Rogier van Dinther
Supervisor: Dr. Frank Dignum
11 August 2016

Using Biosocial Theory as a Model for Agent Emotion

Abstract: With increased demand for agents that exhibit emotion-driven behaviour, models are required that facilitate implementation of those agents. A commonly used model for this purpose is the OCC model, but this model falls short in representing the effects and self-regulation of emotion. Linehan's Theory of Emotional Dysregulation is an alternative theory that treats these aspects of emotion. An attempt was made to model this theory into an agent and represent emotional phenomena analogous to those in humans. Several sequential experiments were designed to test if various aspects of the theory could be replicated. While some real-life phenomena seemed to emerge, more research will be required to verify the feasibility of Linehan's theory.

Table of Contents

1. Introduction

The demand for agents that can display realistic emotion-driven behaviour has risen in recent years. There have been developments in the field of serious games using virtual reality and agent architectures to simulate an environment that players can directly interact with. These serious games are often used as training tools for high-risk jobs like soldiers, policemen and firemen. Agents play the role of other people in a virtual world, acting and reacting to the player. For this, emotions are important. Emotions are an integral part of human behaviour, especially so in the sort of situations that policemen and soldiers are trained to deal with. Providing simulated people with no or poor emotion-driven behaviour would leave the characters rather flat and fail to provide the training that serious games aim to give. As well, virtual agents need to have an understanding of how emotions work to recognise them in the human player and to act on them appropriately.

Biosocial Theory is a collective term for psychological theories that have in common that they look for the causes of psychological phenomena in biological factors and the social environment. Many of these theories lie close to the field of Evolutionary Psychology, stating that their particular phenomenon has naturally evolved to fill a function in social behaviour. A subset of biosocial theories is dedicated to the function and workings of emotion. There are separate theories on personality, emotion elicitation and what effect emotions have on behaviour and decision-making. Of particular interest is Linehan's Biosocial Theory of Emotional Dysregulation (Linehan, 2014). Though targeted at clinical therapy and psychological disorders, it makes some assumptions and claims that are interesting from an AI perspective.

The aim of this project is to investigate the feasibility and effectiveness of using Biosocial Theory, and Linehan's theory in particular, as a model for implementing believable emotion-driven behaviour in agents. As an extension of this, it will take a look at Dialectical Behaviour Therapy (DBT), a method of therapy developed by Linehan based on her theory. This method of therapy teaches ways in which emotions may be controlled when they become dysregulated, something which should not be overlooked for an implementation of emotion.

When talking about implementing emotion in an agent, it is impossible to avoid the debate on whether giving agents consciousness or 'feelings' - which emotions are near synonymous with - is ethical. Two things should then be said about this project. Firstly, the experiments here are little more than a proof of concept and are not advanced enough for ethics to be a concern. Secondly, it is not the aim to make agents feel the emotions, but only to make them give the impression on the outside that they experience them by acting on them in a believable way.

The question which this project will try to answer is: Does using Linehan's Theory of Emotional Dysregulation as a model for implementing emotion in an agent lead to behaviour that is analogous to real-life emotional phenomena in humans?

To answer this question, it is necessary to first explain in detail what Linehan's theory of emotion says about their causes and functions, how DBT relates to this and what skills it teaches. Furthermore, a short explanation needs to be given of the *facial feedback hypothesis*, as it will be used in this project. All that will be covered in Section 2. Next, a description is given of the custom-made architecture used in this project in Section 3. Some brief details are given of the actual implementation using Joost van Ooijen’s CIGA framework in Section 4. Finally, the experiments that were done are described in Section 5 with a discussion of their results.

2. Theory

There are many different theories of emotion. While there are differences in the way they define and treat emotion, some commonalities can be seen. Steunebrink writes: “Emotions typically have specific objects and give rise to action tendencies relevant to these objects” (Steunebrink 2010). In other words, emotions are always in a direct connection to some recognisable cause and a related effect. We will see later that Linehan follows this definition and makes it more explicit for individual emotions. Emotions should be distinguished from *moods*. Moods are “more diffuse and last longer than emotions” (Steunebrink, 2010). Their distinguishing factor is duration. Emotional expressions last in the order of seconds or minutes, hours in extreme cases. Moods can last weeks or months. We are not concerned with modelling moods here but it is then important to note where the limit of our interest lies.

A model commonly used for emotions in agents is the OCC model. Briefly summarised, the OCC model recognises 22 emotions. Which emotion is elicited is dependent on which aspects of a situation are appraised and how (Steunebrink, 2010). Situations can be appraised in three ways: 1) consequences of events, which can be *desirable* or *undesirable* in relation to the agent’s goals; 2) Actions of other agents, which can be *praiseworthy* or *blameworthy* in regards to the agent’s standards, and 3) Aspects of objects, which can be *appealing* or *unappealing* in regards to the agent’s attitudes or tastes. This divides the 22 emotions in three categories based on their category of causation. Furthermore, emotions are differentiated by other factors, such as whether they apply to the agent’s own actions or those of other (e.g. pride vs. admiration) and whether events apply to the self or to others (sadness vs. pity).

What makes the OCC model popular in the field of AI is that it gives clear definitions for which appraisals lead to which emotions (Steunebrink, 2010). This provides a straightforward model for at least the processing of emotion, if not for what the agent’s goals, standards and attitudes are and how situations relate to them. However, while the OCC model works well to specify how emotions are caused and how they relate to each other, it appears to say less about their effects and how agents can affect or regulate these emotions. This project will be more interested in these latter aspects of emotions, making the OCC model less useful. Linehan’s Biosocial Theory is focused on emotional regulation and how emotions affect behaviour, which is why the model presented here is primarily based off it. In the next section, this theory will be explained in more detail.

2.1. Biosocial Theory

Linehan’s Biosocial Theory of Emotional Dysregulation (Linehan, 2014) describes emotional dysregulation as “the inability (...) to change or regulate emotional cues, experiences, actions, verbal responses and/or nonverbal expressions under normative conditions” (Linehan, 2014). In a project aiming to build emotion into an agent it might seem strange to focus on theories describing, in a sense, how *not* to handle emotion. The reason for this focus is that, by looking at dysregulation and its effects, we gain an understanding of how emotion is normally regulated in humans. In other words, you can't grasp the concept of light if you don't have a concept of darkness. The understanding is in the contrast.

