 2012 [46]), so not having certain emotional reactions might say something about the agent's moral character.

For how such an inconsistency between behaviour and self-image might cause distress, it's also interesting to look at cognitive dissonance, which I will detail when discussing the Sense of Normalcy and Order factor. Additionally, experiencing one emotion might "dampen" another emotion, like fear decreasing the intensity of joy, as implemented in the FLAME model [44]. This would be observed by the Emotion Monitor itself.

#### 4.3.12 (Changes in) Evaluation

In the Long Term Memory, the agent may have stored certain Standards-derived evaluations of other agents and itself, and possibly the world. They might influence emotions in the way detailed under Attribution, or they could function in the same way as Memory Association does. The OCC model [33] uses the similar "fortune-of-others" as a determinant in emotion generation. In Roseman's model [35], different negative emotions are elicited if performing an action is intrinsic to an agent's or object's character than when the action was seen as a behaviour by that agent or object. When these evaluations change, that might elicit certain emotions, like surprise or disappointment.

The Change in Evaluation factor is partly a subset of Expected Consequence factors in a way: an evaluation of an agent indicates what can be expected of them in the future, and if the evaluation changes, so does the expectancy. It may also interfere with a Sense of Normalcy and Order, which I will detail later. Scherer's model [38] partly has this in the form of a "internal standards" check, but that only relates to the evaluation of the self.

The Inferencer is the one that finds out there is a change, and the Memory Manager is the one to apply it, so one of them is best equipped to notify the Emotion Monitor. The current evaluation as part of a more complex trigger can be read by the Emotion Monitor from the Long Term Memory.

#### 4.3.13 (Changes in) Knowledge of Another Agent

Aside from how the agent evaluates another agent, an agent also has knowledge about another agent. Emotions that may be about how others are feeling, like pity ("sympathetic sorrow for one suffering, distressed, or unhappy", MWD[25]) or schadenfreude ("enjoyment obtained from the troubles of others", MWD [27], might be caused by new information about another agent's Motivational or Emotional State. Information on what another agent might find desirable is one of the determinants in the OCC model's emotion generation. The impact a change in another agent's evaluation of the agent might have is noted by Scherer [38], as an "external standards" check. As explained by Scherer, no longer conforming to others is often considered quite negative for humans. Like the previous factor, the Memory Manager is best equipped to notify the Emotion Monitor.

#### 4.3.14 Emotion Signalling by Other Agents

Emotion also has a strong social function, as I will detail later, which requires the agent to be able to see (and send) information on an agent's Emotional State. While the previous factor is about knowledge to be reasoned about, the Emotion Signalling by Other Agents factor is more about the immediate reaction to seeing others express emotion, namely, to copy them. This "copying" of the emotions signalled by others is referred to as mood contagion. Research by Neumann and Strack [30] found that mood contagion was not reliant on Goals relating to emotion, and that the participants were not actually aware that their mood had changed (so it was not caused by assumptions or suggestions either). They also found that increased cognitive load did not diminish the effect, suggesting mood contagion can occur without cognitive resources.

In practice, this effect can be modelled by slightly increasing the intensity of emotions an agent can see the signals of. The reason I say "slightly" here is that the participants of the experiment were not aware of their mood change: a sudden dramatic increase of intensity is therefore not likely to have occurred. To align with the non-cognitive aspect of mood contagion, the Emotion Signal would be observed by the Sensory Filter.

#### 4.3.15 Body Signals

According to Izard's model of emotion activation [18], outwardly displaying signs of an emotion via body posture and facial expression can elicit the emotion itself. Specifically, according to the facial feedback hypothesis, supported by research by Adelmann and Zajonc (1989[2]), facial expression increases an emotional experience if the expression belongs to that emotion, and not showing the related facial expression inhibits it.

As the Output Generator determines which Body Signals to use, it can inform the Emotion Monitor. Something to note here: there has been a replication study of one of the studies about facial feedback hypothesis (2016 [1]). The findings of that study could not be replicated, which does support my point that there is merit in comparing different emotion models, and allowing different emotion models to be implemented in one's system, rather than committing to any one emotion model.

### 4.3.16 Sense of Normalcy and Order

While already touched upon slightly when discussing Likelihood, what an agent considers the "normal" or expected situation can impact the emotions felt by an agent. Scherer's model [38] alludes to this when discussing "novelty" as a check for emotion appraisal, which includes how familiar something is, relying on the schemata people have formed throughout their lives. To give an idea of how formed schemata and other learned or accepted pieces of information (and particularly subversions of them) can influence emotion, I'll give some examples.

Firstly, there is the psychological theory of cognitive dissonance (Festinger, 1957 [15]). Cognitive dissonance, according to Festinger, occurs when there is an "existence of non-fitting relations among cognitions", where cognitions are "any knowledge, opinion or belief about the environment, oneself, or one's behaviour". This is experienced as a sense of discomfort. This discomfort will then drive the person to somehow remedy the non-fitting relations. Someone might encounter dissonance when reconciling new information with previous assumptions and dealing with choices made while the other options were better in some respects. There are many ways to solve dissonance, and implementing cognitive dissonance as an emotion in its own right, or, as Festinger suggests, as a Motivational Value like hunger, might be very useful when thinking about how to implement the Inferencer. However, I would specifically like to focus on the initial feeling of dissonance.

In a practical sense, whenever the Inferencer notices an inconsistency in cognitions of the Short and Long Term Memory, the Sense of Order (of cognitions, in this case) factor comes into play, signalling that the order is not there. Additionally, when an agent is prompted to reconcile the new information, the explanations that it forms might trigger emotions indirectly. For instance, the explanation might be in the form of formulating Expected Consequences. The Consequences can also be that the agent will learn certain new information about past events.

To illustrate: the agent arrives at its home, and announces its arrival. Normally, its spouse answers back, and the spouse is expected to be home. The agent starts to wonder why the dissonance is there, and one of the explanations it considers is that its spouse got in an accident. By creating that Expected Consequence, the emotion "fear" is triggered. On the other hand, it could also just prompt the agent to remember information it already knows, but which has a Memory Association factor. For instance, the agent remembers its spouse died, prompting grief, or that its spouse promised to buy it a present and is therefore still out shopping, prompting joy.

Something else that causes emotions due to not being normal is the Uncanny Valley effect (Mori, 2012 [translation] [29]). This hypothesis states that, the more something resembles a human, the more humans like it, right up to the point where it resembles a human almost completely, except for some small detail. Whether this observation is related to cognitive dissonance or not, this observed effect also relies on something not quite fitting what the agent sees as "normal". Of course, effects like that can only be modelled if the agent is able to classify things as "normal", and if it is able to recognize when something is not part of that classification.

## 4.4 Effects Of Emotions

The effects emotions have on human behaviour are, as mentioned before, not entirely clear yet. However, there are some theories or indications, and I have written suggestions on how to implement these into a model. These effects may require an emotion to have an extra property, but they will often manifest in adjusted algorithms or formulas used in other processes, so that these use the properties of the emotions.

### 4.4.1 Emotions as Information

As detailed in "Feelings-as-Information Theory" by Norbert Schwarz (2012 [43]), the emotional state is used as information by the agent, to evaluate current situations. Unlike the Sensors, and the Motivational Values an agent is aware of, information provided by emotions extends beyond information about what is happening right here and right now (think of the emotions elicited by Expected Consequences, Sense of Normalcy and Order, Memory Association, etc.).

Another important distinction is that humans are not always clear on where their emotions are coming from, as detailed in ”Mood, Misattribution, and Judgments of Well-Being: Informative and Directive Functions of Affective States” by Schwarz and Clore (1983 [41]). When no cause of an emotion is inferred, people tend to assume it must be due to the current situation. What this means is that multiple causes may be ignored in favour of assigning the full emotion intensity to one cause, or the assumed cause may not be the (most) contributing cause at all.

For instance, in Schwarz and Clore's paper, they found people's judgement of their overall happiness in life depended on their mood at the time. However, when they were informed their bad mood was due to bad weather, they seemingly stopped using their bad mood as information. The same was not found when people in a good mood learned about the cause of their good mood, which was hypothesized to be due to bad moods requiring more scrutiny, as deviations from the norm.

When it comes to implementing such effects, the first important part is that the Inferencer (and other Processes) can easily read the Emotional State. This requires a ”translation” of all the intensities into more distinct states. How this translation would work depends on what emotion model is implemented: can an agent feel more than one emotion at a time, or does it feel a mixture of all emotions above a certain intensity threshold? Does the model differentiate between moods and emotional episodes, and does it make a difference between moods and emotional episodes, when it comes to feeling multiple at the same time? For instance, can two moods be felt at the same time, but not two emotional episodes? And, for individual agents, how skilled is the agent at determining which emotion(s) it is feeling? In all cases, the intensity values first need to be translated into discrete (or fuzzy) categories like ”high” and ”low”, which can then be interpreted further.

When it comes to determining causes, this would be done by the Inferencer in much the same way as it might determine the cause of Motivational values like pain. However, especially for positive emotions, assigning a cause does not need to be a priority. These causes can be stored in the Short Term Memory, and potentially move to the Long Term Memory as part of information about events, and as Memory Associations.

### 4.4.2 Memory

One function of emotion that I have already referred to in regard to Memory Association, but not yet expanded upon, is the effect on Memory. A recent review of memory and its relation to emotion (Levine and Pizarro, 2004 [22]) showed that current research seems to indicate that feeling strong emotions during an event, or as reaction to an event, can make the event easier to remember (although the event, including the emotion felt during it, is liable to be misremembered through the lens of information gained after it). And, conversely, that memories related to the emotion a person is currently experiencing are easier to remember, and therefore closer at hand when any judgements need to be made or conclusions need to be drawn. This effect on memory will likely be relevant to some of the other effects of emotions I will be discussing later, but for now I will be focusing on how to model such a memory effect on increasing ease of remembrance, drawing conclusions, and forming plans.

When I discussed the emotion elicitation factor of Memory Association, memories of any type can have an emotional association. The way that could be implemented in practice is by attaching an emotion tag to any memory or learned information. This way, other processes can filter and sort memories and information by emotion tag. For instance, the Memory Manager decides which information to transfer to the Long Term Memory based on their relevance. In this, an emotional association can be considered as something that makes the information relevant.

When it comes to recalling the memories, in what order information is considered can be quite relevant when there is a time limit, as is the case for the Inferencer and Planner. In an complex environment, not all possibilities can be considered, and not all information can be processed until there is no new information to derive from it. If the order in which information is accessed and used is influenced by the current emotion intensities (or rather, how these intensities are interpreted), then this gives the effect of emotionally relevant information being easier to recall.

In the case of the Planner, it depends on how that component is implemented how much of an effect the current Emotional State could have. For instance, in my implementation I'll be using Goal Oriented Action Planning (GOAP) (Orkin, 2004 [32]) to determine the agent's Plan. GOAP works by creating a path from the solution to the current situation: it looks at the situation of the solution, and finds an action that has the solution as its consequence and a situation one step closer to the current situation as a prerequisite. In other words, it reasons backwards from the solution, using some variant of a path-planning algorithm. In this case, if it prioritizes considering actions or the objects those actions are used on, that have a relevant emotion tag, the Plan it makes will be influenced by the current Emotional State. Previous Plans stored in the Long Term Memory can also be abstracted, which can make them more generally applicable and therefore reusable. And if the Planner is checking if it already has an abstract Plan to use as a staring point, the retrieval process can also use the emotional tags to sort for relevance.

### 4.4.3 Emotions as Focus

Emotion also seems to function as a way to focus the agent's attention. For instance, in the study "Emotion Drives Attention: Detecting the Snake in the Grass" by Öhman, Flykt, and Esteves (2001 [31]), it was found that while non-fear related stimuli were processed in a serial manner, a fear related stimuli was noticed in parallel. This suggests that another, presumably emotion-related, process took note of the fear stimuli.

Obviously, the most accurate way to model such an effect would be to mimic how perception processes work in the brain, but that might be a bit too ambitious for this thesis. Rather, I would like to suggest how we can get the effect of emotion causing some information to "jump out". When information is received by the Sensory Filter, it sorts/filters information and then sends it on, and if there is a strong Memory Association or the direct emotion causing factor Standards Rule is implemented, this can either simply influence the sorting or filtering, or cause that piece of information to be sent to the Inferencer and so on immediately. Of course, an implementation of the sort/filter process could already include sending information along in batches or however else could be useful (in other words, the Sensory Filter is not inherently required to send all information along at once).

Additionally, related to the research on emotion and memory previously mentioned, which pieces of information are still considered relevant enough to retain in the Short Term Memory by the Memory Manager could also be influenced by their associated emotion, meaning the information might be considered relevant to the current situation for a longer amount of time.

### 4.4.4 Emotions for Communication

Another theory on the purpose of emotions is that they are used to enhance communication: they serve a social function. A paper by Keltner and Kring, "Emotion, Social Function, and Psychopathology" (1998 [19]), collected research that indicated emotions and emotion signalling play a role in communication and social functioning, and showed that mental disorders that disrupt emotion functioning also disrupt social functioning. Emotions can have a social effect due to their causes, and the subsequent desire to avoid or achieve those causes. Emotion Signalling can quickly confer information on a person's state to others, aiding in communication. Some emotions can be beneficial to maintaining positive social relations, due to the context in which they occur. For instance, if people express embarrassment or guilt in reaction to one of their actions, that signals to others that they are aware their action was undesirable. And that gives others a reason to assume they will try to avoid the action in the future, so no (drastic) correction needs to be made.

To model this, the agent's emotions need to be communicated to other agents by the Output Generator, and the other agents need to be able to receive and process the signals. When I introduced the Output Generator, I already discussed this to some extent, including possibly implementing an ability of the agent to send Emotion Signals not in line with the current Emotional State. As discussed, this can be extended up to the point where Emotion Signalling is its own action.

### 4.4.5 Goal Managing and Priority

When the Goal Manager decides which Goal(s) to focus on first, there needs to be some way of ranking the Goals. In order to do so, they need to be evaluated on something (urgency perhaps). Additionally, before ranking them by evaluation, they could also be ranked by type (Goals that are punishment minimizing before Goals that are reward maximizing, to give an example). In either manner, the current Emotional State might influence this ranking. For instance, humans have been observed to be more likely to give into hedonistic impulses when they experience negative emotions (Tice and Bratslavsky, 2000 [48]). The intensities of undesirable emotions might be part of the function that calculates a Goal's evaluation, and the Goals could be divided by which Emotion they are trying to increase or decrease, or by whether they pertain to Emotion or a Motivational State, and so on.

Emotions could also play a role in how you implement a particular Goal Manager architecture. For instance, I'll be using a BOID architecture in my implementation, based on work by Jan Broersen et al. (2001 [8]). BOID stands for Beliefs-Obligations-Intentions-Desires. The BOID architecture relies on an agent that has at least those four components, which output their respective influences on decisions under certain inputs. Any conflicts between these outputs are resolved according to the agent's type. The BOID architecture introduces several agent types: realistic agents prioritize beliefs over other components (so they adjust their Goals and Plans according to new information); selfish agents prioritize desires over obligations (in the context of MEIA, an example of how to do that would be prioritizing Goals based around increasing positive emotions over Goals arising from certain Standards or promises made to other agents); social agents prioritize obligations over desires; simple-minded agents prioritize prior intentions over desires and obligations. Integrating emotion into this architecture can be done by letting the Emotional State influence what type of hierarchy is used. An agent that has a negative emotion with a high intensity might change from social to selfish, for instance, to model increased hedonism.

### 4.4.6 Time Pressure and Creativity

According to a review of research into the effect of emotion on cognition by Isen (2001 [17]), there have been many observations that positive feelings increase creative problem solving. According to research by Isen, Daubman, and Nowicki (1987 [16]), this may have a lot to do with positive emotions increasing how related different concepts are considered to be. If someone is able to link the information given to them with more knowledge and other information, that allows a wider variety of solutions to problems or new ideas to occur.

To model such an increase in creativity, the Inferencer could be allowed to run longer, or preferably, use methods for reasoning that use more of a top-down approach than a bottom-up one. Creativity can also be modelled in less direct ways. For instance, an agent, when forming a plan, will also look at plans they made in similar situations in the past, and might use these as a starting point. If the agent has little time pressure due to non-emotional factors and is feeling a creativity promoting emotion, they might try to find a solution from scrap as well, to see if they can find an ever better solution, while a non-creative emotion could make the agent "settle" for the known plan. This requires an emotion to have either a creativity-promoting property or not.