Before talking about the parts of Linehan's Biosocial Theory relevant to this project, it may be useful to explain the context in which it was intended to be used. Central to Linehan's theory is that emotional dysregulation is caused by two main factors: 1) a biological predisposition towards high emotional sensitivity and impulsivity and 2) invalidation of emotion during childhood (Linehan, 2014). Emotional sensitivity is the degree to which a person experiences emotion. Impulsivity here means that a person acts more directly on emotions. Both are personality traits that someone is essentially "born with" and aren't bad or good per se (Linehan, 2014). An invalidating environment can obviously be a situation where caregivers simply ignore expressions of emotion, either forcing the child to go to extremes to be noticed or teaching them that expressing emotion doesn't lead to results. Invalidation can also happen when a child - or adult, for that matter - is told that their emotions are bad ("That's a stupid thing to feel") or invalid ("Nobody else feels like you do, so stop it"). Linehan does point out that both factors - biological and social - are not required. Someone with normal emotional sensitivity who grew up in a neglectful environment can still develop emotional dysregulation.

Linehan has developed a form of counseling based on her theory to teach clients with emotional dysregulation to cope. Dialectical Behaviour Therapy (DBT) was originally aimed at people who suffer from Borderline Personality Disorder (BPD), though it has since come in use for treating other disorders. BPD is characterised by pervasive dysregulation of affect, relationships, self-concept and cognition. It is related to self-injury, suicidality and substance abuse, leading to frequent hospitalisations (Reeves, 2010). Linehan believes emotional dysregulation lies at the basis of that disorder (Linehan, 2014). What makes the things taught in DBT interesting here is that it makes explicit how Biosocial Theory views the function of emotions and how they are elicited and expressed. The next section provides an overview of Linehan's view on emotion, followed by a more in-depth explanation of some of the techniques used in DBT and their potential use in multi-agent systems. A third section goes into *facial feedback*. This psychological phenomenon is not strictly related to biosocial theories but will be used in the experiments to help model one of Linehan's DBT techniques.

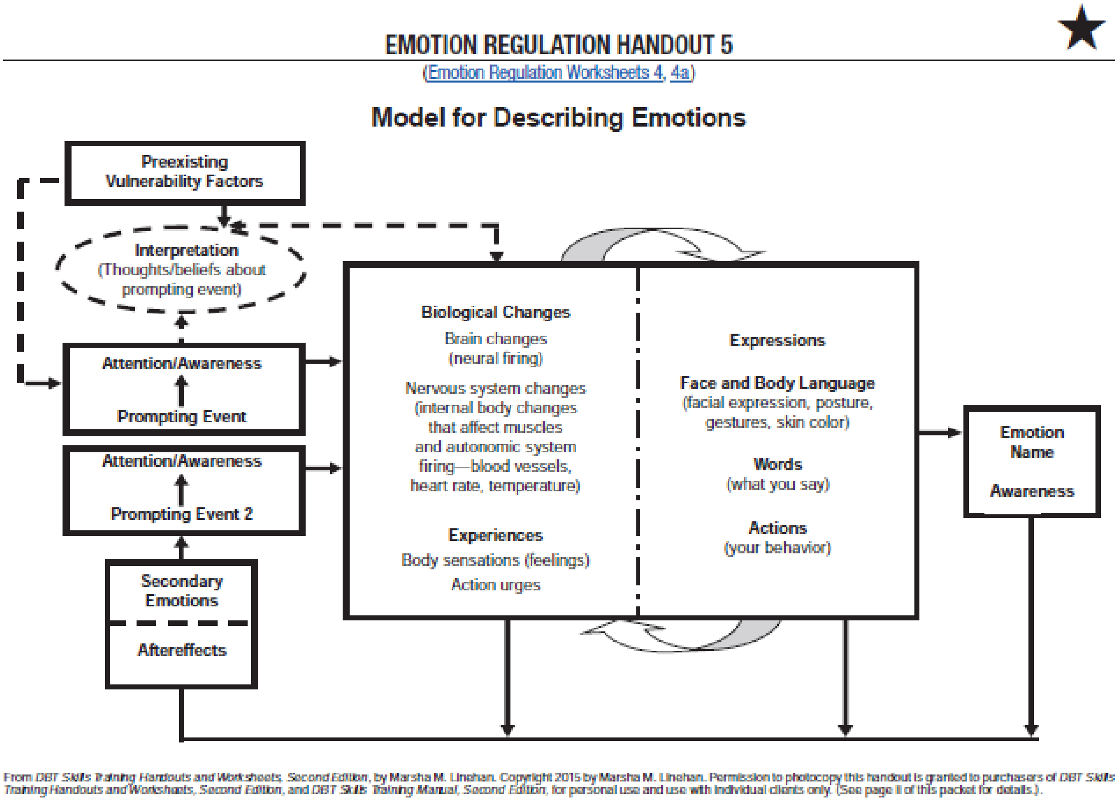
2.1.1. Overview

A biosocial view of emotions, by definition, assumes that emotions exist because of biological and social factors. They are a psychological mechanism that evolved through natural selection to organise physiological, cognitive and action patterns that facilitate adaptive responses to the environment (Izard, 1992). Indeed, Linehan takes the view that emotions are not something to be avoided but only something to be regulated. We need emotions in life. Emotions serve three main purposes: 1) motivating action, 2) communication to self, and 3) communication to others (Linehan, 2014).

Emotions motivate us to action. This is true in a few different ways. First of all, an emotion can prepare us physically for action. Secondly, emotions save time by biasing behaviour towards actions pertinent to the situation. From an AI perspective, this is rather like a filtering heuristic. When under attack, anger drives you to attack back in self-defence, faster than if you had to rationally consider your options. As a more positive example, many people like feeling anxious for a test (a form of fear), because it motivates them to keep studying (Linehan, 2014). These responses are mostly “hardwired” and are strongly related to the situation. As general examples, anger is a reaction to the blocking of goals and threats posed to the self or important others, and motivates towards self-defence and control. Sadness, by contrast, is a response to a loss of important objects or goals and motivates towards priorities and communicating to others that you need help.

Emotions are a way to communicate to ourselves. They can focus attention towards important detail in the stream of input that enter our senses in daily life. Fear directs our attention to physical threats, disgust directs it to possible poisoning - a rotting smell on fruit we’re about to eat (Linehan, 2014). This function of emotion does imply some independence of emotions from conscious perception. The emotion is elicited before we are consciously aware of it, so that it can register relevant information with our conscious awareness. Reisenzein et. al posit that theories of emotion elicitation can be classified as cognitive or noncognitive (Reisenzein et. al, 2013). Cognitive theories assume that emotions require “higher-order” mental representations and cognition, beliefs and desires in particular. Non-cognitive theories assume that emotions - at least some of them - have a more direct route to elicitation that bypasses cognition. The hypothesis that certain kinds of affects (joy at smelling a pleasant smell) are non-cognitively generated is “intuitively plausible”, but the other hypothesis that non-cognitive theories make, that basic emotions like fear, joy and anger can be non-cognitively caused, does not have a lot of support (Reisenzien et. al, 2013). Linehan seems to be mostly on the non-cognitive side of this divide, judging by this second function attributed to emotion.

As well as communicating to ourselves, emotions are also a way to communicate to others. There is much psychological and evolutionary evidence to suggest that emotions have a large social role. There are universal and innate facial expressions belonging to specific emotions (Izard, 1992; Linehan, 2014). These expressions are used and responded to by babies within the first few months of life and are even similar to expressions in nonhuman primates. The uniformity of responses to these facial expressions suggests that recognition of them is biologically innate (Izard, 1992). Aside from innate ways of expressing them, emotions also have a direct effect on others. An example of this is a baby’s spontaneous response to an adult’s smile without having learned to do so. “Emotional contagion” is also an example and a well-studied phenomenon. Emotional contagion is the susceptibility of people to mirror an emotion someone else near them is expressing (Lundqvist, 2008). This is different from simply reacting to an expressed emotion, as it only works for that same emotion. For example, feeling guilt when someone expresses anger towards you does not qualify. Guilt as a response to anger does qualify as an example for more indirect ways emotions affect others. They communicate information about a person’s perspective. If you see that your actions make another person angry, you would surmise that they consider them against their best interest or blocks their goal in some way. You might then adjust your behaviour accordingly. This is different from something like emotional contagion in that it happens on a rational level, whereas emotional contagion is subconscious. A big purpose of expressing sadness, according to Linehan, is communicating to others that we need help (Linehan, 2014).



EMOTION REGULATION HANDOUT 5

(Emotion Regulation Worksheets 4, 4a)

Model for Describing Emotions

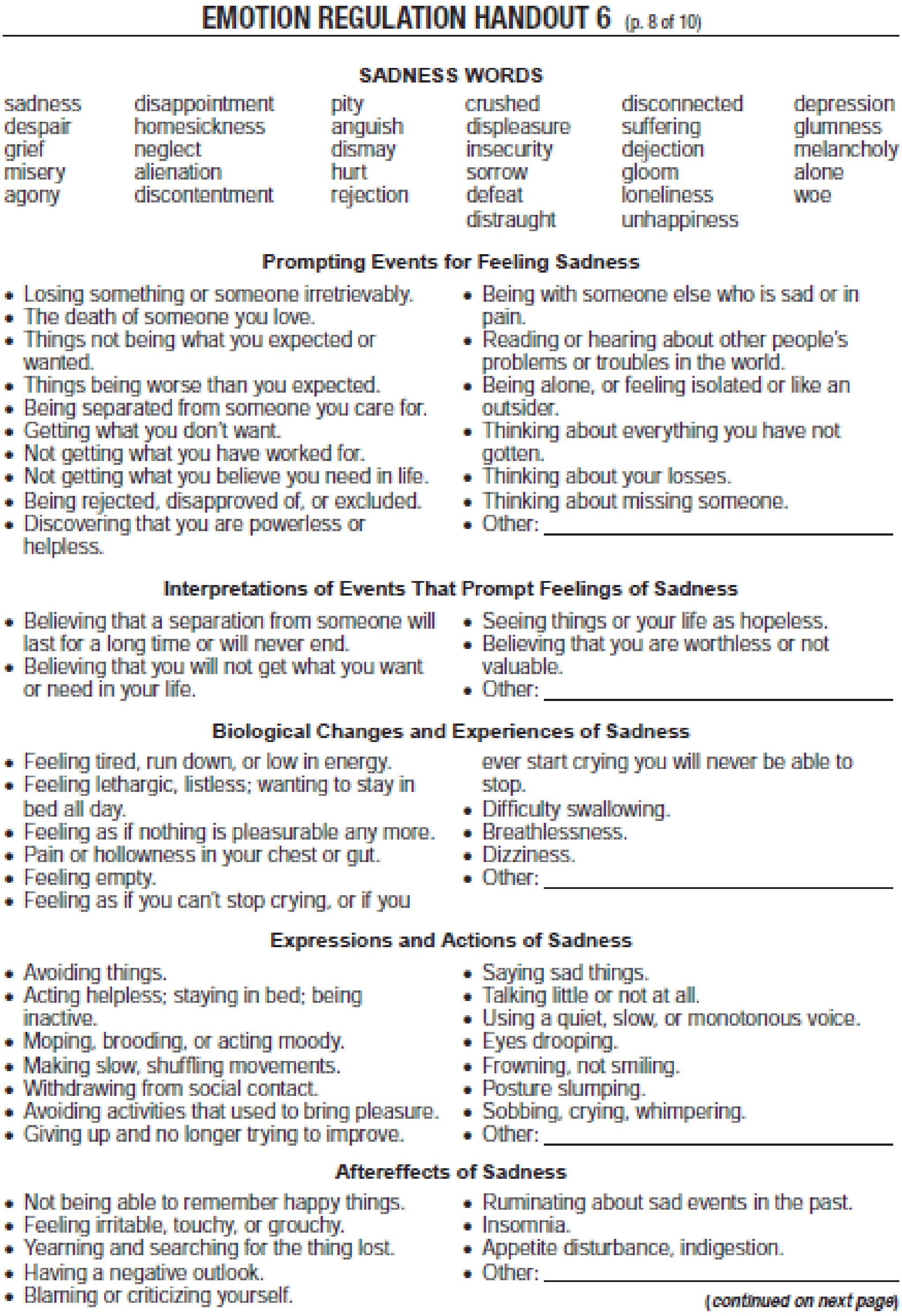
Preexisting Vulnerability Factors
Interpretation
(Thoughts/beliefs about prompting event)
Attention/Awareness
Prompting Event
Attention/Awareness
Prompting Event 2
Secondary Emotions
Aftereffects
Biological Changes
Brain changes (neural firing)
Nervous system changes (internal body changes that affect muscles and autonomic system firing—blood vessels, heart rate, temperature)
Experiences
Body sensations (feelings)
Action urges
Expressions
Face and Body Language
(facial expression, posture, gestures, skin color)
Words
(what you say)
Actions
(your behavior)
Emotion Name
Awareness

From *DBT Skills Training Handouts and Worksheets, Second Edition*, by Marsha M. Linehan. Copyright 2015 by Marsha M. Linehan. Permission to photocopy this handout is granted to purchasers of *DBT Skills Training Handouts and Worksheets, Second Edition*, and *DBT Skills Training Manual, Second Edition*, for personal use and use with individual clients only. (See page ii of this packet for details.).