#### 4.4.7 Risk Taking and Outcome Assessment

There have been a number of studies on how emotions influence risk-taking. Some of these studies only looked at whether the felt emotion was positive or negative, but others also looked at the effects of specific emotions, and the intensity of the emotion. I will describe some of these studies, and then suggest a way they can be implemented.

The first is a study by Lerner and Keltner, "Beyond valence: Toward a model of emotion-specific influences on judgement and choice" (2000 [21]), which examined the effects of anger and fear on the decision making process. They did this to show that emotions with the same valence (positive or negative) could have different effects. They found that anger made someone more likely to judge a risk as low, while fear did the opposite. They raised the possibility that this was due to the informative properties of the emotion and/or due to the emotion's influence on memory recall (as described before), where anger made people recall times they were in control.

Another study that compared two negative emotions by Raghunathan and Tuan Pham, "All Negative Moods Are Not Equal: Motivational Influences of Anxiety and Sadness on Decision Making" (1999 [34]), found that anxiety caused people to prefer the low risk/low reward decisions, while sadness made people prefer high risk/high reward. They theorized this to be because anxiety makes a person focus on preventing negative events and sadness on feeling better than they do now. However, they also noted that their findings were not entirely compatible with previous research, and theorized this was due to heightened arousal also increasing risk-taking behaviour, irrespective of the particular emotion. They also referred to other studies that seemed to indicate that the Emotional State is still used as information in judgement even when the cause for a particular intensity has nothing to do with what is being judged. Namely, they found that an angry mood would bias people to judge an event to be caused by other people (because anger tends to be caused by people), whereas sad people would be biased towards considering situational factors.

Isen's review on positive affect (2001[17]) indicated that positive emotions made people more inclined to take risks, but only if the stakes were low, making it seem like happy moods inspire a desire to keep them. People were also found to rate bad consequences as worse, which may be related.

Another study, "Mood and Persuasion: Affective States Influence the Processing of Persuasive Communications" by Schwarz, Bless and Bohner (1991 [42]), aiming to find out if emotion could influence information processing style, found that people in a good mood relied more strongly on heuristic and experience, as opposed to people in a bad mood, who were more likely to be motivated to look at an argument critically (possibly because the bad mood indicated there was reason to be on their guard). This seemed to be due to the amount of processing the people were inclined to do: when the people were distracted, the people in a bad mood were affected, but not the people in a good mood, because they hadn't used the processing time when not distracted either. They did note that people in a good mood could be persuaded to use a more thorough processing method, if they considered doing so more important.

Now, it should be noted that all these studies include the disclaimer that their results are not conclusive: a lot more research needs to be done, and they sometimes seem to contradict with previous research. However, they do seem to indicate that emotion affects risk taking and judgement, so what I will do is propose ways to adjust these processes using the current Emotional State of the agent.

Firstly, when determining possible outcomes to a Plan by the Planner, the formula to calculate the chance of these outcomes happening can be adjusted. Every Emotion could have a property indicating whether they skew the formula to optimism or pessimism. Secondly, once these chances are calculated, the thresholds for "acceptable risks" should be determined by the emotion (and its intensity, by simply increasing the risk threshold as intensity increases). As it's not really clear yet how this works in humans, I'm going to suggest a method to model this based around its effects being similar to what the research has found, not necessarily the workings. This requires considering how "wanted" the current Emotional State is in comparison to the possible Emotional State.

To do so, consider for each Emotion: what tends to cause the Emotion to increase in intensity, and would it make an agent want to experience another emotion? For instance, sadness tends to be caused by non-positive things happening, and an agent would probably prefer most other emotions. So, we can simplify it as an Emotion with the property "want-to-lose". On the other hand, while anxiety is also an emotion that an agent would prefer not to have, the anxiety may be precisely because bad things could happen, so we can simplify it as "want-to-keep", or as "want-to-lose-to-positive" (to allow more rankings and difference). This is just a suggestion, however, and undoubtedly a more accurate way to model such an effect can be thought of, especially as more research is being done.

However, what to do when there is too little time to calculate something? For instance, let's say there's no time to calculate possible outcomes. In that case, agents would just expect either the intended outcome (from the Planner) or the first associated outcome (from the Inferencer, which might use an Emotion-sorted list) is going to occur (although they might experience cognitive dissonance due to not matching this expectancy with other beliefs). But let's say the outcomes can be considered, but not their chance of happening? In that case, you could have the agent expect the outcome that matches their Emotional State.

On the other hand, let's say we have plenty of time and the agent can get a bit more complex in its considerations: if we go by the studies, we should perhaps involve the Long Term Memory in risk calculation. The Long Term Memory has learned behaviour stored, possibly as part of a classical conditioning learning method. In the less time-consuming method above, you would just use those behaviours and their likeliness, and modify according to the Emotion properties later.

However, if we want to have a way of calculation that is closer to how the brain is theorized to do it, then the Long Term Memory needs to store the expectancies by Emotion tag. And then the current Emotional State would determine which expectancy to consider the most applicable. You might not want to completely ignore the learned expectancies from other Emotion associations, so then there would need to be a formula that combines these expectancies. They would then have to be weighted by the current Emotion's property modifier, as well as of course with how many examples an expectancy was formed (as in most learning algorithms). If we then also want experiences that occurred further in the past to have less influence (which many learning algorithms do), then we are looking at a quite expensive way of calculating risk allowance and outcome assessment. Considering that, as well as the fact that it is unknown how much value such a method would add, I will not be using this in my implementation, rather preferring to simply adjust the formula at the last step.

The second adjustment emotion seems to make to judgement is which type of processing information and plan-making method to use (heuristic-reliant, or critical and less reliant on heuristics). While the Planner and Inferencer might make different decisions regarding when to stop searching and considering due to a time limit, we can also have that decision be influenced by the current Emotional State. If the Emotional State indicates there is no need to worry because things are going well (aka the agent is experiencing positive emotions), there is less pressure on an agent to try to see if any information doesn't match their heuristics, and less pressure to look for alternative solutions. Note that this seems to contradict the notions about creativity and good moods from the previous section: however, they can be easily combined by making sure the creative solutions are only attempted if it is deemed important (however that is decided).

# 5 Emotion Integration in a Specific Agent Architecture

## 5.1 Setup

In the preceding chapters, I have already given several suggestions on how MEIA can be implemented, and how to use it in order to implement particular emotion models. In order to properly illustrate how the suggested emotion integration can be implemented, I will provide some more extensive examples in this chapter. In order, I'll describe an agent architecture that allows for all the different emotion integrations I've introduced, specific emotion models for that agent architecture, and some resultant behaviours of different emotion models. Initially, I wanted to write an actual program. For several reasons, which I will discuss more in depth in chapter 6, I have decided to restrict myself to a description of an implementation.

I will detail how each process as described in chapter 3 could be implemented, with all suggested emotion integration from chapter 4. For each process, I'll provide a UML activity diagram for clarity. I would like to preface this description with saying that this implementation was chosen to be relatively simple, while still allowing examples of how to implement the different causes and effects I've introduced. Undoubtedly, these causes and effects could be implemented in different and more interesting ways in a more complex implementation, but that would quickly make the illustrative examples more complex as well, which might make them less illustrative.

Whenever an activity uses the data "Operations Allowed", that activity should be implemented in such a way that it can terminate and still continue on as if it had been allowed to finish all operations in the loop. In other words, it uses an "anytime algorithm", as coined by Thomas Dean in 1986 [12]. This "Operations Allowed" data is part of the implementation of chapter 4.4.6 and 4.4.7. To summarize the relevant parts of those chapters: People become more creative when experiencing certain emotions, and more or less thorough in their reasoning in others. We can have this influence how much time an agent is allowed to spend on cognitive processes.

The total "Operations Allowed" should therefore be some function of time pressure (the higher the pressure, the lower the value), desire to be thorough (higher desire means a higher number of operations allowed, dependent on Emotional State), and creativity (likewise, and also dependent on the Emotional State, but only not zero when the current Process was called with the mark "important"). A Process would be marked as important if it was called as part of trying to fulfil a Goal (for instance). Calculating the "Operations Allowed" should not involve deriving any new facts.

In the UML diagrams, these "Operations Allowed"-related elements will be coloured red, while elements relating to integrating emotion causes will be green, and those related to emotion effects will be blue. I will conclude describing the Processes by giving a list of all the variables that can be adjusted to arrive at a different emotion model. I will also give some examples of agents with differing emotion models that use the detailed implementation, which will involve some suggestions on how a character writer could use a MEIA implementation for creative purposes.

## 5.2 Processes

### 5.2.1 Sensor

In this implementation of MEIA, the Sensor is the only process that receives input from the environment. Therefore, I will first describe the assumed environment. The Environment consists of (or is translatable into) a set of logical facts. These facts are about which objects exist, what properties these objects have, which actions can be performed on these objects, and which relations between objects exist. At a regular time interval, the Environment will send a subset of these facts that the agent can observe to the Sensor. Which facts the agent can observe would be decided by the Environment based on the relation of the agent to the objects. This subset of facts is what I'll refer to as an Update On Environmental Facts.

Another Update the Environment sends the agent is on its Motivational Values. While some of these Values may be part of the Update On Environmental facts (in other words, be observable), some may not be, and they can nevertheless influence the agent (as discussed in chapter 4.3.2). Lastly, the Sensor informs the agent when an Action of it has been completed, and with what outcome. Effectively, this leaves us with three Sensors, two of which I've illustrated in Figure 1 below. The Action Sensor simply sends the information directly along to the Planner, no matter the contents.

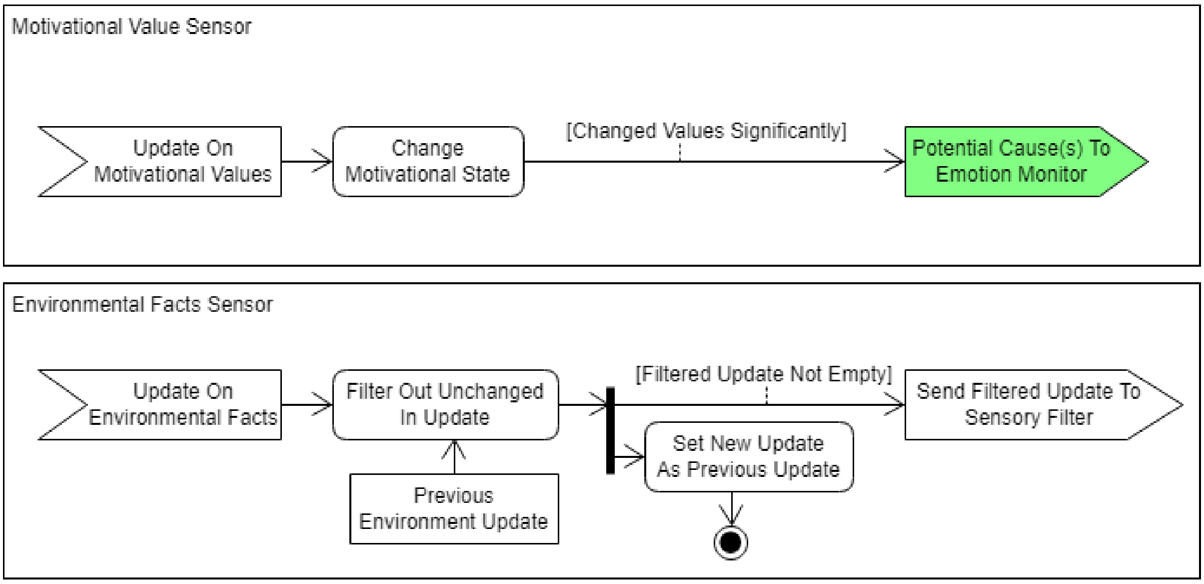


Figure 1: UML diagram of two of the Sensors.

When the Motivational Value Sensor receives its Update, it changes the Motivational State where necessary. If the changes are "significant", it then sends the Emotion Monitor a signal, so it can see if the Emotional State needs to be adjusted. This takes care of implementing the first part of chapter 4.3.2. When I use the word "significant" here, I leave it open to interpretation what that means. Any change could be significant, but signalling the Emotion Monitor can also be limited to when a Motivational Value changes from being inside one predefined range to being in another. For instance, a hunger value might have the predefined ranges "starving", "satisfied", and "overstuffed".

The Environmental Facts Sensor also checks if there are any changes. However, it doesn't use the Long or Short Term Memory for this, as those are the beliefs of the agent, which do not necessarily map directly to what was last observed. If there are changes, they are sent on to the Sensory Filter.

### 5.2.2 Sensory Filter

The job of the Sensory Filter is basically to make sure that the new information is processed in order of importance. Exactly how this importance is decided, I've left vague, but it should be some function of, potentially, four factors. First, to what extent the new information relates to the current Plan. As the Sensory Filter is not supposed to perform extensive reasoning, this calculation should be simple and straightforward. Possibly, it could be a simply check of whether the new information matches any of the requirements or effects of any of the steps of the Plan.

The other factors influencing the outcome of the function are dependent on the emotion model. One might be a rule in the agent's Standards regarding direct emotion causing, another might be an emotion tag. These factors are both a result of wanting to implement chapter 4.4.3. Lastly, the Emotional State could modify how much weight is placed on an Emotion rule or Emotion tag, depending on whether it matches the current Emotional State, as part of implementing chapter 4.4.2 (namely, memories relating to the current Emotional State being easier/quicker to access).

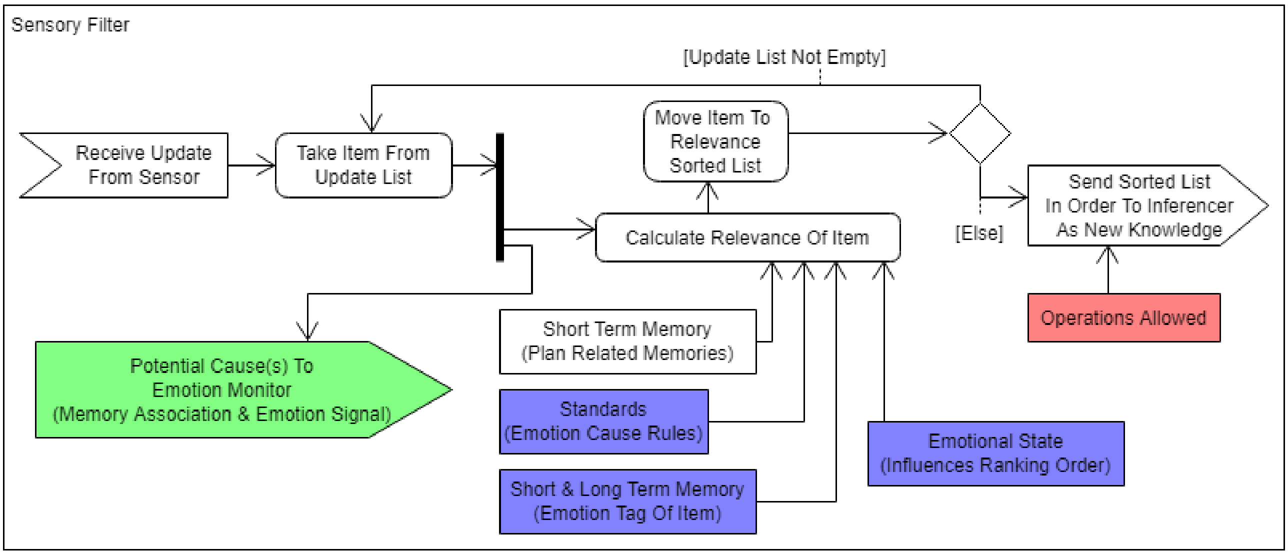


Figure 2: UML diagram of the Sensory Filter.

As can be seen in Figure 2, before a fact's relevance is assessed, the Emotion Monitor is sent potential emotion causes. Namely, Memory Association (chapter 4.3.8) and Emotion Signalling (chapter 4.3.14). When the Update has been sorted by relevance, the facts are sent to the Inferencer, one by one. In order to keep my example relatively simple, I'm not going into great detail how to deal with time limits unless these are directly relevant to the emotion integration. Obviously, if the agent truly has limited time, we would want the agent to start processing new information as soon as possible, in which case sending the information on to the Inferencer should not take place after sorting, but somehow during it. The Inferencer could then put any processing on hold if the next piece of information was more relevant. For the purpose of this example, however, I'll pretend the agent has unlimited time, and any time limit is artificial for the purpose of simulation. However, if the agent runs out of time and/or attention (referring to the "Operations Allowed" data), not all information is necessarily processed, making the ordering even more relevant.