Linehan goes into detail on emotions affect the person experiencing them. Above is a schematic taken from her DBT Skills Training manual (Linehan, 2014). Likely, as it is a handout given to clients, it is not intended to be an exact model but rather an impression. Still, a few things are worth noting here. Looking at the causes for the Biological Changes/Expressions part, there are two: prompting events, and the *interpretation* of prompting events. Prompting events are more or less treated as objective facts, whereas the interpretation of them involves reasoning and are subjective. Pre-existing factors, as discussed before, influence both these causes. On the other end, the effects of emotions are split in three parts. There are the immediate biological changes, which interact with the actual expression of emotion through body language, speech and actions - expression can also again cause biological changes. These two lead into the aftereffects, by which Linehan means more long-term changes in behaviour. These aftereffects in turn can also draw attention to other prompting events that cause more emotion.

A main element of biosocial theories is that there are a finite number of basic emotions that are universal among all humans. Others are learned or a combination of these basic emotions (Linehan, 2014). Linehan recognises 12 emotions: Anger, Disgust, Fear, Guilt, Joy, Jealousy, Envy, Sadness, Shame, Surprise, Interest, and Love. For each of these, she lists prompting events and interpretations of prompting events that might cause the emotion. Similarly, she lists biological changes, expressions and aftereffects relating to that emotion. A list of synonyms for the emotion is also given, to aid in understanding for clients, but that is not relevant here.

On the next page, one such description is given for Sadness. This, too, is taken from the DBT Skills Manual (Linehan, 2014). The description strikes a balance between being general enough to be applicable to many situations, while staying specific enough that it clearly distinguishes the emotion from others. Note that even the causes listed as direct prompting events sometimes require beliefs about the world or some interpretation. “Being with someone who is sad or in pain”, for example, requires that you recognised that state in the other person - and perhaps also some definition of what qualifies as being “with someone”. The underlying idea seems to be that the event as is would be recognised by any person who was there, thus making it a direct prompting event. “Believing that you will not get what you want” is more subjective. Another person could disagree or not share your perspective.



EMOTION REGULATION HANDOUT 6 (p. 8 of 10)

SADNESS WORDS

sadness
despair
grief
misery
agony

disappointment
homesickness
neglect
alienation
discontentment

pity
anguish
dismay
hurt
rejection

crushed
displeasure
insecurity
sorrow
defeat
distraught

disconnected
suffering
dejection
gloom
loneliness
unhappiness

depression
glumness
melancholy
alone
woe

Prompting Events for Feeling Sadness

- Losing something or someone irretrievably.
- The death of someone you love.
- Things not being what you expected or wanted.
- Things being worse than you expected.
- Being separated from someone you care for.
- Getting what you don't want.
- Not getting what you have worked for.
- Not getting what you believe you need in life.
- Being rejected, disapproved of, or excluded.
- Discovering that you are powerless or helpless.
- Being with someone else who is sad or in pain.
- Reading or hearing about other people's problems or troubles in the world.
- Being alone, or feeling isolated or like an outsider.
- Thinking about everything you have not gotten.
- Thinking about your losses.
- Thinking about missing someone.
- Other: ______________________

Interpretations of Events That Prompt Feelings of Sadness

- Believing that a separation from someone will last for a long time or will never end.
- Believing that you will not get what you want or need in your life.
- Seeing things or your life as hopeless.
- Believing that you are worthless or not valuable.
- Other: ______________________

Biological Changes and Experiences of Sadness

- Feeling tired, run down, or low in energy.
- Feeling lethargic, listless; wanting to stay in bed all day.
- Feeling as if nothing is pleasurable any more.
- Pain or hollowness in your chest or gut.
- Feeling empty.
- Feeling as if you can't stop crying, or if you ever start crying you will never be able to stop.
- Difficulty swallowing.
- Breathlessness.
- Dizziness.
- Other: ______________________

Expressions and Actions of Sadness

- Avoiding things.
- Acting helpless; staying in bed; being inactive.
- Moping, brooding, or acting moody.
- Making slow, shuffling movements.
- Withdrawing from social contact.
- Avoiding activities that used to bring pleasure.
- Giving up and no longer trying to improve.
- Saying sad things.
- Talking little or not at all.
- Using a quiet, slow, or monotonous voice.
- Eyes drooping.
- Frowning, not smiling.
- Posture slumping.
- Sobbing, crying, whimpering.
- Other: ______________________

Aftereffects of Sadness

- Not being able to remember happy things.
- Feeling irritable, touchy, or grouchy.
- Yearning and searching for the thing lost.
- Having a negative outlook.
- Blaming or criticizing yourself.
- Ruminating about sad events in the past.
- Insomnia.
- Appetite disturbance, indigestion.
- Other: ______________________

(continued on next page)

Any implementation of an emotion that follows Linehan's biosocial model of emotion should at least strive to represent the three main functions and should represent causes and effects described by Linehan for that emotion. For complex or non-basic emotions a good practice would be to break down what basic emotions underlie the complex emotion and use elements from those basic emotions. In this project we will assume that the causes, effects and functions define what an emotion is. To be considered a basic emotion at all, a system in the agent should demonstrably have facets of action motivation, communication to self and communication to the outside world, and have causes and effects that are a subset of those described.

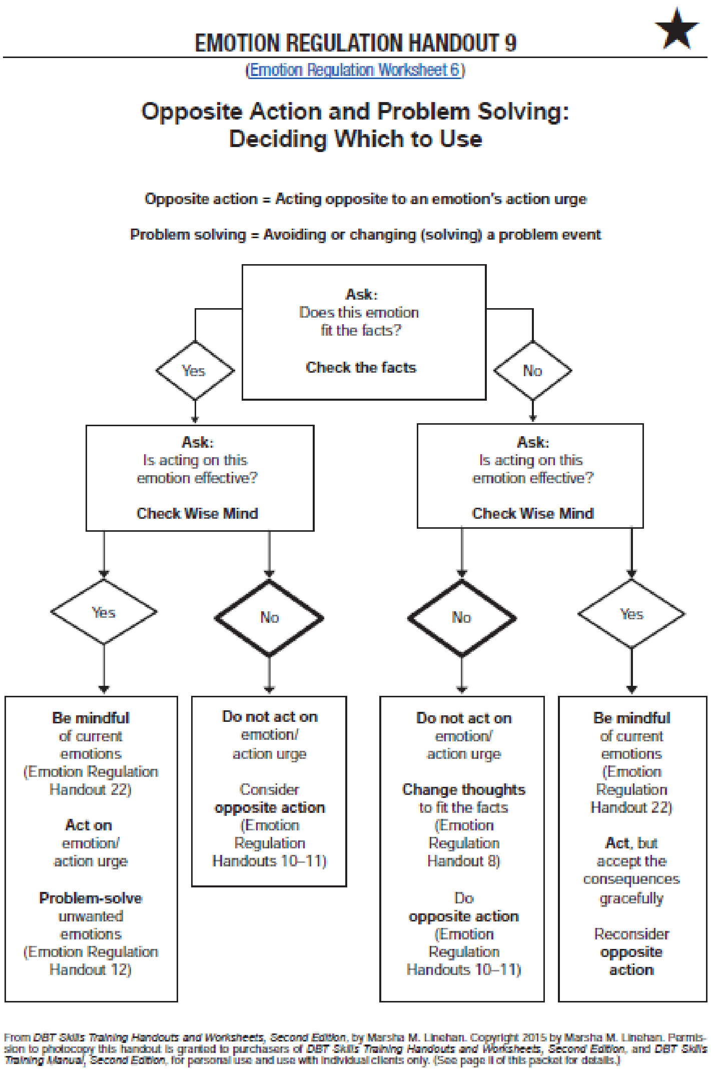
2.1.2. Dialectical Behaviour Therapy

As was said before, DBT is a form of therapy developed by Linehan based on her Theory of Emotional Dysregulation. It was originally designed for treating people with Borderline Personality Disorder, but has since come in use for treating other disorders and is even used in non-clinical settings (Linehan, 2014). The therapy consists of teaching skills to clients that they can apply in situations of emotional duress. What skills are taught in which session is strictly dictated, with half a session - 45 minutes - spent on skill training and the other half available for free discussion. There are different skills for regular emotion regulation, self-evaluation, and distress tolerance in crisis situations, all described in teaching notes and handouts for clients. Describing them all would go beyond the scope of this paper. A small set of related skills were selected to focus on in this project.

What makes DBT interesting for implementing emotion in the field of AI is that it makes things explicit that ‘normal’ people (i.e. without particular emotional dysregulation) do automatically. The therapy is targeted at people for whom emotional regulation does not come naturally. It is important to note that emotional regulation does not mean *removing* emotions. In Linehan’s view, emotional behaviours evolved as “immediate, automatic and efficient ways to solve common problems” and suppressing them entirely is only detrimental (Linehan, 2014). Rather, the goal is to keep emotions in a balance that lets them serve their function while getting out-of-hand emotions under control.

Perhaps we should wonder if a simpler solution could not achieve the same goal. One such solution to keeping emotions in check is to put a hard ceiling on their intensity. Say we measure every emotion as a numerical value, and cap values at 100. This would prevent any emotion from dominating too far over any other emotion. Another, perhaps more dynamic, solution is to give these same numerical emotion values a tendency over time towards the average emotion level. Relatively low emotions rise, and relatively high emotions fall. While both these solutions may emulate the same effect in some cases, the goal of this project is to find a way to model human behaviour. Where possible we should strive to stay close to how real humans function. Therefore, modeling emotion regulation based on a therapeutic method like DBT that is shown to work is preferable over mathematical capping or weighting that has no basis in psychological theory.

The DBT skills that will be explored here are the three focused around changing the emotional response: Checking the Facts, Opposite Action, and Problem-solving. On the next page is the flowchart Linehan uses to explain the link between the three. The skills in question or used when something is causing an emotion and you’re not sure if that’s justified. The first step is to check the facts. Is feeling this emotion reasonable given what events led up to it? For people, and especially for those with Borderline Personality Disorder, it can be easy to misinterpret a situation or to jump to conclusions. Consciously running back through what objectively happened can solve a problem before it starts.



EMOTION REGULATION HANDOUT 9 ★

(Emotion Regulation Worksheet 6)

Opposite Action and Problem Solving: Deciding Which to Use

Opposite action = Acting opposite to an emotion's action urge

Problem solving = Avoiding or changing (solving) a problem event

Ask: Does this emotion fit the facts? **Check the facts**

Yes

Ask: Is acting on this emotion effective? **Check Wise Mind**

Yes

Be mindful of current emotions (Emotion Regulation Handout 22) **Act on** emotion/ action urge **Problem-solve** unwanted emotions (Emotion Regulation Handout 12)

No

Do not act on emotion/ action urge Consider **opposite action** (Emotion Regulation Handouts 10–11)

No

Ask: Is acting on this emotion effective? **Check Wise Mind**

No

Do not act on emotion/ action urge **Change thoughts** to fit the facts (Emotion Regulation Handout 8) Do **opposite action** (Emotion Regulation Handouts 10–11)

Yes

Be mindful of current emotions (Emotion Regulation Handout 22) **Act**, but accept the consequences gracefully Reconsider **opposite action**

From *DBT Skills Training Handouts and Worksheets, Second Edition*, by Marsha M. Linehan. Copyright 2015 by Marsha M. Linehan. Permission to photocopy this handout is granted to purchasers of *DBT Skills Training Handouts and Worksheets, Second Edition*, and *DBT Skills Training Manual, Second Edition*, for personal use and use with individual clients only. (See page ii of this packet for details.)

When checking the facts, there are two questions to answer: 1) does the emotion fit the facts? For example, if something scares you, is there a rational reason to be scared? 2) Is acting on the emotion effective? (i.e. will acting improve the situation in any way?) As seen in the flowchart, Linehan recommends different solutions depending on the answers. What Problem-solving and Opposite Action are will be explained in a moment. The need for checking the facts relies on the assumption that perception is imperfect and/or that there are multiple inferences that can be made from the same percepts that can be narrowed down with hindsight, which is of course true for the real world. The model used in this project is relatively simple. Knowledge of the physical state of the world is near-perfect and events aren't complex enough to lead to large ambiguity. Therefore, modeling Checking the Facts as a skill for agents is of limited value. It's safe to assume that elicited emotions always fit the facts. That narrows the flowchart down to the left two outcomes. If acting on an emotion is effective, then there is little problem acting on it, though Problem-solving can help avoid it in the future. If acting on the emotion would be ineffective, it's better to not act at all or act opposite.

Problem-solving in this context doesn't quite have the meaning that it has in AI. To an AI researcher, 'problem-solving' typically refers to the general task for an agent of analysing a problem and possible actions, choosing a course of action through some planning or search algorithm, and executing that plan. Linehan uses the term to mean: avoiding or solving the cause of the emotion (Linehan, 2014). This can be leaving the room to cool down if you're angry, stopping to play a game that frustrates you, asking someone to stop an activity that bothers you, etcetera.

Opposite Action is little more than what it sounds like: doing the opposite of what the emotion urges you to do. The rather extreme example Linehan gives is when you're on a mountain and an avalanche is coming down the slope while you're standing by a jumpable gap. Even though you're afraid to jump, and feel the urge to stay there, it's better - safer - to jump anyway.