### 5.2.3 Inferencer

The Inferencer is where the logical reasoning takes place. The input this Inferencer receives is always some sort of first-order logic statement. The main purpose of the Inferencer is to provide an answer to a query, using the facts in the Long and Short Term Memory and the rules in Standards as its knowledge base. The implementation I had started building used a modified Prolog interpreter, which may provide some useful context in understanding the Inferencer as described here.

When the Inferencer receives a query, this query can be with or without variables. If it has no variables, the answer given can either be `true` or `false`, where false may just mean it couldn't find a line of reasoning under which the query was true, not necessarily that none exist, just like Prolog. A difference however, is that this Inferencer remembers the line of reasoning that led it to conclude the query was `true` or `false`. In the latter case, what is actually remembered is the line of reasoning up until the point where the knowledge base was found not to contain a fact that would've led to the query being `true`, or when it ran out of time.

I'll give an example using Prolog notation, where terms that start with a lowercase letter are constants and those that start with an upper case are variables, `\+` is a negation, and `A :- B` means "if B, then A". Let's say the query received is `rich(me).` The only thing in the knowledge base is the Standards rule `rich(Person) :- hasLotsOfMoney(Person).` (Someone knowledgeable about Prolog might note that this could lead the program to crash, but let's assume we can deal with dynamic knowledge bases). The Inferencer would first match the query with the one rule, and find it can unify it to become `rich(me) :- hasLotsOfMoney(me)`. Now, it will start a sub-query, in which it needs to check whether `hasLotsOfMoney(me)` can actually be derived from the knowledge base. Of course, it can't, so it returns `false`, with the reason `\+(hasLotsOfMoney(me))`. It will also return that if it ran out of time before it could find the knowledge `hasLotsOfMoney(me)`. Either way, this indication on how to make the query `true` will be used later in the Planner. If the query contains variables, it will return for which constants the query is `true`, or if `false`, it might indicate which constants would have made the query `true`. For each answer to a query or sub-query that is `true`, the answer should also be added to the Short or Long Term Memory. This line of reasoning can be remembered for a later purpose, basically being remembered alongside the fact itself.

Aside from regular first-order logic rules, Standards also contain default rules. A default rule, in this Inferencer, takes the form: `A :- B`, but only if justification `C` is not provably `false`. What this means is that the Inferencer will only apply the rule `A :- B` if the query `\+ C` returns `false`. That means that the order in which default rules are considered may decide what the agent decides is `true` or not, if a default rule's justification is only provably `false` after applying a different default rule. These order dependent solutions are called extensions. If the agent cannot find an answer after a default rule was applied, it may be able to find an answer in a different extension. However, whether it is allowed to adjust the ordering is dependent on whether the default rules to be switched around are in the same Standards category or not, which is related to implementation of chapter 4.4.5, "Goal Managing and Priority". I will expand on that later.

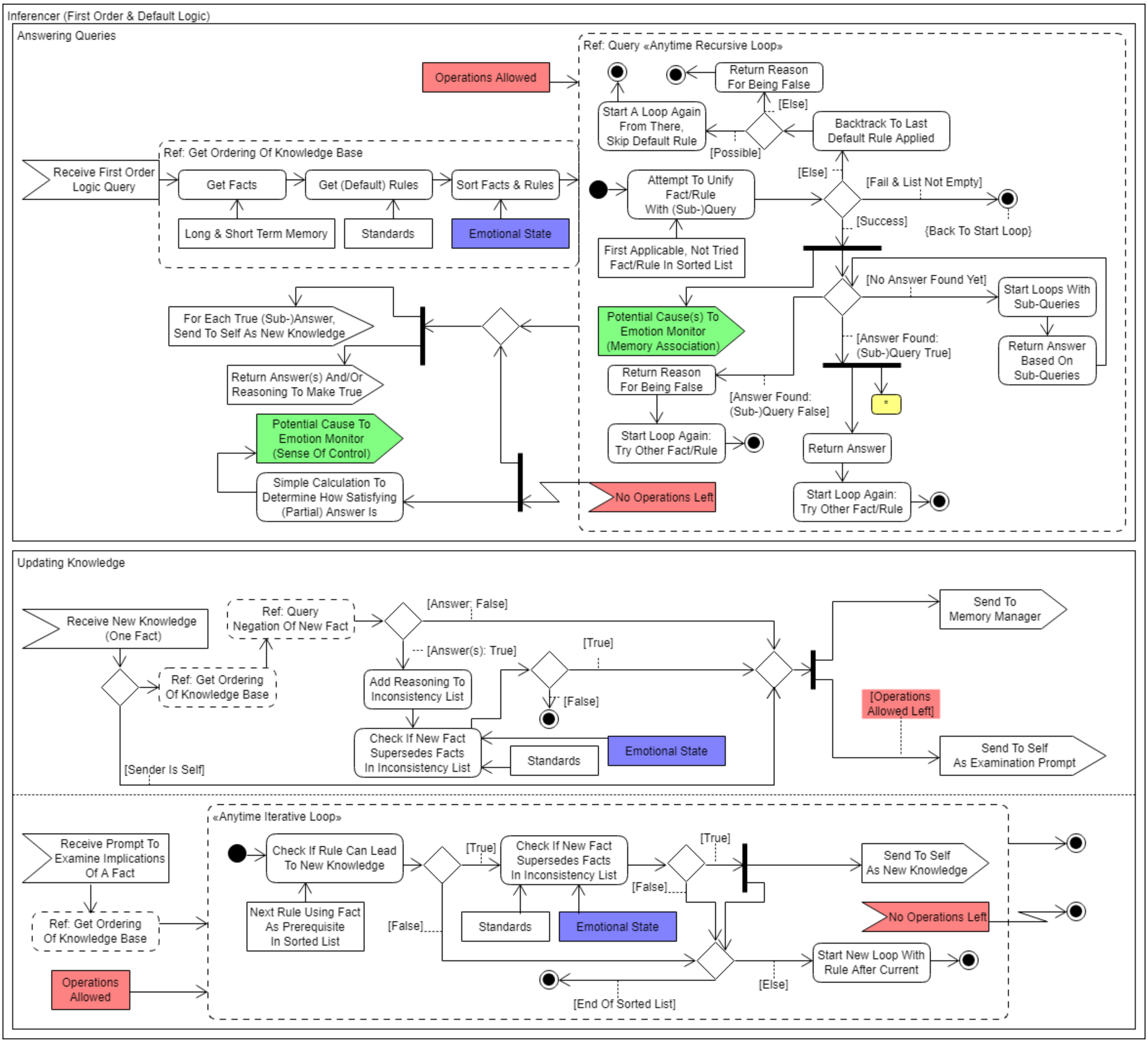


Figure 3: UML diagram of the Inferencer.

Now that I've explained how the Inferencer performs its reasoning, I'll walk through the UML diagram in Figure 3. First, there is the main method of how the Inferencer answers queries by other Processes. It will first sort the knowledge base, which can be influenced by the Emotional State, as per chapter 4.4.2 on the effects of emotion on Memory. Because the Inferencer's attempts to find an answer to the query can be stopped at any time, the order in which it searches the knowledge base and starts sub-queries (in a depth-first manner) matters in regard to what answer can be found. The order of default rules also matters, even if the agent is allowed to infinitely search for an answer. The sorting is done as follows. First, there is the same type of sorting that the Sensory Filter did, namely, relevance. This can be applied to rules by simply treating the individual terms as the factors in sorting. For the sorting of default rules, we can also sort them according to types. In what order to apply these two types of sorting, I'll leave up to the reader, as I've found no evidence yet in favour of either sorting resulting in more believable behaviour.

The types of rules can be anything, but they are specifically meant to allow implementation of something like the BOID architecture (2001[8]) mentioned earlier. So, to use that architecture as an example, the types of default rules would then be Beliefs, Obligations, Intentions and Desires. In the BOID architecture, the order in which the rules are applied, and consequently, which extension is arrived at, is dependent on the type of agent. A BIOD-ordered agent will first try to apply default rules belonging to the category of Beliefs, then Intentions, then Obligations, and lastly Desires. Applying these rules results in a new extension, on which it again tries to apply the default rules in order, until no rules can be applied any more.

To quote the 2001 BOID paper, "beliefs are informational attitudes – how the world is expected to be – obligations and desires are the external and internal motivational attitudes, and intentions are the results of decision making". BOID was specifically intended for resolving conflicts when an agent decides which action to take. In this case, the action is deciding on what it believes. This may not seem directly relevant to categories of Obligation or Desire, but if the agent is deciding on subjective matters, it might. For instance, let's say an agent is deciding on whether another agent is trustworthy. The agent might desire to think positive, and has a default rule that people are trustworthy unless proven otherwise. However, the agent also promised itself it would not trust people who have not earned that trust, which is the exact opposite rule. If the agent applies Desire rules before Intentions, it will conclude the other agent is trustworthy; otherwise, it will conclude the opposite. As mentioned in chapter 4.4.5, negative emotions might cause an agent to prioritize hedonistic desires more, which we can mimic by placing Desire type rules earlier in the ordering when an agent's Emotional State is negative.

Of course, there is no need to restrict oneself to those four types: you might want to split Desire into multiple categories, ranging from desires of self preservation and protection of loved ones, to desires of which colour clothing the agent prefers. One could also imagine that Obligations to society and Obligations to friends might get sorted differently for different agents in different moods.

Once the knowledge base is sorted, the Inferencer attempts to find the answer as described earlier. When a fact or rule is successfully unified with the query, the Emotion Monitor is sent the Memory Association of that fact or rule as a possible emotion cause. Basically, if a unification is successful, I consider the agent to have actively remembered that fact or rule. Of course, one could just as easily decide to send the Emotion Monitor a Memory Association for each attempt, or only for those facts and rules that are part of the answer. Which implementation turns out to be the most useful would need to be researched.

Once an agent has found an answer, it will return that answer. If there is still time/attention left, it will try finding an alternative answer by attempting to unify the query with another fact or rule instead (especially useful if the query contained variables). If the answer to a query is found to be true (for a certain set of variables), this information could be immediately added to the knowledge base (at the place of the yellow asterisk in the UML diagram), and therefore trigger emotion causes immediately. However, I figured it was best to keep the example simple, and not adjust the knowledge base in the middle of using it. Rather, it adds this information after the query has been answered.

If the Inferencer is stopped from trying to find an answer because there was no time/attention left to spend, the Emotion Monitor is informed of how satisfying the partial answer is, as a function of how important getting the answer right is, and how well it was answered. This is one part of how one can implement the Sense of Control emotion cause of chapter 4.3.7.

The Inferencer is also in charge of checking whether new knowledge can be added to the knowledge base without inconsistencies, and seeing whether the current knowledge base can be used to derive more information. If the Inferencer receives the prompt to add a new fact from itself, it will assume it is consistent (at least with the current ordering of the knowledge base). Otherwise, it will basically perform a reverse query of the fact. If it finds that the current knowledge base allows it to derive the opposite of the new fact (a contradiction), then the new fact and the facts that contradict with it are compared. This comparison is based on how trustworthy these facts and rules are, which requires knowing what the source of the knowledge was. For instance, the line of reasoning that led to it, or its rule type, or how long it has had this knowledge. These sources can then be compared using Standards rules, whose order can of course be adjusted by the Emotional State again, depending on the emotion model.

If the new fact is considered okay to add, it, and any facts that need to be discarded, are sent to the Memory Manager. If there's any time/attention left, it will also prompt itself to consider the implications of the new fact: maybe even more facts can be derived using it. If it does do so, the Inferencer basically does the reverse of finding an answer to a query. Instead, it sees if the fact to examine can be unified with any of a Rule's prerequisites. In the example rule given earlier, `hasLotsOfMoney(Person)` was the prerequisite of the rule `rich(Person) :- hasLotsOfMoney(Person).` So examining the fact `hasLotsOfMoney(bill)` and uniting it with that rule would lead to the new fact `rich(bill).`

### 5.2.4 Memory Manager

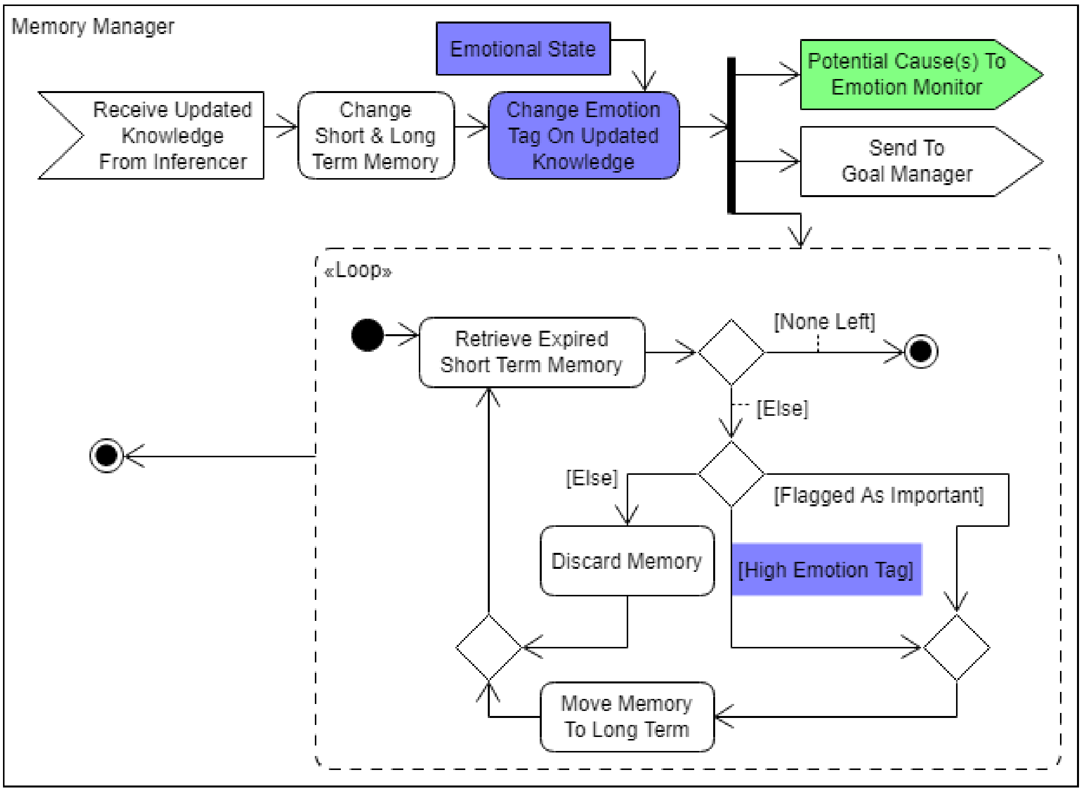


Figure 4: UML diagram of the Memory Manager.

After the Inferencer has confirmed that new information is consistent with the the rest of the information (as far as the agent can derive), the Memory Manager actually adjusts the Short and/or Long Term Memory. It also adjust the Emotion tag associated with the updated knowledge, depending on the current Emotional State, as part of implementing chapter 4.4.2. This adjustment could be adding a small increase in the associated Emotion tag intensity for any Emotion that currently has a very high intensity, or it could be a copy of the intensities of the current Emotional State; that would depend on the emotion model.

The Goal Manager is then informed that the knowledge base has changed, so it can check if any Goals have been fulfilled and so on. The Memory Manager will also send the change on to the Emotion Monitor, in case it's a fact about any of the causes, like Attribution (chapter 4.3.1), Likelihood (4.3.3), Predicted Alternative

Events (4.3.4), Expected Consequence (4.3.6), Sense of Control (4.3.7), Standards Rule (4.3.9), Confirmation or Disconfirmation of Current Emotion Cause (4.3.10), Changes in Evaluation (4.3.12), Changes in Knowledge of Another Agent (4.3.13), and Sense Of Normalcy and Order (4.3.16). The Emotion Monitor has the means to recognize when a fact is about any of these possible emotion causes.

The Memory Manager also checks if any memory that has been in the Short Term Memory without getting updated for a certain amount of time (while not being part of the current Plan) can be removed, or moved to the Long Term Memory. It moves it to the Long Term Memory when it passes some sort of check that flags it as important (however one chooses to define that), or, as part of implementing chapter 4.4.2, if the memory has a strong association with an emotion.