2.2. Facial Feedback

The facial feedback hypothesis was suggested as far back as James (1884) and Darwin (1872). Duclos describes it well in a nutshell as "When people are induced to act happy, they feel happier. When people are induced to act angry, they feel angrier". In numerous studies, participants were shown to experience an emotion more when made to adopt the expression or posture related to it (Doclus, 1989). Despite the name, it is not limited to facial expression. While the underlying cause of the effect is unclear, it is widely accepted and used in various forms of therapy (Linehan, 2014). Linehan shows in the schematic of her model of emotion that there is a feedback from expression of emotion through verbal and nonverbal means to biological changes, suggesting that she too accepts some form of the facial feedback hypothesis. It is not strictly speaking a part of biosocial theory. The reason it's discussed here is that it will be adopted into the agent architecture used in this project to implement one of the DBT skills into agents.

3. Conceptual Architecture

For this project, it was chosen to use a custom-made architecture. To explain why this choice was made, it is necessary to look at two existing architectures that were considered: BDI and FAtiMA Modular. It's worth noting that these architectures are not entirely mutually exclusive. A combination of both is conceivable.

BDI, Belief-Desire-Intention, is an architecture made by Bratman in 1987 (Oijen, 2014). It gives the agent three kinds of internal attitudes: beliefs, desires and intentions. Beliefs are what the agent thinks is true about the world and essentially form the agent's internal model of the world state. Desires are states the agent would like to be true, and that it will work towards achieving. A desire is fulfilled if it lines up with an identical belief. Intentions are a sort of commitment to a desire, making the agent pursue it above the others. This latter is necessary to provide consistent behaviour from the agent. BDI is popular because it provides concepts that are easy to relate to when needing to explain behaviours or reasoning processes (Oijen, 2014). For example, it has been used for modelling virtual team-members where domain-specific knowledge could be translated directly into agent knowledge (Oijen, 2014). Where BDI is less strong is when it comes to simulating some human-like behaviours. The abstraction of BDI has been found too high to accurately represent instinctual or physiological factors and when additional processes such as memory, emotion or learning are involved (Oijen, 2014).

FAtiMA Modular, Fearnot Affective Mind Architecture, is an agent architecture with planning capabilities, designed to use emotions and personality to influence the agent's behaviour (Dias et. al, 2014). It splits the process from perception to action into multiple stages, each of which can be specifically implemented by one or more components. Perceptions of events are stored in memory and also put through a sequence of Appraisal Derivation and Affect Derivation. Appraisal Derivation looks at the incoming percept and distils appraisal variables from it - e.g. desirable/undesirable and praiseworthy/blameworthy from the OCC model. Then, Affect Derivation takes these appraisal variables and generates affects - emotions and moods – from them. These are stored in the Affective State. The Affective State then plays a factor in Action Selection, which chooses the agent behaviour. As said, this architecture is not entirely mutually exclusive with BDI. Components could be implemented in FAtiMA that use – parts of – BDI.

Whereas many implementations of emotion use the OCC model, this project specifically aims to try another approach. FAtiMA, while ostensibly a core that any model or implementation can be laid on to, seems to assume some parts of the OCC model in the way it is set up. Biosocial Theory shares some similarities with the OCC model, but forcing it to interact with the architecture in some of the same ways that the OCC model does might undermine the differences between the two models. As for BDI, it does not currently seem suitable for modelling complex emotional interaction in an agent. It could be adjusted to have better support for this – and this is in fact being done (Oijen, 2014) – but doing that here would add another layer of complexity on top of what is already provided by the agents themselves and the mind-embodiment interface, potentially slowing down real-time processing. It seemed more efficient to make a custom architecture for this project.

The architecture used here draws inspiration from FAtiMA Modular in its use of interchangeable modules but moves away from the explicit extra layer of appraisal. In the sections below, a specification of the architecture is given.

3.1. Embodiment

A first design choice made in this project is to represent agents in the multi-agent system as an embodiment in a virtual environment, what is called an Intelligent Virtual Agent. Important characteristics of IVA's are that embodied in a real-time and virtual environment and that it has social abilities to interact with other IVAs or humans (Oijen, 2014). Agents occupy a physical space and can only act according to the capabilities of their body. Any interaction between agents should take place in the virtual world. This virtual world provides a place for emotion-related body language to take place and a physical environment for agents to emote about. Often, IVAs are also expected to have human-like ways of expressing and perceiving, but that is not necessarily the case in this project.

In more practical terms, there are two directions of communication between an agent's 'mind' in the MAS and its embodiment in the virtual world.

Percepts are collected by sensors in the embodiment and sent to the mind. For truly believable perception, sensors should be constrained in three ways: *situatedness, sensory capabilities* and *environment physics* (Oijen, 2014). Situatedness relates to the position of the sensor in the virtual world. Realistically, a visual sensor can't see things beyond a certain range. Sensory capabilities are the types of information that a sensor can take in, like the field of view of an eye limiting what it can see. The third constraint, environment physics, is that sensors should obey the virtual world's laws of physics, like how objects that are obscured by other objects can't be seen. In this project, the virtual world is sufficiently small that a reasonable maximum range of the sensor covers practically everything and for simplicity no maximum range is checked. The agent is treated as only having a visual sensor, although they are considered to have 360 degree vision rather than a view cone. This too is for simplicity. To take in the relatively sparse relevant events in the world used here, the agent would simply need to turn constantly to scan the world. Dealing with missed events and intelligent use of senses is not the focus of this project, so it is simpler to assume full vision of the world in an abstraction of an agent making a conscious effort to stay informed. In theory, agents *do* require line of sight to see an event or object. If situations can arise in the experiments where objects might end up being obscured for an extended period of time then this constraint should be enforced. However, since that state of affairs is very nearly impossible in the setup, it is safe to assume that no object or event is ever truly obscured and avoid potentially expensive computations on visibility.

Action messages are sent from the mind to the embodiment, instructing it how to behave. The embodiment can have a certain amount of autonomy in performing these actions, allowing the action message to be relatively high level. The available actions are chosen to give the mind meaningful control over its body while leaving the minutiae to the embodiment that has direct knowledge of the virtual world. For example, rather than make the mind choose speed and direction for every step the body takes, instead the mind can dictate a destination and the body will find some optimal way to move there. The embodiment reports back to the mind about every action in progress, so monitoring is still possible.

3.2. Modularity

The minds of agents are made up of a collection of modules that interact to define the behaviour of the agent as a whole. The agent only provides a framework to regulate these modules and provide a way to communicate with the embodiment. Though different modules serve different purposes, they are effectively created equal in how they can affect the agent. Modules are intended to be as self-contained as possible. This makes it simpler to add specific behaviours or remove them in a way that leaves the agents otherwise unchanged. This will be helpful in creating agents for changing scenarios, as changes can be explained in terms of the modules that were added or removed.

Modules as defined here have some constraints, both to allow efficient computation in a complex agent and to allow the kind of self-containment we strive for. Modules are passive, independent and only make suggestions for behaviour.

Modules are passive. They only run when an update is requested. This is a matter of computational efficiency. In complex agents, it’s possible that only a few are active at any one time while others are only interested in relatively rare events, like for example a message from another agent. As we will see below, update requests almost always happen as a result of changes in the agent’s state that are of interest to the module.

Modules are independent. They have no information on which other modules are present in the agent and can therefore only act on information provided by the agent. Communication to and from modules happens indirectly through variables. The agent keeps a map of *internal variables*, which can be read and written to by modules. A module interested in changes to an internal variable can register to it. The module will then be updated when the value of the variable changes. Similar to internal variables, *percept variables* are linked to perceived events in the world, like seeing another agent or object. Modules can likewise register with these to be updated on changes. Percept variables can only be read, not written to, by modules. Their values are changed by the agent upon receiving percepts from the embodiment. Another slight difference to internal variables is that when a percept variable is updated, any registered module immediately gets a chance to act on the new information. This design choice stems from a difference in what internal and percept variables represent. While a module waits to be updated, it is possible that the value of a variable changes more than once. That would mean that the module, which presumably reads the variable it is registered to, skips an earlier value. Internal variables represent goals, and higher-level beliefs. Only the most recent knowledge of those is relevant. Percept variables are the agent's model of the current state of the world. Events in the world happen only once, and repetition of the event is itself meaningful as a negative change in the state of the world. The possibility of missing a percept value is unacceptable.

So far we've only talked about how modules get information about the world and interact inside the agent. Obviously, modules also need the ability to dictate actions to the agent's embodiment. However, since modules have no knowledge of each other's existence, they can't be allowed to execute these actions directly. Conflicts would arise. Instead, all a module can do is to suggest an action to the agent. It is up to a resolver in the agent to decide which actions to execute. To give some hint to the resolver of an action's importance, the requesting module provides a bid value. What constitutes a conflict can vary between resolver implementations, but a few rules are followed:

1. If there is no conflict, an action is always executed
2. If there is a difference in bid value, the action with the highest bid is accepted, with the other action(s) rejected or aborted as the case may be
3. If the bids are equal, the most recent request is accepted

As an example of conflicting action requests, one can think of two modules suggesting different moves in space. The resolver would accept one of them and reject the other. Modules are provided with a way to track the status of their suggested actions first through the resolver and then during execution in the world. When the state of an action changes, its requesting module is scheduled for an update. In this way, actions function similar to variables.

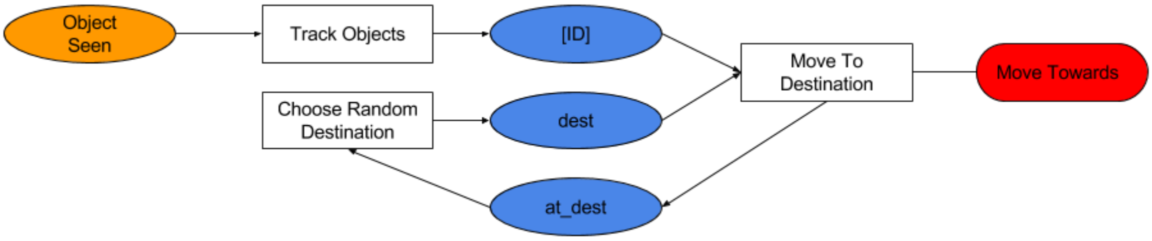
A simple example of such a modular agent would be one that walks randomly through the world. It picks a destination and walks there, after which it picks another destination and repeats the cycle ad infinitum. A schematic of this agent can be seen below. Arrows indicate modules (white boxes) reading and writing percept and internal variables (orange and blue ovals respectively). Actions (red tags) are connected to modules by a line since their relation is necessarily bidirectional. This agent has three modules.

The "Track Objects" module listens to ObjectSeen percepts from the embodiment, which report the position of objects in the world, including the agent's own embodiment, and stores object information in an internal variable corresponding to a unique identifier.

The "Move To Destination" module reads the agent's own position from this information and also reads a destination from another variable and requests for the agent to move if he is not currently at the destination. For the use of other parts of the agent, another internal variable is maintained that reports whether the agent is at the destination.

The "Choose Random Destination" module forms the heart of this agent. As the name suggests, it chooses a new destination at random when the agent is at its current destination.

It can be seen that each module performs a self-contained task, together resulting in the overall behaviour of walking between random points in the world.



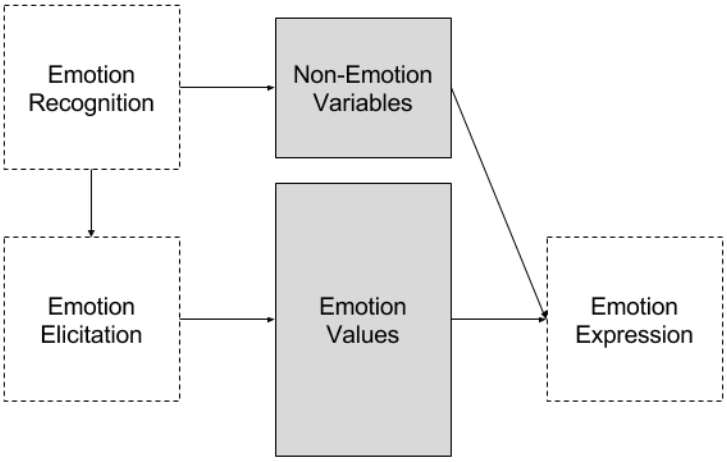
3.3. Model of Emotions

In attempting to model emotion in an agent, it is important to define what constitutes an emotion. Adding a counter ranging from 0 to 100, say, that represents the amount of anger an agent is experiencing could be said to serve that purpose. Intuitively, emotions are more than that. This project takes a functionalist approach. Some subsystem in the agent is only an emotion if it fulfills the functions ascribed to emotions in humans. The internal representation is secondary. Based on the function of emotion according to biosocial theories, an emotion must motivate action in the agent and serve in some way to communicate to the agent itself and to others. What makes two emotions different from each other is the way they meet these criteria, and also what external and internal factors cause them in an agent.