### 5.2.5 Goal Manager

In order to describe the Goal Manager, I'll first describe how a Goal is constructed. A Goal is a set of logical facts that the agent should make sure are true (in which case it's fulfilled), and conditions under which the Goal can be removed. These remove conditions do not need to be achievable: for instance, one usually would not want the Goal "stay alive" to ever be removed, in which case the the condition to remove it should be set to `false`. Additionally, every Goal has a type, which can be used when the Goals are sorted, in the manner that was also used in the Inferencer. Some Goals, the agent might start out with, in which case the type can be predefined. Other Goals may arise as part of long-term planning, in which case they would be the same type as the Goal that lead to the Plan that generated the Goal. Yet other Goals might arise from the Inferencer deriving that a Goal should be added. For instance, from the rule "if I make a promise, I should keep it" and fact "I promised I would do X", the Inferencer could derive "I should do X". For this implementation, I'll assume such facts are easily recognizable by the Goal Manager. In this case, the type of the Goal might depend on the type of the rule. To simplify Goal generation, I will assume all Goals are part of the Short Term Memory, and can be read by the Inferencer as a fact. If an Action from the Planner adds a Goal, that means it's adding a Memory, meaning the Goal Manager will be prompted in the same way as when the Inferencer derives that a

Goal should be added.

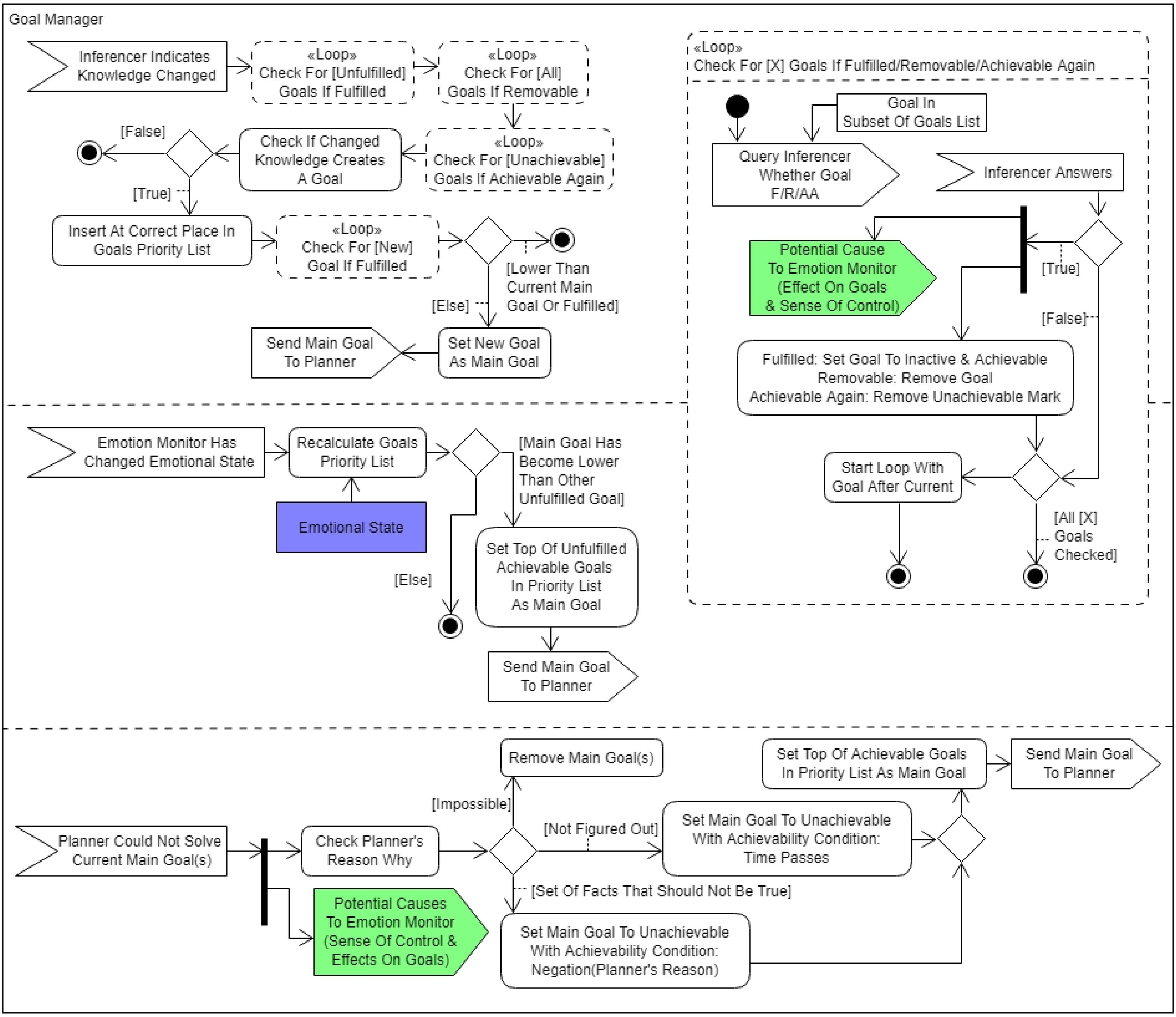


Figure 5: UML diagram of the Goal Manager.

The Goal Manager is called upon in three different situations. The first is when the Memory Manager has changed a fact in the Memory. The Goal Manager will first check if any unfulfilled Goals have been fulfilled and then if any Goals can be removed. It does so by sending a query to the Inferencer. In the same way, the Goal Manager will also check whether any Goals that had been marked as "Unachievable" when the Planner was unable to construct a Plan for them, can have that mark removed. You may note that a cycle between the Goal Manager, the Inferencer and the Memory Manager has now been made, however, an infinite loop is avoided as the Inferencer is beholden to time/attention limits. When a Goal is fulfilled, removable or achievable again, the Emotion Monitor gets informed of this, including which type of Goal it was. The type might be relevant to which Emotion is caused, if the types are differentiated by whether the Goal was positive maximizing or negative minimizing, for instance.

After that, the Goal Manager checks if the new knowledge leads to adaptation of a new Goal. If it does, it needs to be placed at the right point in the Goals List so it remains sorted correctly. If the new Goal is higher in the list than the current Main Goal, and not fulfilled, the Planner needs to make a Plan for this new Goal instead.

When the Emotional State changes to another State, the ordering of the Goals may need to be altered, if the Emotional State has an influence on that ordering.

As mentioned, the Goal Manager can also be called upon by the Planner when the current Goal cannot be planned for. In that case, the Emotion Monitor is informed and the Goal is either removed (if it was deemed impossible), or it is marked as Unachievable. The condition under which this mark can be removed can either be a timer, if the Planner did not give a reason, or it can be the negation of the reason (in other words, when the reason ceases to exist). Lastly, the next unfulfilled Goal not marked Unachievable on the Goals List should be set as the Main Goal, after which the Planner should do its job.

### 5.2.6 Planner

As mentioned before, I will be describing how to integrate emotions into a Goal Oriented Action Planner. A Plan in this implementation consists of a sequence of (abstract) Actions. In the case of an abstract Action, it can be broken down into another sequence of Actions, which can also be abstract. If all abstract Actions are broken down, the Plan consists of a sequence of concrete Actions, each associated with the object it can be performed on. As an example, let's consider a Plan to make a pillow fort. The sequence of abstract Actions might then be "collect pillows", "collect blanket", "place pillows and blanket". The abstract Action "collect pillows" can be broken down into "walk to pillow A", "pick up pillow A", and so on. Each Action has a precondition (a set of facts that need to be true if the Action is to be performed) and a set of possible outcomes. These outcomes, or postconditions, are a set of facts that are true if that outcome occurs.

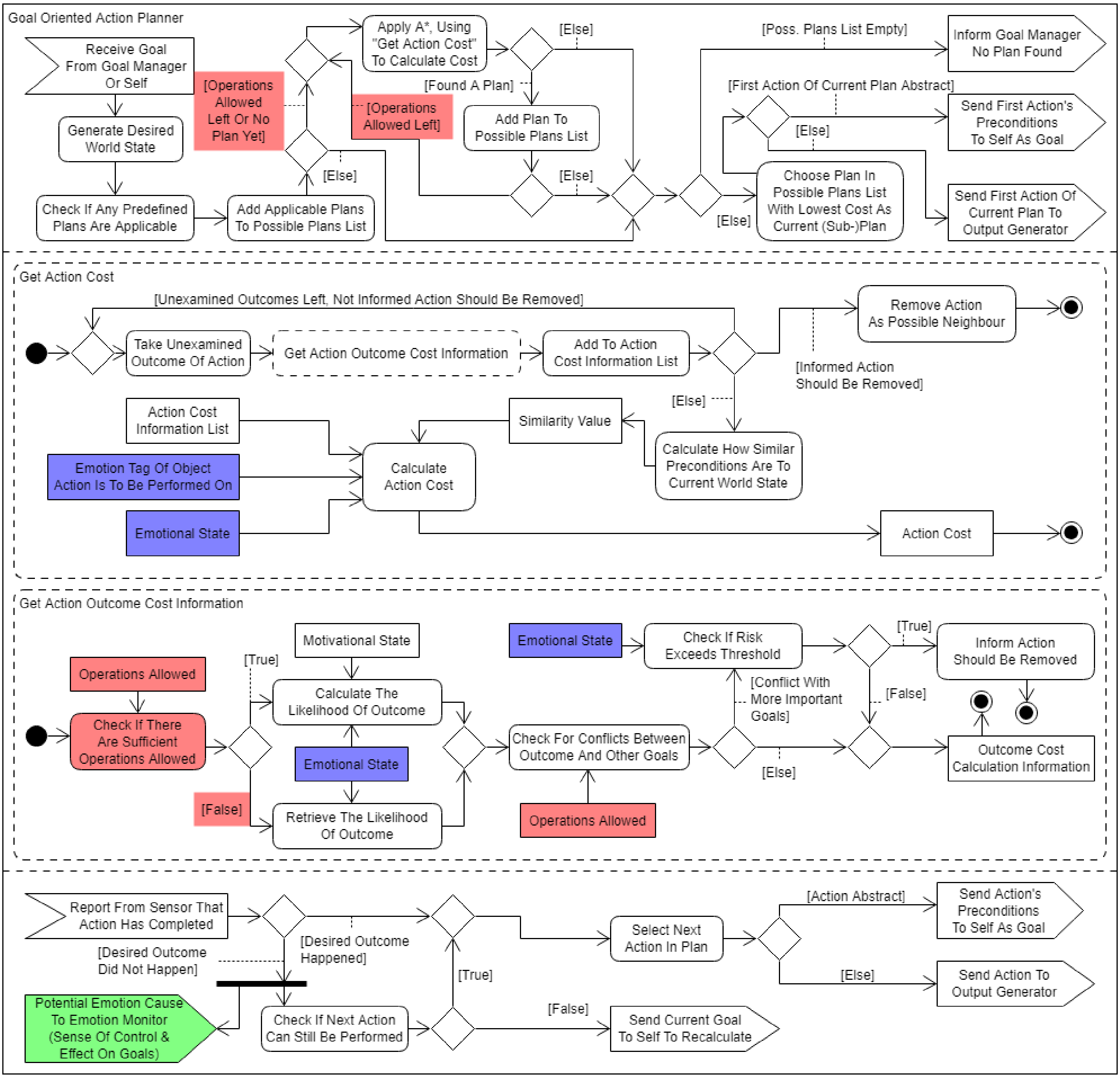


Figure 6: UML diagram of the Goal Oriented Action Planner.

The Planner is activated when it receives a Goal from the Goal Manager. It uses this Goal to generate a desired world state, namely, the set of facts that would cause the Goal to be fulfilled. The current world state consists of all facts in the Long and Short Term Memory. The Planner is going to find a path between the two states using the A* algorithm, starting from the desired world state. I will assume the reader is familiar with this algorithm. To find the world states adjacent to the desired state, it will use the preconditions and sets of postconditions of Actions as the paths. However, it will first check whether a predefined Plan could be used: these Plans are added to the list of possible Plans to choose from. If the agent has enough time and attention, or if no predefined Plan was found, it will then try to think of a new Plan.

For all facts that are in the desired world state, but not in the current world state, the Planner finds those Actions that have that fact as a result of one of their outcomes. For simplification purposes, I'm assuming the agent has access to a list of all possible Actions that could be performed on all possible objects. A precondition common to all non-abstract Actions should therefore be that the object to perform the Action on exists. It should also be noted that the agent itself, and other agents, count as objects in this case. The costs of the Actions are calculated as some function of, for instance, how close the preconditions of the action are to the current world state, the likelihood of the desired outcome, and the emotion associated with the object. That last factor's weight could of course be dependent on the current Emotional State (per chapter 4.4.2), but generally, emotion tags matching the current Emotional State, or just emotion tags with high values, should make the cost lower. In other words, they should be preferred over other actions, to mimic the effect of emotion on memory (chapter 4.4.2), and of emotion as focus (4.4.3).

The likelihood can also depend on the Emotional State, as discussed in chapter 4.4.7, about the effect of emotion on risk taking and outcome assessment. If the agent has little time or attention to spare on calculating the likelihood, it could use some predefined estimate. In that case, the action might have different predefined estimates associated with different Emotional States. If the agent does have time or attention to calculate the likelihood, some weight in the function could be different depending on the Emotional State.

The likelihood of the other outcomes should also be considered, in regard to risk-taking. As described in chapter 4.4.7, the amount of risk an agent should be willing to take can be dependent on the Emotional State. The Planner should therefore quickly be able to evaluate how bad an outcome would be, in terms of violating more important Goals than the Main Goal. In other words, if the outcome includes a fact that is the negation of one of the fulfilment facts of a more important Goal, the Action is considered risky. The thresholds for when a risky Action's outcome has too high a likelihood to perform the Action can be dependent on the Emotional State.

The A* algorithm is then applied until a complete Plan is formed, although abstract actions do not need to be broken down until the agent tries to perform them. If the agent has enough time and attention, the Planner will try to find some alternative Plans to add to the list of possible Plans. If no Plan was found at all, and there was no predefined Plan, the Goal Manager is informed that the current Main Goal cannot be fulfilled right now. If the Planner noticed a reason why the planning failed, this reason is sent along as well.

If there are multiple possible Plans to choose from, the Plan that has the highest chance of success with the least amount of cost should be chosen. The likelihood that the Plan succeeds is some function of the likelihood that each individual Action succeeds; this calculation can be influenced by the Emotional State in the same manner. What balance between cost and success chance should be preferred can again be influenced by the Emotional State, as part of emotion's influence on risk taking. The first concrete Action of the chosen Plan is then passed along to the Output Generator. If the first Action of the Plan is abstract, it first needs to break down that Action.

The Plan, including its likelihood calculations per outcome, is stored in the Short Term Memory, and interpretable by the Inferencer as a set of facts. This means that these facts can trigger emotions as part of the causes Likelihood (chapter 4.3.3), Predicted Alternative Events (4.3.4), and Expected Consequences (4.3.6), especially once the actual outcome of the Action is observed.

The Planner will be told by the Sensor when an Action has been completed. If the Action was successful, the next concrete Action should be sent to the Output Generator. If it was not, the Plan will need to be recalculated by resending the Goal to the Planner. Of course, in an actual implementation, one could first try recalculating only within an abstract Action, or only part of the Plan, to save some time. The Emotion Monitor is also informed if the Action did not go as hoped, as part of implementing emotion causes Effects on Goals (chapter 4.3.5) and Sense Of Control (4.3.7).

### 5.2.7 Emotion Monitor

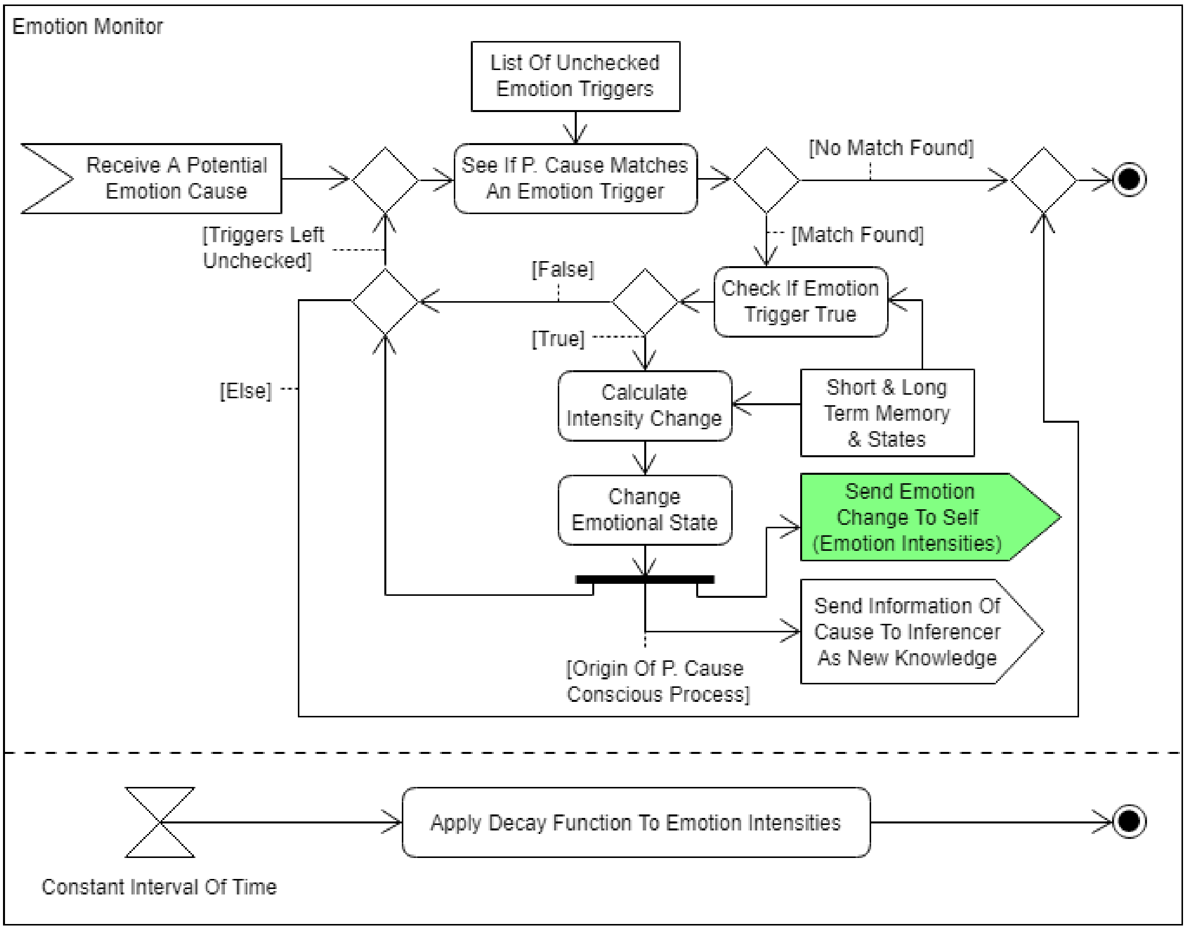


Figure 7: UML diagram of the Emotion Monitor.

The Emotion Monitor needs to adjust the Emotional State whenever an Emotion is triggered. So whenever a potential Emotion cause is sent to the Monitor, it will check if the cause is part of any Emotion triggers. If it is, it needs to use the Emotion cause together with the Long and Short Term Memories, as well as the Motivational and Emotional States, to see whether a trigger condition is true. To keep in line with the idea that emotions are a result of other processes, and not a cognitive process itself, the Standards are not involved in checking whether the trigger condition holds. Of course, if the emotion model requires it, the Inferencer can be called instead to check whether a trigger condition holds. If a trigger condition for an emotion is true, then the intensity change is calculated with the factors that can cause the emotion (although not necessarily the same ones as in the trigger condition). When this intensity change is applied, that change is sent on to the Emotion Monitor itself, as part of the Emotion Intensities cause (chapter 4.3.11). It's important to consider whether trigger conditions involving Emotion intensities might cause unwanted feedback loops. If it can be assumed the potential Emotion cause was known, this information should be sent to the Inferencer. Whether the cause is known, depends on its origin, and whether the emotion model considers that origin something the agent is conscious of or not.

For instance, the Sensor and Sensory Filter are probably subconscious processes. On the other hand, failing to construct a Plan is something the agent is probably aware of. Basically, each origin of a potential Emotion cause should have some translation into a fact, and whether that fact (linked to Emotion intensity increase or decrease) is sent along to the Inferencer (and subsequently, Memory) is dependent on the emotion model.

Something to consider here, is that a trigger condition cannot have more than one cause whose information is not stored in the Memory or a State somehow, or it will never trigger (within this implementation). So, for instance, the (usually) subconscious causes of a Motivational Value changing significantly and the agent displaying a particular Emotion (as part of the Body Signals cause) can not be in the same trigger condition. Of course, you can make a trigger condition of Emotion display while a Memory exists of the Motivational State changing recently, or just the Motivational State being a certain way while the Emotion was displayed. Simply put, the point at which the Emotion Monitor is prompted to consider the trigger conditions coincides with only one particular change: any other factors in the trigger condition should be stored in some manner. If the Long and Short Term Memory, Emotional State, and Motivational State are not enough to model the emotions the way that is desired, the Emotion Monitor should also keep track of other recent prompts. This does not really change how the Emotion Monitor works significantly, so I will leave it out of considerations from now on.

The Emotion Monitor should also apply an Emotion decay function to the intensity values of the Emotion State at some constant time interval. This function depends on the emotion model, and is not required to actually do anything, should the emotion model not have emotion decay.

### 5.2.8 Output Generator

The Output Generator receives the next Action to perform from the Planner and sends that Action to the Environment, as well as the current Emotion Signal, as per chapter 4.3.14 and chapter 4.4.4. For this implementation, I assume each Action is short enough that the Emotion Signalling only needs to be adjusted between actions. However, the Action is not necessarily sent as-is, and the Emotion Signal is not necessarily the one corresponding to the current Emotional State.

The outcome of an Action may depend on an agent's proficiency at something. An agent's proficiency can be generalized as a physical reality of the agent, meaning it would be part of the Motivational State. Other values in the Motivational State could also affect the outcome of an Action, as well as any other informational component of the agent. If one wants to keep calculation to a minimum, however, it is advisable to keep the outcome dependent only a limited amount of values that are easily retrieved. I'll therefore keep it to the Motivational and Emotional State in this implementation.

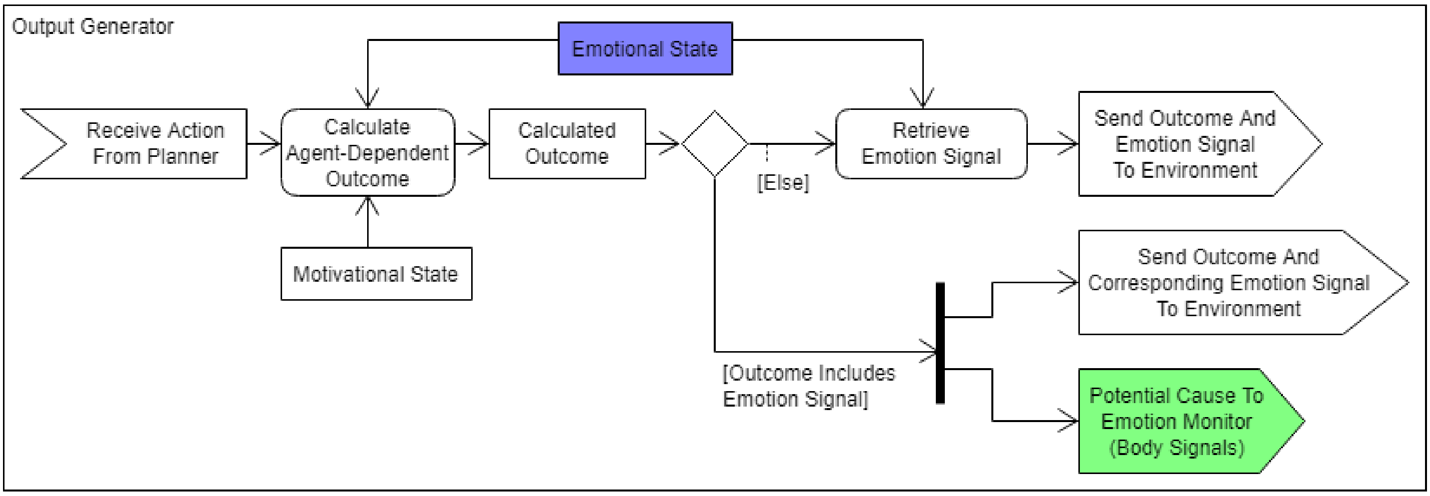


Figure 8: UML diagram of the Output Generator.

For an Action whose outcome is dependent on other values of an agent, multiple versions of it exist to be sent to the Environment. As an example, let's say the Action is "take a step". There are two versions of the Action that could be sent to the Environment: "take a step", and "trip over own feet". Another outcome may be "fall in a pothole", but that depends on the Environment and is not decided by the Output Generator. The Output Generator takes a look at the current skill level, other Motivational Values and the current Emotional State, and calculates which version of the Action to send on. This calculation may or may not include a random number.

The Emotion Signal that is sent on may be part of the Action. The example Action above would just send the Emotion Signal corresponding to the current Emotional State, but for the Action "take a step happily", the Emotion Signal is "Happy". Provided, of course, that the Action succeeds. Other versions of that Action that might be sent could be "take a step sadly, with faked happiness", or "take a step with neutral expression". Of course, varieties of Signals are only useful if other agents have the ability to reason with those Signals as input, so for every type of Emotion Signal, the Standards of at least one other agent should contain some sort of rule on how to reason with them.

As part of implementing the possible Emotion cause Body Signals of chapter 4.3.15, if the agent is sending an Emotion Signal that is incongruent with the current Emotional State, the Emotion Monitor is informed of this.

## 5.3 Emotion Model Variables and Examples

In order to give some examples of the emotion models that can be implemented using the provided implementation, I'll provide lists of all the variables that make up an emotion model, as referred to in 5.1 and 5.2. The first list will be about the Emotional State, the second about individual Emotions. I will indicate what its most basic version is, which essentially means not using that part of the emotion integration.

**Emotional State Variables:**

1. Which Emotions are part of the Emotional State? If no Emotions are given to the agent, you essentially have no Emotional State, which allows for a

non-emotional agent.

2. The translation method of the Emotional State. Given a list of all the chosen emotions, with their current intensities, that list needs to be translated to a set of facts that can be interpreted by the Inferencer. Standards rules and Memories should use this same notation to refer to the Emotional State. At its most basic, no translation method is given and the agent cannot reason about their Emotional State. This variable relates to chapter 4.4.1.

3. Reasoning about Emotion Signals from other agents. This takes the form of Standards rules that use knowledge of observed Emotion Signals. At its most basic, no such rules are part of an agent's Standards, and the agent cannot use observed Emotion Signals for anything. It's probably useful to consider variable 1 when writing these reasoning rules. This variable relates to chapter 4.4.4.

4. Assigning types to the Standards rules. As part of the next variable, every rule needs to have a type for the Inferencer, and a type for the Goal Manager. This type is allowed to be the same. As assigning types to begin with is part of implementing a BOID-like architecture, the most basic version of this can be only using types unrelated to emotions, for instance implementing BOID itself as closely as possible. This variable relates to chapter 4.4.5.

5. Sorting of Standards rule types per Emotional State. Namely, the sorting used by the Inferencer, and possibly separately by the Goal Manager, of the rules in Standards. For every type of Emotional State (presumably defined in the same way as the translation method of variable 2), there needs to be an associated sorting. At its most basic, every Emotional State has the exact some sorting, meaning the Emotional State has no influence on sorting. This variable relates to chapter 4.4.5.

6. The translation function of the current Emotional State to a change in the Emotion tag of a Memory, as used by the Memory Manager. At its most basic, no change to the Emotion Tag is added. This variable relates to chapter 4.4.2.

7. Related to the previous variable, the Emotion tags Memories start out with. At its most basic, the Emotion tags all have an intensity of zero, meaning they are effectively absent. This variable relates to chapter 4.4.2.

8. Whether the Emotion Monitor calls the Inferencer to check if an Emotion trigger is true, or whether it only uses Memories. This relates to how emotion is conceptualized in the model: as reaction to cognitive processes or a cognitive process itself. This part can't really be turned off, as it is part of a purely Emotion-related Process.

9. The effect each type of Emotional State has on the time-intensive likelihood calculations made by the Planner. In the most basic case, no Emotional State has any influence. This variable relates to chapter 4.4.7.

10. The effect of the highest intensity emotion on calculating the likelihood in a time-intensive manner, if the different Emotional States as used in the previous variable do not discern between high intensity and low intensity in some way. At its most basic, or if the previous variable already took care of that aspect, the intensity has no influence. This variable relates to chapter 4.4.7.

11. The likelihoods linked to the outcomes of the Actions, for all the different Emotional States, for situations when there is no time to calculate the likelihood of an Action outcome by the Planner. At its most basic, each Emotional State has the same likelihoods, meaning the Emotional State has no influence on the heuristic. This variable relates to chapter 4.4.7.

12. The threshold of allowed risk associated with each Emotional State, for determining when an Action is deemed "risky" by the Planner. At its most basic, each Emotional State has the same threshold, meaning the Emotional State has no influence on the heuristic. This variable relates to chapter 4.4.7.

13. For each cause origin, whether that origin is considered something the agent would be conscious of or not. This variable relates to chapter 4.4.1.

14. For each type of Emotional State, how it would influence the "Operations Allowed" calculations. This variable relates to chapter 4.4.6 and 4.4.7.

**Emotion specific variables:**

1. The trigger conditions. If this Emotion is a parent Emotion of another Emotion, make sure this Emotion always triggers when its child does.

2. Associated with the trigger conditions, the calculation of the intensity change upon being triggered. If this Emotion is a parent Emotion of another Emotion, make sure this Emotion does not increase less or decrease more than its child.

3. The decay function of the Emotion's intensity. If this Emotion is a parent Emotion of another Emotion, make sure this Emotion does not decay faster than its child.

4. For the Emotion Tags, how high the intensity of an Emotion tag for a particular Emotion needs to be before the Memory Manager flags a Short Term Memory as important (meaning it will get moved to the Long Term Memory). At its most basic, the variable can be set to unreachable, meaning that particular Emotion will never lead to a Memory being flagged as important. This variable relates to chapter 4.4.2.

5. The weights per Emotion in the functions used to sort the observed facts and Memories in the Sensory Filter, Inferencer and Planner. Particularly, the weight given to the Emotion tag associated with the Memory (or Memory/Standard related to an observed fact). At its most basic, the Emotion tags are not given any weight for any Emotion and have no influence on the outcome of the function. This variable relates to chapter 4.4.2 and 4.4.3.

Using these lists, I'll now provide an example of how to implement the Cathexis model as defined in the 1997 paper[50] into this version of MEIA, allowing for some differences resulting from Cathexis being a reactive agent (and other, non-emotion-related differences).

**Cathexis Emotional State:**

1. Emotions: Anger, Fear, Distress/Sadness, Enjoyment/Happiness, Disgust, and Surprise.

2. Emotional State translation: All different combinations of the Emotions with high and low intensities can be labelled as a different emotion blend. The paper doesn't go into detail.

3. Reasoning about Emotion Signals from other agents: Exactly what the Cathexis agent can infer is not defined in detail. As such, no reason not to include rules regarding Emotion Signals.

4. Assigning types to the Standards rules: Cathexis was written with a reactive agent in mind, where the emotion intensities contributed to the values given to the behaviours (the behaviour with the highest value was performed). So the emotion intensities directly contributed to sorting the behaviours. The sorting of the Standards that lead to adapting a Goal for the Goal Manager could then be given a similar sorting. The Cathexis paper also alludes to perceptual biases and selective filtering being a possible influence of emotion, but does not define it further, so for this example I'll leave Inferencer sorting unrelated to the Emotional State. The types for the Goal Manager would then be "made more likely by Emotional State X", for all possible X, and the other Processes would have just one type.

5. Sorting of Standards rule types per Emotional State: As mentioned above, only the Goal Manager has different types, already defined per Emotional State. The sorting for each Emotional State would then be the linked type first, all other types second (using a possibly random ordering).

6. The translation function of the current Emotional State to a change in the Emotion tag of a Memory: Not alluded to in the Cathexis paper, so no Emotion tags are added.

7. The Emotion tags Memories start out with: As above, no Emotion tags are added.

8. Emotion Monitor calls the Inferencer or only uses Memories: Only uses Memories, as Cathexis uses sensors that update the Emotion intensities according to what was observed, and no allusion is made to the intensity updaters making further inferences.

9. Effect each type of Emotional State has on the likelihood calculation: Not described beyond allusions to it in Cathexis, so no effect.

10. Effect of intensity on the likelihood calculation: Like above, no effect.

11. For each Emotional State, for each Action, for each possible outcome, the linked likelihoods: Like above, should all have the same likelihood.

12. For each Emotional State, the threshold of allowed risk: Like above, should all have the same threshold.

13. For each cause origin, whether that origin is considered something the agent would be conscious of or not: The emotion causes in Cathexis are defined as neural, sensorimotor, motivational, and cognitive. Neural and motivational are the causes originating from the Sensor (about the Motivational State) and the Emotion Monitor (about Emotion intensities). Sensorimotor causes come from the Output Generator (about Body Signals). Based on just the Cathexis paper, I can't divide all four categories into conscious and subconscious, but one could do so.

14. For each type of Emotional State, how it influences the "Operations Allowed" calculations: Time pressure is not alluded to in the Cathexis paper, so emotion should have no influence.

As the Cathexis paper doesn't define the particular causes and effects per emotion, I'll give a general list of emotion-specific variables:

1. The trigger conditions: The emotion causes in Cathexis are defined as neural, sensorimotor, motivational, and cognitive. Neural causes fall under Motivational Values, as do most motivational causes. The other motivational causes fall under Emotion Intensities. Sensorimotor falls under Body Signals. The cognitive emotion causes are those defined by Roseman [35], which I've incorporated as Attribution, Likelihood, Effect on Goals, Expected Consequence, Sense Of Control, and Changes in Evaluation. Non-cognitive causes are not alluded to being combined with any other causes, so I'll assume each cause is a trigger by itself.

The cognitive causes as defined by Roseman are always in combination with other causes: most combinations are a trigger for a different emotion. However, the basic emotions as defined by Roseman are not the same as those six defined in the Cathexis paper, so providing an exact list with no element of guesswork in it isn't possible. They both use "disgust" though, so I'll use that as an example. The trigger for "disgust" would consist of the following factors: a punishment-minimizing Goal was affected badly, which was not considered unlikely, with the affect attributed to the World, and which the agent had little Sense of Control over.

2. Intensity calculation: To quote the 1997 paper [50], "in Cathexis, the intensity of an emotion is affected by several factors, including the previous level of arousal for that emotion (which takes into account the mood), the contributions of each of the emotion elicitors for that particular emotion, and the interaction with other emotions (inhibitory and excitatory inputs)".

   The function given calculates intensity by taking all the different emotion elicitations together, whereas this implementation of MEIA calculates change in intensity per emotion change triggered. Simply put however, each emotion trigger should have an associated increase or decrease to its intensity, which can be made greater or smaller by the intensities of other Emotions, and the previous intensity.

3. The decay function: The Cathexis paper only mentions that each Emotion has their own decay function, which should, provided the Emotion is not triggered, decrease the intensity to zero after some amount of time.

4. How high the intensity an Emotion tag needs to be, to be an important Memory: Emotion tags are not implemented in Cathexis, so this variable is never used.

5. The sorting weights in the Sensory Filter, Inferencer and Planner: As above, Emotion tags are not implemented, so this variable is never used.

As you can see, Cathexis is fairly complex, and its implementation as described in the 1997 paper [50] was not completely set in stone, but a version of this Cathexis-inspired emotion model can be used to create an agent. It could be compared against agents with other emotion models, or an agent which uses no emotion model, to see which agent comes across as the most believable.

Now, let's consider a very simple emotion model in terms of how emotions are caused, with slightly more complex emotion effects:

1. Emotions: Happy, Sad, and Angry.

2. Emotional State translation: The Emotion with the highest intensity determines the State. If none of the intensities are above a certain limit, the State is Neutral. If intensities are equal, Angry is chosen over the other two, and Sad over Happy. Meaning there are four possible states.

3. Reasoning about Emotion Signals from other agents: The four possible States are used in some rules in the Standards.

4. Assigning types to the Standards rules: The types for the Inferencer are "Optimistic Assumptions", "Pessimistic Assumptions", and "General Beliefs". The types for the Goal Manager are "Beliefs", "Obligations", "Desires", and "Intentions".

5. Sorting of Standards rule types per Emotional State: When feeling Neutral, Happy, or Angry, the order in the Inferencer is "General Beliefs", "Optimistic Assumptions", and then "Pessimistic Assumptions", but when Sad, "Pessimistic Assumptions" are placed before "Optimistic Assumptions". In the Goal Manager, Happy and Neutral agents use the ordering "Beliefs", "Obligations", "Intentions", and then "Desires", while Angry agents become rule breakers, placing "Obligations" last, and Sad agents just want to feel better, and place "Desires" right after "Beliefs".

6. The translation function of the current Emotional State to a change in the Emotion tag of a Memory: If an Emotional State other than Neutral is the current State, the related Emotion tag receives a very small increase in intensity, up until some maximum.

7. The Emotion tags Memories start out with: There are a few object-archetypes and Plans with a tag for each Emotional State that is not Neutral.

8. Emotion Monitor calls the Inferencer or only uses Memories: Only uses Memories.

9. Effect each type of Emotional State has on the likelihood calculation: Each Action has a success outcome and a failure outcome. When the agent is Happy or Angry, each success outcome is considered more likely than when Neutral, and less likely when Sad.

10. Effect of intensity on the likelihood calculation: No effect.

11. For each Emotional State, for each Action, for each possible outcome, the linked likelihoods: The same idea as for the likelihood calculation.

12. For each Emotional State, the threshold of allowed risk: When Happy, the allowed risk threshold should be lower than for the other States, as the agent wants to keep being Happy. And for Sad, the threshold should be higher, as the agent doesn't feel like it has much to lose. Angry and Neutral's thresholds are in between.

13. For each cause origin, whether that origin is considered something the agent would be conscious of or not: In line with my personal intuitions, let's say anything sent by the Sensor, Sensory Filter, and Output Generator is subconscious, as well as Memory Associations. The agent is conscious of all other cause origins.

14. For each type of Emotional State, how it influences the "Operations Allowed" calculations: Happy and Sad both allow more time than Neutral (because there is more creative impulse, or because the agent "needs to get it right", respectively), and Angry allows less time, to simulate hot-headedness and jumping to conclusions.

**Simple Happy Emotion:**

1. The trigger conditions (with intensity calculations): A Goal was fulfilled (large increase), or a Happy-tagged Memory was remembered (small increase), or the Body Signal for Happy was performed (small increase), or another agent was observed to be Happy (small increase), or there was an

increase in Sadness or Anger (small decrease, relative to increase of other Emotion).

2. Intensity calculation: Already defined above.

3. The decay function: The intensity of this Emotion is slightly adjusted to be closer to a value that is just below the value required to change the State from Neutral to Happy (provided the other Emotions are not higher in intensity). In other words, the agent is Neutral with no stimuli, but can easily become Happy.

4. How high intensity of an Emotion tag needs to be, to be an important Memory: Happiness was said to increase creativity in part because more connections were made between less related concepts [16], so let's say that it only needs to be above zero, to hopefully mimic these many connections.

5. The sorting weights in the Sensory Filter, Inferencer and Planner: Based on the above variable, as long as the current Emotional State is Happy, it should be easier recalling many recently observed and learned facts. Or at least, we can test if it does. Being Happy should make it easier to recall many things not necessarily relevant to the current Plan, to mimic creativity, so being Happy places Happy tagged Memories quite high in the sorting.

**Simple Angry Emotion:**

1. The trigger conditions (with intensity calculations): An Action has failed and the agent knows it was because of another agent (large increase), or an Angry-tagged Memory was remembered (small increase), or the Body Signal for Angry was performed (small increase).

2. Intensity calculation: Already defined above.

3. The decay function: The intensity of this Emotion is slightly adjusted to be closer to zero, meaning the agent only gets and stays Angry if a reason has recently occurred.

4. How high intensity of an Emotion tag needs to be, to be an important Memory: Let's say the agent holds a grudge, so any Angry Emotion tag flags a memory as important.

5. The sorting weights in the Sensory Filter, Inferencer and Planner: It should be lower than that of the Happy Emotion (to make Happy stand out), but it should still make Angry tagged Memories rank a little higher when the Emotional State is Angry.

**Simple Sad Emotion:**

1. The trigger conditions (with intensity calculations): An Action has failed and the agent doesn't ascribe this failing to another agent (large increase), or a Sad-tagged Memory was remembered (small increase), or the Body Signal for Sad was performed (small increase), or another agent was observed to be Sad (small increase), or there was an increase in Happiness (small decrease, relative to the increase in Happiness).
2. Intensity calculation: Already defined above.
3. The decay function: The intensity of this Emotion is slightly adjusted to be closer to zero, meaning the agent only gets and stays Sad when a reason to be has recently occurred.
4. How high intensity of an Emotion tag needs to be, to be an important Memory: To be different from the other Emotions, a Sad tag only flags a Memory as important if it is really high.
5. The sorting weights in the Sensory Filter, Inferencer and Planner: It should be lower than that of the Happy Emotion (to make Happy stand out), but it should still make Sad tagged Memories rank a little higher when the Emotional State is Sad.

Hopefully, this simpler example than the Cathexis model was illustrative in showing how to create an emotion model in this implementation. These examples were made with comparing multiple emotion models in mind, which is why one was an existing model, and the other a model with all the types of emotion integration. However, that was not the only use of MEIA that I proposed. Another is using its modular nature to create character specific emotion models for creative purposes. For instance, a lot of characterization is implied by what causes a character to feel certain emotions. To illustrate, I'll give an example of some hypothetical story.

Let's say we have a group of characters who are soldiers on a mission, and we give the corresponding agents the same emotion model to start with. In this common emotion model, seeing a toy that can be inferred to have been lost by its owner causes a mild increase in sadness. Already, we are relying on the assumption that abandoned toys in war imply sad situations to most people, and do not attempt to model why. Rather, we assume the story's audience will conclude this for themselves.

Now let's say we change one agent's emotion model to have a far stronger increase in sadness. This could imply all sorts of things about that character: maybe the character is reminded of his or her own children? Agent could get angry at seeing the doll as well, suggesting they are quick to look for someone to blame and take revenge on rather than feel sad, or maybe they refuse to just accept sad events. And having one agent that doesn't feel sad at all invites the audience to speculate why, exactly, that character is defying expectations. To refer back to the train track analogy used earlier, every difference in an emotional model can be used by writers to suggest a way to build that train track of narrative realism, including not providing a piece of the train track the audience was expecting, forcing them to find another piece.

One can imagine that doing so correctly requires a certain amount of insight into how people work, best to left to writers rather than those who are exclusively programmers. For that, it's useful to be able to hand a writer a list of questions, like the variable list I provided, rather than having to explain the ins and outs of how a character is programmed.

## 5.4 Emergent Behaviour Examples

Having described how to create different emotion models for my example architecture, I'll conclude by giving some examples of how different behaviours could result from having different emotion models (including a model with no emotions). For each of the emotion effects I've introduced, I'll provide at least one example to illustrate how an agent without emotions would behave differently from agents who do have the emotion effects implemented. Of course, it should be acknowledged that an agent without emotion could of course be programmed in such a way that, in the specific examples, it displays the same behaviour as the emotional agents I'll describe. However, the point is to illustrate that adding the emotions is a way to cause a particular type of behaviour in many different situations. To make the examples more easy to read, I'll refer to the non-emotional agent as N, and the emotional agents as E.

### 5.4.1 Emotions as Information

As described in chapter 4.4.1, the Emotional State can be used as information when making decisions. What this essentially means is that all the different emotion causes are implicitly a part of making particular decisions, without the agent necessarily even retaining the knowledge of all those causes occurring. And, like the example given in chapter 4.4.1, knowledge of the causes of the Emotional State may even be inaccurate, yet still used in making decisions. Concrete implementation of this effect takes the form of reasoning rules in the Standards that use the Emotional State as a factor. In other words, reasoning rules that treat the different potential Emotional States as facts in the knowledge base, which can be true or false, and can be used to derive further knowledge.

In this example, the agent is deciding whether buying a house it is viewing would be a good idea. Both N and E would think buying a house is a good idea if its cost is equal or lower than its value. However, E also requires that the current event of viewing the house does not make it feel a negative emotion. To avoid having to define what an event is, to keep this example simple, these rules will assume the question is only asked while viewing the house, and any cause the agent may know about takes the form of an object, not an event. In other words, this is the reasoning rule N uses, in Prolog:

```
goodBuy(House) :- lowerOrEqualCost(House).
```

And these are E's:

```
goodBuy(House) :- lowerOrEqual(HouseCost, HouseValue),
currentEmotion(Emotion), \+negativeEmotion(Emotion).

goodBuy(House) :- lowerOrEqual(HouseCost, HouseValue),
currentEmotion(Emotion), negativeEmotion(Emotion),
causeCurrentEmotion(Cause), Cause \= House.
```

In other words, E thinks the house is a good buy if E is currently feeling a positive emotion, or if the negative emotion has a known cause that is not the house. The agents are then exposed to the following environment:

```
lowerOrEqualCost(house1).
negativeEmotion(fear).
currentEmotion(happy).
```

When asked if house1 is a good buy, both N and E will conclude that it is. If we change the `currentEmotion(happy)` to `currentEmotion(fear)`, however, E will no longer conclude that. However, if E somehow gets the additional knowledge that `causeCurrentEmotion(clownPainting)`, E will once again conclude that house1 is a good buy. This knowledge could have been added by the Emotion Monitor when the fear emotion got triggered by the clown painting, or, like in the example given in 4.4.1, another agent could have pointed it out to E. Or possibly, E is capable of introspection.

Either way, this allows for E to take into account many different factors when making this decision, without the designers needing to explicitly couple these factors to house buying. If the fear E is feeling is due to many small factors, like multiple slightly intimidating neighbours, and speeding cars and aggressive dogs encountered in the vicinity of the house, then buying the house might indeed be a bad idea. However, these causes of the fear on their own would not make for a convincing argument not to buy the house. If we want N to arrive to a similar outcome, we have to explicitly remember all these causes, and add extra rules for when the agent should or should not buy the house. If we then want N and E to be able to decide whether or not to stay at a party they are currently attending, E only requires a rule that it will keep doing what it's doing, as long as it's not feeling a negative emotion, whereas N needs it all written out explicitly once again. If we give N a similar rule, like "are there at least X amount of negative signs" instead of "am I feeling a negative emotion", then we have basically given N an Emotional State that requires N to remember all causes and reason about them for every decision.

To recap, using Emotion as Information allows agents to have a quick way of considering the broader context of their situation in their direct reasoning. In the case of the agent being afraid due to its fear of clowns, this quick way may turn out to have inaccurate results, but it may turn out to be better to have a generally applicable quick way and correct later than to have a costly, less applicable, but more accurate way.

### 5.4.2 Memory

As described in chapter 4.4.2, implementing the effects of emotion on memory consists of two adjustments. The first is that memories gained while in an emotional state are more likely to be remembered. This is taken care of in the Memory Manager, where a memory does not get discarded if the associated emotion tag is high enough. For this example, N and E have some very bad memories related to dogs. Therefore, E has a high "Fear" tag associated with dogs in general. N and E both work behind the counter of a fast food restaurant. Customers, once they leave the restaurant, are no longer relevant to any Goals, so when the memories related to these customers are removed from the Short Term Memory, they are not moved to the Long Term Memory. However, if a customer has a dog, the memory of that dog has a high "Fear" tag, and so does get transferred to the Long Term Memory in E's case. If someone were to ask N and E whether a person with a brown Labrador visited the restaurant, N would not be able to remember, but E would, and would therefore be able to help the asker. In this case, E being able to remember the dog because of E's phobia could be contrasted against E's co-workers, who don't remember the dog (despite being implied to also be emotional agents), giving E some depth of character.

Unrelated to behavioural changes in the agent, but nevertheless interesting to consider, is the fact that the high emotion tag causing memories to be kept takes some pressure off whatever set of rules leads to a memory being flagged as important, in terms of how completely these rules have to encompass all situations. For those vague yet important goals like "staying alive", a fear tag makes a good proxy for checking whether a memory has the potential to be relevant to that goal (or at least one aspect of it), meaning the person writing the rules has to worry less about edge-cases getting forgotten.

The other adjustment required for implementing the effects of emotion on memory takes place everywhere the knowledge an agent has is used for reasoning: it adjusts the order in which the knowledge is processed. In implementations with time limits, or where, out of the best options, the first option found is chosen, this means the agent might display different behaviour than if the knowledge had been sorted otherwise.

I will give an example in which both the Inferencer and Planner are involved. Let's say N and E have the goal to dance the waltz with someone. In the Prolog-inspired terminology I use in my implementation, the Goal is to have the fact `waltzedWith(Person)` be true. The knowledge base N and E share is as follows:

```
person(alice).
person(bob).
person(carol).
person(dan).
canFollowWaltz(alice).
canFollowWaltz(bob).
canFollowWaltz(carol).
canLeadWaltz(dan).
```

They also have access to four Actions. The first is "Waltz As Lead", with the prerequisite `canFollowWaltz(Person)`, and with the single outcome `waltzedWith(Person)` (let's assume that other agents are always up for a little dancing). The second is "Waltz As Follow", with `canLeadWaltz(Person)` as the requisite and `waltzedWith(Person)` as the outcome. The other two are "Teach Waltz (Follow)" and "Teach Waltz(Lead)", both with the prerequisite "person(Person)", and outcomes `canFollowWaltz(Person)` and `canLeadWaltz(Person)` respectively.

E attaches a "Happy" tag to all facts pertaining Carol and Dan, and an "Angry" tag to facts pertaining Bob. E also has a "Happy" association with the Action "Waltz As Lead". In this particular example, E's Emotional State is "Happy". Let's assume that there is only enough time to come up with one Plan.

When the dancing Goal is first created in the Goal Manager (as illustrated in Figure 5), the Goal Manager queries the Inferencer on whether the Goal is fulfilled, meaning it sends these facts to the Inferencer. The Inferencer won't be able to find the fact in the knowledge base, and so return the answer `false`, so the Goal is sent to the Planner. The Planner, as illustrated in Figure 6, will then run a backwards

A* search with the desired world state of `waltzedWith(Person)`. For N, the two Actions that can connect with that world state due to their outcomes ("Waltz As Lead" and "Waltz As Follow"), have the exact same cost, so it decides on which Action to explore first in an arbitrary fashion. For E however, the "Happy" tag of "Waltz As Lead" makes the cost of that Action lower, so it will always attempt to make a path using that one first.

The prerequisite of the "Waltz As Lead" Action is that a potential partner can actually waltz as a follow, `canFollowWaltz(Person)`. The Inferencer is then called to see whether there is a person for whom this is `true`. Let's say the Inferencer only gets allotted time enough to find two people. In N's case, that would be Alice and Bob, if the knowledge base is ordered as above. The Planner would find no objections to dancing with Alice, so that's what N will do. Had N decided to waltz as a follow, N would have danced with Dan. E however, will think of Carol and Bob, because Coral has a "Happy" association, which matches E's Emotional State, and Bob, unlike Alice, has an emotion tag. So E will dance with Coral when Happy, which makes intuitive sense.

But what if E had been "Angry"? Then E would have danced with Bob, who is associated with "Anger". Now, had the goal been "I need to do a sporting activity", and Bob had been "boxing", this would have made intuitive sense. But people that one associates with anger are usually not people one wants to dance with. So let's say that the outcome `waltzedWith(bob)` would have gotten registered as an outcome in conflict with a more important Goal, when the A* algorithm tried to calculate the cost of the outcome, as part of calculating the cost of the Action "Waltz As Lead" with Bob. Then E would have danced with Carol again. But, if Carol had not existed, E would think another Action was required (because Alice was not remembered).

At this point, E's Planner has to either try the Action "Waltz As Follow" instead, or see if the Action "Teach Waltz (Follow)" can be performed previous to "Waltz As Lead". This would depend on the heuristic the A* algorithm uses. Let's say, for the sake of the example, the heuristic was learned in an environment where E usually had to teach people how to waltz before they could start dancing. In that case, trying to "Waltz As Follow" instead does not seem like the lower cost path. So E's Inferencer attempts to find a `person(Person)`, the prerequisite of the teaching Action, and finds Bob and Dan (as Carol doesn't exist now). Bob has already been discounted, so E tries to teach Dan how to be the follow in a waltz.

This does not seem like the rational thing to do, admittedly, but N would actually have tried the same thing if Alice had not existed: the Planner would have discounted Bob, and not remembered Carol, and then attempted to teach Dan how to waltz as a follow. When in a hurry, both agents just go for the first thing they can think of, not giving themselves time to truly consider the best option. However, E would do so in a manner that seems consistent over time and in different scenarios, with only the extra cost of sorting the knowledge base (not inconsiderable, admittedly) and adjusting the cost heuristic based on the current Emotional State (one operation extra for each Action). N's decisions would seem arbitrary.

### 5.4.3 Emotion as Focus

Like the effects on memory, the effects of emotion on focus are primarily implemented through sorting of information, in this case in the Sensory Filter. The Sensory Filter sends the information received through the Sensors on to the Inferencer to be evaluated. If an implementation is made in such a manner that all information is eventually sent through, the order can decide whether certain default logic rules are, or are not, applicable when trying to discover the implications of the new information, meaning the order decides which implications are added to the knowledge base. If only the first part of the list of observed information is sent on, the order of that list decides which knowledge gets added or updated in the knowledge base in the first place. That having a different knowledge base can lead to different behaviour seems rather self-explanatory.

In fact, the situations in which this would change the behaviour are similar to the "remembered dog because agent is afraid of dogs" example. Instead of remembering, the change would be registering the information in the first place. So if N and E had seen the dog on the street, N and their companions would not have noticed the dog at all, because it had no relevance to their goals, but E would have, because the "Fear" tag would place the observation of the dog higher in E's Sensory Filter list.

Also like the memory example, the emotion tags placing the associated information higher in the relevance list can take the place of having to thoroughly check whether a given observation is relevant to broad but important goals like "staying alive", "being happy", or "retain place in a social hierarchy".

### 5.4.4 Emotions for Communication

The implementation of the first part of this effect, namely receiving communication, is actually very similar to that of Emotion as Information. This is not that surprising. After all, using the emotion signals of another agent is just asking "what kind of broader context is the other agent experiencing?" instead of "what kind of broader context am I experiencing?". We can therefore use the exact same example, just with `currentEmotion` being replaced with `realEstateAgentEmotionSignal`. Suddenly, the rule that it is not a good idea to buy a house where one feels afraid becomes a rule that one shouldn't buy a house from a real estate agent that looks afraid, implicitly because that could mean there is reason for the real estate to be afraid, relating to the house (like the neighbourhood) or the showing of the house (like the possibility that the buyer discovers the house is rotting).

Sending emotion signals does not affect behaviour of agents directly, unless the agents have reasoning rules pertaining to what effect emotion signals have on other agents. Concrete examples of this effect very quickly turn into examples of how to model the mind of another agent, which is beyond the scope of this paper. As such, I'll give an example using a very simple agent as the one being modelled by E: that of a very simple cat, who copies the perceived Emotional State of its owner.

E has two Goals: to keep the cat in the Emotional State "Happy", and to watch a film that makes E have the Emotional State "Sad" (prioritized in that order). The first Goal's set of facts to make true consists of `catEmotionalState(happy)`, and the condition under which it can be removed is `false` (in other words, the Goal should never be removed). The second goal has the same fact to make true as the one that allows it to be removed, namely `haveWatchedFilm(film)`. Currently, both E and the cat have the Emotional State "Happy", so the first Goal is inactive. Therefore, the second Goal gets sent to the Planner. I will assume the reader has Figure 6, the UML diagram of the Planner, close at hand for the following part of the explanation.

The desired world state generated by the Planner is `haveWatched(film)`. E does not have access to any predefined plans, and has unlimited time to make a decision, so it will then perform the modified backwards search A* algorithm. Its starting world state is `catEmotionalState(happy)`, `ownEmotionalState(happy)`, and `emotionSignal(happy)`. E has two Actions available to it: "Watch Film" and "Watch Film While Signalling Happiness". "Watch Film" removes `ownEmotionalState(Current)` and `emotionSignal(Current)` (the Action is not dependant on the current Emotional State), and adds `ownEmotionalState(sad)`, `emotionSignal(sad)`, and `haveWatched(film)` to the world state, as its only possible outcome. "Watch Film While Signalling Happiness" does almost the same thing, but instead adds `emotionSignal(happy)`. The two "Watch..." Actions result in the necessary world state (`haveWatched(film)`), so both are considered.

When calculating the Action cost for both of these Actions, and subsequently the Action outcome costs, the likelihood of the desired outcomes are a 100%. For the "Watch Film" Action, it will find that the resultant world state conflicts with the desired world state of the more important `catEmotionalState(happy)` Goal. In this example, the agent is able to find this conflict (with a call to the Inferencer), as it has unlimited time. I will later examine what would happen if it did not have unlimited time. As the chance of the undesirable outcome is 100%, and thus above the allowed risk threshold (more on that later as well), the Action is removed from consideration. "Watch Film While Signalling Happiness" does not have such a conflict, making it the only viable Action to take in this case. It also connects the starting world state and the Goal world state perfectly, so that is what the agent will do.

### 5.4.5 Goal Managing and Priority

When it comes to implementing an agent with multiple goals, these goals need to be ranked somehow. If the way to achieve one goal conflicts with another goal, which goal is more important? And which of the goals should an agent attempt to achieve first? In this implementation, I assume these questions can be answered as though they are the same. It would be possible to make a ranked list of all possible goals and use that. However, consider the following goals: "minimizing money spent" and "being happy". While people would probably consider being happy more important than money, the first goal was actually created as part of fulfilling the goal "long-term survival", and N (who in this example only has an Emotional State that has causes, but doesn't have any effects) would therefore consider saving money more important than the "being happy" goal. In most Emotional States, so would E. However, when E is in the state "Sad", "long-term survival" type goals get lower in the ranking than "being happy". Let's say N and E are playing a technically free mobile game to fulfil or maintain the "being happy" goal. They lose the level, and are given the option to pay money to get a second chance. N, and E in most moods, would never take the Action "Pay For A Second Chance", as it conflicts with the prioritized "minimizing money spent" goal. But a "Sad" E would.

#### 5.4.6 Time Pressure and Creativity

In the implementation I've outlined with UML diagrams, the effects of emotions on Time Pressure and Creativity relies entirely on the "Operations Allowed" factor. To illustrate how N and E might display different behaviours due to allowing themselves a different number of operations, I'll give an example situation for each of the occurrences of the "Operations Allowed" factor in the UML diagrams.

In the Sensory Filter (Figure 2), the same example as in Emotion as Focus can be used, only instead of information being registered or not registered due to its placement in the ordering, it is registered or not registered due to the amount of information sent along being different. So if, for instance, the house keys are placed 8th on the Sensory Filter's list, it only gets registered if the agent allows itself 8 or more observations to be sent along to the Inferencer. N allows that, because the estimated time before N and E have to go to work allows that, but a "Happy" E allows itself less observations (because there is no need to waste cognitive energy if everything is fine), and so does not notice the house keys being left behind.

In Figure 3, of the Inferencer, there are multiple places where the Operations Allowed factor has influence, but they all amount to the same thing: how much of the knowledge base is it allowed to use (and possibly expand) before it stops itself? To refer back to the cat example from Emotion as Communication: if the agent does not feel pressure (does not allow itself much time) to examine the implications of watching the sad film, it will not derive that the cat will be sad, and will just watch the sad film without pretending to be happy.

Changing the Operations Allowed in the Planner (Figure 6) can make an agent just choose a predefined Plan, or the first discovered Plan, without examining whether any other Plans can be made. For instance, let's say N and E have a predefined Plan to walk to the refrigerator to get a drink. They have recently broken their legs, so the cost of doing that to fulfil the Goal "get a drink" is very high currently. If the agents allow themselves time to make another Plan (which, purely based on the amount of time they have, they can), the new Plan would almost certainly have a lower cost. So N comes up with the Plan to use crutches instead. However, an E that is "Happy", and does not judge getting a drink as something to pay attention to (low creativity required), will just go with the predefined Plan (which turns out to be rather painful). The effects of emotions are sometimes most obvious when they cause behaviour that is less than ideal.

The Planner has two other instances where the Operations Allowed play a role. One, regarding checking for conflict between an Action outcome and a Goal, was already discussed in the Emotion as Communication example. The other decides whether to use the predefined likelihood for the outcome or to calculate it. This one is mostly there for consistency's sake. It made sense to me that, given enough time, you would want an agent to use more accurate methods to estimate how likely an outcome is (in the same way that the Inferencer is allowed to perform more operations). It would be interesting to see whether it would have an observable, consistent effect, if the decision for one method or the other was influenced by the Emotional State.

#### 5.4.7 Risk Taking and Outcome Assessment

To give an example on how to implement the effect of emotion on risk taking and outcome assessment, I'll refer back to the example from Emotions as Communication of the cat that should not see its owner be sad. Previously, the Action "Watch Film While Signalling Happiness" had only one outcome. Let's add another outcome, where the agent does not succeed, with a 50% chance of happening. The effect of that outcome is that `ownEmotionalState(Current)` and `emotionSignal(Current)` get removed from the world state, and that `ownEmotionalState(sad)`, `emotionSignal(sad)`, and `haveWatched(film)` are added to the world state. In other words, the same outcome of the "Watch Film" Action.

Now, when the Planner checks if the risk of that outcome exceeds the threshold, how high the threshold is could decide whether or not the N and E would allow themselves to use the Action in a Plan. Let's say the usual threshold is 50%: anything at or above that is too risky. So for N, the risk is not worth it. However, for E, this depends on the Emotional State. In a "Happy" state, the risk threshold depends on whether the outcome would change the Emotional State. As it happens, it would (even without the cat's influence, incidentally), so the threshold gets lower, meaning E definitely wouldn't risk it. But a "Angry" E takes far more risks, regardless of consequences, and has a threshold of 60%. So an "Angry" E would risk it. However, a 50% chance of an undesirable outcome would significantly increase the cost from the initially estimated cost, so given the option, an "Angry" E might still prefer another Action that was previously considered too costly instead.

With similar behavioural effects, the calculation of the likelihood in the first place can also be affected, making a "Happy" E judge the likelihood of the negative outcome to be lower (possibly even below the lowered threshold), or a "Sad" E to judge the likelihood higher (while also having a higher risk threshold). In other words, the two factors play off each other. In this example, for "Happy" and "Sad", the two factors temper each other. For "Angry" and "Afraid", the two factors could strengthen each other: "Angry" E's judge bad outcomes as less likely, while also allowing more risk, and "Afraid" E's judge bad outcomes as more likely, while allowing less risk.

# 6 Future Work

I have suggested ways to use and implement MEIA, but an obvious next step to this would be an actual implementation. I have actually started to make an implementation like the one suggested in chapter 5. At first, I wanted to find an implementation of a BDI agent that was already written and adjust it so that it used the suggested emotion integration. However, the requirement that I had to be able to change the order in which information was used by the Inferencer, Planner and Goal Manager (or analogous parts of an agent), and have a knowledge base that is allowed to be inconsistent (due to different Emotional States causing different subjective judgements), meant I couldn't find an agent implementation that matched my exact needs. And that meant I had to adjust the program at its core functionality. Namely, the way it handled its knowledge base, and how it applied reasoning.

In order to make such adjustments well, you have to be quite familiar with how a program works. At the time, I did not want to invest a lot of time into learning about any particular agent implementation, only to possibly find out the adjustments I implemented had made other parts of the agent malfunction, or that the adjustments needed required so many changes, I would spend more time trying to keep the rest of the program working than on actually implementing the emotion integration. So, I figured I would make a "simple" agent myself.

Of course, I did not want to reinvent the wheel entirely, so I used a GOAP library and a Prolog interpreter library. I spent quite a bit of time wrangling with the Prolog interpreter, in order to get it to work with dynamic knowledge bases, and adjusting the order of the knowledge base, and dealing with default rules. The solutions I had found to these problem worked (insofar as I have been able to test), but they were mostly a matter of making the Prolog interpreter register everything it was doing, and then derive the needed information from that afterwards, or just restarting the Prolog interpreter with a slightly different knowledge base each time. Undoubtedly, I could have done these things within the Prolog interpreter for a more elegant solution, but then I ran into more roadblocks.

I realised that the best GOAP library I could find did not allow for abstract actions, or desired world states with negative statements in them. I realised I would have to spend even more time on adjusting code made by other people, instead of working on the actual emotion integration. I also realised that I was going to need significantly more time to produce a program than I had, even if I used a lot of sloppy code that would make me unhappy with the final product. So, I came to the conclusion that trying to implement a general agent architecture by myself, and then also add some extras to it, in just a few months, was probably overambitious. Hence, I gave a description of what I had tried to implement instead.

So an implementation has not been made yet, and it would be very useful to have one, in order to see if using MEIA adds anything in regard to scientific research or creative endeavours. It would also be very interesting to see the emotion integration applied to any existing agent implementation, especially ones that have already been proven to be relatively believable. That would really allow one to test whether emotion integration improves believability or not. And if it does, if it does so significantly enough to be worth the cost of implementing.

It would also be interesting to find out if people in the creative industry find MEIA a useful framework for creating emotional agents for character driven narratives, or stories in general. This would require an implementation of MEIA that allowed someone to create agents with different emotion models on the fly, with a sufficiently intuitive User Interface instead of code.

Beyond implementing MEIA and researching how useful it is, it's important to evaluate whether any aspects of MEIA could be modularized even further. After all, if it turns out an emotion model uses only a part of some aspect that I've defined as one all-or-nothing component, that component needs to be broken down in order for that emotion model to be implemented in MEIA. With the causes of emotions, I don't think it can be broken down even further, as each cause is just one particular fact or step in a process. With the effects however, as the reasons for these effects become clearer, a way to modularize them even further might become clear. Additionally, new causes and effects may be discovered as well.

The other way around should not pose an issue: if an emotion model considers parts I've modularized as parts of the same whole, one can simply implement all those modularized parts to arrive at the same model. Of course, if the implementation of one modularized part relies on how another modularized part was (or was not) implemented, that would need to be taken into account when using MEIA to compare such a model to other models. For instance, how the Memory Association cause is implemented and how the Memory effect is implemented may rely on each other, in the sense they can both use the same emotion tag. So it would bear examining whether that negatively impacts the ability to compare models. Another thing to research would be whether implementing an emotion model where parts I've modularized are conceptualized as one part in MEIA, while trying to keep it as adjustable as possible, is somehow inferior to implementing it outside MEIA. For instance, maybe the implementation in MEIA is slower.

In terms of expanding MEIA, one obvious factor that I haven't discussed is learning. While the agent as described in the implementation of chapter 5 does learn new facts about its environment, and potentially has Emotion tags that are adjusted over time, I have not really discussed how emotion could influence learning new Plans or Standards rules. As emotions seem to have quite a bit of influence on memory recall, which is related to learning, it would surprise me if emotions had no further influences on learning. Aside from learning, new developments in emotion research might mean more causes and effects could become apparent, and any new insight could inspire new conceptualizations of emotions.

# 7 Conclusion

In this thesis, I have provided a general agent architecture, with a modular way of integrating emotion into such an architecture. This Modular Emotion Integration Architecture, or MEIA, can be used to create agents with the same general architecture, but different emotion models, so that different emotion models can be compared for research purposes. Additionally, MEIA can be used to create agents with personalized emotion models, so that in creative endeavours, agents can be given different personalities, or have emotion models that are pragmatic (focus only on resulting in desired behaviour) rather than accurate. So, I have answered my research questions as follows:

- **In what way does emotion influence the decision making progress, according to the current psychological theories and studies (beyond being a value to minimize or maximize)?** Emotion seems to influence the decision making process in several, possibly related, ways. The influences I've come across are on the information used in cognition, on which and how memory is recalled, on the focus cognitive input gets, on how communication between agents is conducted, on which type of goals are given priority, on how much time and cognitive energy is spent on problem solving, and on how risks and outcomes are evaluated.
- **How do I model these influences in an emotional computational agent for the purpose of creating more believable computational agents?**
  - **What is it that makes behaviour seem more believable, accord-**

**ing to current literature?** The information the fiction provides not causing an unsolvable contradiction with the other information available, including context outside the piece of fiction itself. The information available includes preconceived notions about how the world in general, or the genre of fiction, usually functions. As such, the information provided by the agent's actions does not have to be only or completely in line with what they would do in reality, but also or instead with what is seen as consistent with cultural and fiction-contextual information.

– **How do I model an emotional computational agent (so that I can integrate the emotion influences into it)?**

  * **How are computational agents in general modelled?** There are logical, reactive, BDI and multi-layered types of agent architectures. Logical agents are unsuited for games and simulations, but the other types of agents are used for those purposes, and might therefore benefit from being extended with further emotion integration. I identified different components which these types of agent architectures have in common, which are all described in chapter 3.

  * **What general emotion theories are suited for emotional computational agent modelling?** In regard to the purpose and source of emotions, there are appraisal theories, evolutionary theories, constructionist theories, and potentially others, which are sometimes but not always compatible. They also have differing sets of the "basic" emotions, as well as different ideas on what makes something an emotion to begin with, and whether there is a difference between an emotion and a mood. I identified these differences as several meta-structures, emotion causes, and emotion effects (the last of which I identified as part of answering the first of my main research questions).

  * **How have others modelled emotional computational agents?** There have been many emotional architectures proposed by others, most based on some type of appraisal theory, like FLAME, FAtiMA, EMA, and Cathexis. The last two I've examined in greater detail.

Together with the observations I made while answering the previous question, I described the components which I believe make up these models in chapter 4.

Chapter 3 and 4 together describe MEIA, an architecture that allows any emotional agent model to be constructed that is needed. In other words, I've answered this question by outlining all the ways it could be answered. Chapter 5 gave some concrete examples, to illustrate how to use MEIA in practice, and ended with some examples of how behaviour is changed when implementing these examples.

– **How do I account for all the different views in psychology within one emotional computational agent model?** How to model the influences exactly depends on which broader emotional agent architecture was chosen, and which constraints need to be worked with. But for each of the influences, I've described which general agent architecture component would be affected and how it interacts with the broader emotion model chosen. I have primarily described this in chapter 4.4, with more concrete examples in chapter 5.

Overall, I would say I have laid the groundwork for what I attempted to do with this thesis. Namely, to find a way to integrate emotion further into the decision making process, which I have done when examining the effects emotions can have, and how to implement those. Additionally, to provide an emotion architecture which assumes as little as possible about the emotion model used, which I have done by making the emotion architecture as modular as possible. However, while I think I have given enough arguments for MEIA being useful for research and creative purposes, the only way to really test that, is to use it. And MEIA has not been implemented yet.

Aside from an implementation being needed, new research or insight on emotion could always lead to MEIA needing to be expanded, as the causes and effects I've listed are not necessarily exhaustive. It could turn out that the causes or effects can be modularized further, in which case MEIA in its current state is restricted in the amount of emotion models it allows. It may also turn out that a modular approach has some unforeseen negative side effects when trying to implement an emotion model that is far less modular in design, for instance due to what I consider different parts being strongly dependant on each other.

So, I would say that, much like research into emotions and their effects on cognition, there is a lot of work still to be done when it comes to emotional computational agent architectures, but I believe MEIA is a step in the right direction.

# References


[1] A. Acosta, R.B. Adams (Jr.), D.N. Albohn, E.S. Allard, T. Beek, et al. *Registered Replication Report: Strack, Martin, & Stepper (1988).* Perspectives on Psychological Science, Volume 11, Issue 6, Pages 917 - 928, 2016.

[2] Pamela K. Adelmann and R B. Zajonc. *Facial efference and the experience of emotion.* Annual Review of Psychology, Volume 40, Pages 249-280, 1989.

[3] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (5th ed.).* Arlington, VA: American Psychiatric Publishing, 2013.

[4] Lisa F. Barrett. *Solving the Emotion Paradox: Categorization and the Experience of Emotion.* Personality and social psychology review 10.1, Pages 20-46, 2006.

[5] Henning Boecker, Till Sprenger, Mary E. Spilker, Gjermund Henriksen, Marcus Koppenhoefer, Klaus J. Wagner, Michael Valet, Achim Berthele, and Thomas R. Tolle. *The Runner's High: Opioidergic Mechanisms in the Human Brain.* Cereb Cortex Volume 18 (11), Pages 2523-2531, 2008.

[6] Deanna Brewer, Eugene B. Doughtie, Bernard Lubin. *Induction of mood and mood shift.* Psychotherapeutic Processes, Volume 36, Issue 1, Pages 215226, 1980.

[7] Rodney Brooks. *A robust layered control system for a mobile robot.* IEEE journal on robotics and automation, Volume 2, No. 1, Pages 14-23, 1986.

[8] Jan Broersen, Mehdi Dastani, Joris Hulstijn, Zisheng Huang, and Leendert van der Torre. *The BOID Architecture.* AGENTS '01 Proceedings of the fifth international conference on Autonomous agents, Pages 9-16, 2001.

[9] Rick Busselle and Helena Bilandzic. *Fictionality and Perceived Realism in Experiencing Stories: A Model of Narrative Comprehension and Engagement.* Volume 18, Issue 2, Pages 255280, 2008.

[10] Jianxia Chen, Yixin Chen, You Xu, Ruoyun Huang, and Zheng Chen. *A Planning Approach to the Recognition of Multiple Goals.* International Journal of Intelligent Ssystems, Volume 28, Pages 203216, 2013.

[11] Charles Darwin. *The expression of the emotions in man and animals.* New York: Filiquarian, 2007 (1872).

[12] Thomas L. Dean. *Intractability and Time-Dependent Planning.* In Proceedings of the 1986 workshop on Reasoning about Actions & Plans, Pages 245-266, 1986.

[13] Joo Dias, Samuel Mascarenhas, and Ana Paiva. *FAtiMA Modular: Towards an Agent Architecture with a Generic Appraisal Framework.* Emotion Modeling, Pages 44-56, 2014.

[14] Louise Ellison and Vanessa E. Munro. *Reacting to Rape.* British Journal of Criminology, Volume 49, Issue 2, Pages 202-219, 2009.

[15] Leon Festinger. *A Theory of Cognitive Dissonance.* California: Stanford University Press, 1957.

[16] Alice M. Isen, Kimberly A. Daubman, Gary P. Nowicki. *Positive Affect Facilitates Creative Problem Solving.* Journal of Personality and Social Psychology, Volume 52, No. 6, Pages 11221131, 1987.

[17] Alice M. Isen. *An Influence of Positive Affect on Decision Making in Complex Situations: Theoretical Issues With Practical Implications.* Journal of Consumer Psychology, Volume 11 Issue 2, Pages 7585, 2001.

[18] Carroll E. Izard. *Four Systems for Emotion Activation: Cognitive and Noncognitive Processes.* Psychological Review, Volume 100, No. 1, Pages 68-90, 1993.

[19] Dacher Keltner and Ann M. Kring *Emotion, Social Function, and Psychopathology*. Review of General Psychology, Volume 2, No. 3, Pages 320-342, 1998.

[20] James D. Laird, J. John Wagener, Mark Halal, and Martha Szegda. *Remembering What You Feel: Effects of Emotion on Memory*. Journal of Personality and Social Psychology, Volume 42, No. 4, Pages 646-657, 1982.

[21] Jennifer S. Lerner and Dacher Keltner. *Beyond valence: Toward a model of emotion-specific influences on judgement and choice*. Cognition and Emotion, Volume 14, Issue 4, Pages 473-493, 2000.

[22] Linda J. Levine and David A. Pizarro. *Emotion and Memory Research: A Grumpy Overview*. Social Cognition, Volume 22, No. 5, Pages 530-554, 2004.

[23] George Loewenstein and Jennifer S. Lerner. *The Role of Affect in Decision Making*. Handbook of affective science, Pages 619-642, 2003.

[24] Stacy C. Marsella and Jonathan Gratch. *EMA: A process model of appraisal dynamics*. Cognitive Systems Research, Volume 10, Pages 7090, 2009.

[25] Merriam-Webster. *Pity*. Retrieved from "www.merriam-webster.com/dictionary/pity" on 28-1-2017.

[26] Merriam-Webster. *Relief*. Retrieved from "www.merriam-webster.com/dictionary/relief" on 28-1-2017.

[27] Merriam-Webster. *Schadenfreude*. Retrieved from "www.merriam-webster.com/dictionary/schadenfreude" on 28-1-2017.

[28] Agnes Moors, Phoebe C. Ellsworth, Klaus R. Scherer, and Nico H. Frijda. *Appraisal Theories of Emotion: State of the Art and Future Development*. Emotion Review, Volume 5, No. 2, Pages 119124, 2013.

[29] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. *The uncanny valley [from the field]*. IEEE Robotics & Automation Magazine, Volume 19, No. 2, Pages 98-100, 2012.

[30] Roland Neumann and Fritz Strack. *"Mood Contagion": The Automatic Transfer of Mood Between Persons.* Journal of Personality and Social Psychology, Volume 79, No. 2, Pages 211-223, 2000.

[31] Arne Öhman, Anders Flykt, and Francisco Esteves. *Emotion Drives Attention: Detecting the Snake in the Grass.* Journal of Experimental Psychology: General Issue, Volume 130, No. 3, Pages 466478, 2001.

[32] Jeff Orkin. *Applying Goal-Oriented Action Planning to Games.* AI Game Programming Wisdom, Volume 2, Pages 217-227, 2004.

[33] Andrew Ortony, Gerald L. Clore, and Allan Collins. *The cognitive structure of emotions.* Cambridge university press, 1990.

[34] Rajagopal Raghunathan and Michel Tuan Pham. *All Negative Moods Are Not Equal: Motivational Influences of Anxiety and Sadness on Decision Making.* Organizational Behavior and Human Decision Processes, Volume 79, No. 1, Pages 5677, 1999.

[35] Ira J. Roseman. *Appraisal Determinants of Emotions: Constructing a More Accurate and Comprehensive Theory.* Cognition and Emotion, Volume 10, Issue 3, Pages 241-278, 1996.

[36] James A. Russell. *Culture and the Categorization of Emotions.* Psychological Bulletin, Volume 110, No. 3, Pages 426-450, 1991.

[37] James A. Russell and Lisa F. Barrett. *Core Affect, Prototypical Emotional Episodes, and Other Things Called Emotion: Dissecting the Elephant.* Journal of Personality and Social Psychology, Volume 76, No. 5, Pages 805-819, 1999.

[38] Klaus R. Scherer. *Appraisal Considered as a Process of Multilevel Sequental Checking.* Appraisal processes in emotion: Theory, Methods, Research, Pages 92-120, 2001.

[39] Norman B. Schmidt, J. Anthony Richeya, Michael J. Zvolenskyb, and Jon K. Manera. *Exploring human freeze responses to a threat stressor.* Journal of Behavior Therapy and Experimental Psychiatry, Volume 39, Issue 3, Pages 292304, 2008.

[40] Wolfram Schultz. *Neuronal Reward and Decision Signals: From Theories to Data*. Physiological Reviews, Volume 95, No. 3, Pages 853-951, 2015.

[41] Norbert Schwarz and Gerald L. Clore. *Mood, Misattribution, and Judgments of Well-Being: Informative and Directive Functions of Affective States.* Journal of Personality and Social Psychology, Volume 45, No. 3, Pages 513-523, 1983.

[42] Norbert Schwarz, Herbert Bless, and Gerd Bohner. *Mood and Persuasion: Affective States Influence the Processing of Persuasive Communications.* Advances in experimental social psychology, Volume 24, Pages 161-199, 1991.

[43] Norbert Schwarz. *Feelings-as-Information Theory.* Handbook of theories of social psychology, Pages 289-308, 2012.

[44] Magy Seif El-Nasr, John Yen, and Thomas R. Ioerger. *FLAMEFuzzy Logic Adaptive Model of Emotions.* Autonomous Agents and Multi-Agent Systems, Volume 3, Pages 219-257, 2000.

[45] Craig A. Smith and Richard S. Lazarus. *Emotion and adaptation.* Handbook of personality: Theory and research, Pages 609637, New York: Guilford Press, 1990.

[46] Deborah C. Stearns and W. Gerrod Parrott. *When feeling bad makes you look good: Guilt, shame, and person perception.* Cognition and Emotion, Volume 26, Issue 3, Pages 407-430, 2012.

[47] Square Enix and Jupiter. *The World Ends With You [Video game for Nintendo DS].* Published by Square Enix, European version, 2008.

[48] Dianne M. Tice and Ellen Bratslavsky. *Giving in to Feel Good: The Place of Emotion Regulation in the Context of General Self-Control.* Psychological Inquiry, Volume 11, No. 3, Pages 149159, 2000.

[49] Tzvetan Todorov. *The poetics of prose* . Ithaca, NY: Cornell University Press, 1977.

[50] Juan D. Velsquez. *Modeling Emotions and Other Motivations in Synthetic Agents.* AAAI-97 Proceedings, Pages 10-15, 1997.

[51] Michael Wooldridge. *Intelligent Agents*. Multi-Agent Systems, Pages 17-40, MIT Press, 2013.

The template used for this paper comes from: http://www.latextemplates.com/
The UML diagrams were drawn using http://www.umlet.com/umletino/umletino.html