To meet the functional requirement, and in keeping with the aim of modules being self-contained, generally three types of emotion-related modules can be identified:

- **Emotion Elicitation Modules:** These modules define what causes emotion in an agent. What actions of others are considered “infuriating” (i.e. a cause for anger), what makes the agent happy, and so forth. They tend to listen to events in the world and some low-level internal variables and use them to compute emotion changes. These modules are also likely to listen to the output of emotion recognition modules mentioned below.
- **Emotion Expression Modules:** Modules in this category are on the other end of the diagram if the agent is drawn schematically. They define how an agent reacts to different combinations of emotions. They tend to suggest actions to the agent framework and rarely if ever changing emotion values themselves.
- **Emotion Recognition Modules:** Likewise to the emotion elicitation modules, these modules tend to listen to external percepts, but rather than focus on events and general behaviour of other agents, these modules are concerned with guessing at the emotions of others based on specific behaviours. This information can then be passed to other modules and used to decide how to behave. Modules using information on the emotions of other agents aren’t emotion-related in the way that it is used here. They treat information on emotions the same way as other internal variables.

The typical flow of information between these three types of modules is shown in the diagram below.



These three kinds of emotion-related modules do not directly correspond to the criteria for defining an emotion. They combine to meet these criteria. Elicitation and expression need to match to create an emotion that motivates actions relevant to its cause. Behaviour linked to an emotion by expression modules needs to correspond to behaviour recognised as indicative of an emotion by the recognition modules to be effective as a means of communicating anything. The emotion as a phenomenon emerges from this interplay of modules.

Internally, every emotion is represented by a numerical value. These are always positive values but are not otherwise constrained. Modules can adjust the value, usually increasing but potentially decreasing the emotion, and read out the current value much the same as internal variables. Unlike internal variables, emotion values are not passive states. Over time, the value of an emotion decays towards 0. The higher the value, the faster this decay happens. This models the fact that emotions are temporary and fade unless the cause of them persists or another cause presents itself. The speed of decay is different for different emotions. The emotions used in this project are Happiness/Joy, Sadness, Anger and Guilt. Happiness was set to have a decay rate such that it halves after 60 seconds. The negative emotions of Sadness and Guilt were given the longer half-life of 120 seconds. Cloninger supports the claim that humans are more averse to negative emotions like sadness than they are drawn to positive emotions like happiness (Cloninger, 1986). Slower decay reflects this. Anger fits sort of a special niche in that it is not a negative or positive emotion as such but rises and falls quickly. Its decay time is short at 30 seconds. What effects these exact decay times have on any real simulation depends on the time between emotional impulses and their intensity. Emotion decay is not to be confused with emotion regulation. Regulation is a deliberate controlling of emotions by the agent, whereas the emotion decay described here is analogous to a physiological property of emotion outside the agent’s control.

Emotion regulation is a separate system altogether from emotion elicitation, expression and recognition. It should only become active when emotions get out of bounds and need to be controlled. The techniques used for this are as described in the Dialectical Behaviour Therapy manual by Linehan (Linehan, 2014). Modules concerned with emotion regulation affect emotion values in specific ways, or directly guide behaviour to indirectly change emotions, but should ideally only do so when necessary.

3.4. Classifying Modules

The following is not so much a defining part of the architecture used in this project, and rather an observation of how agents constructed in this way can be viewed on a higher level. All modules individually guide agent behaviour and are essentially created equal in doing so. However, there are two ways to categorise modules. The first is what sort of behaviour they are a part of. Generally modules fall in one of three systems of behaviour, which I shall call the Goal, Emotion and Reaction systems:

1. **Goal:** Modules in the Goal system handle an agent's rational beliefs and goals, separate from emotion. This is the part of the agent that is akin to a BDI agent or task planner. If it were the only system, the agent would simply be a task-oriented automaton.
2. **Emotion:** Modules in this system are concerned with how the agent responds to its own emotions, including body language, changes in decision-making and explicit actions to express emotion.
3. **Reaction:** These modules react to the emotions of other agents, choosing behaviours that communicate a response to them or aim to increase or decrease that emotion in the other agent.

The Emotion and Reaction system both include parts of the model of emotions as described above, possibly overlapping a little with some modules defining both Emotion and Reaction type behaviours. To keep the agent organised even when it becomes more complex, these three systems should be kept separate. If a module handles more than one type of behaviour then it might be better to split it into multiple modules, each fitting within one system. Communication between systems should ideally happen on a high level, passing meaningful interpretations of the world rather than raw percepts and observations.

Another way to categorise is by where modules fit in the chain of information from percepts to actions. Here too there are three classes that almost all modules fit into, somewhat arbitrarily called Processing, Interpretation and Decision:

1. **Processing** modules listen to percepts from the embodiment, interpret the information into a more useful form and write this to internal variables for use by other modules. A prime example of this is the Track Objects module from the example, which reads an information dump about objects in the world and extracts the perceived agent's position from it.
2. **Interpretation** modules read and write only internal variables or emotions. They take variables written by Processing or other Interpretation modules and translate it to a more complex belief that can in turn be used by other modules. These modules will become more prevalent as an agent becomes more complex. In the example, the Choose Random Destination module is a Processing module.

3. **Decision** modules are at the end of the chain of information. They read internal variables and emotions and suggest actions based on that data. The Move to Destination is this in the example - though it also fits in Interpretation on account of updating a variable representing if the destination has been reached

Contrary to the classification in behaviour types, the distinction between classes isn’t as clear-cut, and that is fine. Many modules in this project have marks of multiple classes. Emotion Elicitation modules generally are either Interpretation or Processing modules. Interpretation-class Emotion Elicitation modules model the interpreted causes of emotion described by Biosocial Theory, while Processing-class Emotion Elicitation modules model direct causes.

4. Implementation

To implement the architecture described above, this project makes use of Joost van Ooijen’s CIGA framework as a middleware application. CIGA forms a bridge between a Java implementation of the agent mind on one side, and an implementation in Unity3D[1] of the agent embodiment and virtual world. Communication between the two takes place through a common ontology constructed with Protege[2].

4.1. Java-side

A singleton MAS class in Java is in control of when agents get a chance to update their modules and systems. It also relays events and action updates to the appropriate agent and passes messages from agent to agent. An abstract Agent class defines all standard operations described in the previous section, like updating modules that were scheduled to do so and calling the action resolver at an appropriate time. Concrete types of agents are defined as subclasses of this Agent class, choosing an action resolver and emotion system implementation in the constructor. Typically whatever modules are needed are also added in the constructor. Modules can also be added and removed later - though this functionality is not used in this project.

Modules are likewise subclasses of an abstract Module class. This class defines three methods for defining module behaviour: 1) an update() method that is called some time after the module was scheduled for an update; 2) an onPercept() method that is called when an event occurs that the module is registered to; 3) an onMessage() method that is called for every module when another agent sends a message to the containing agent. By default, these methods do nothing, so concrete modules should override them to define behaviour. Typically, a module registers to relevant variables and events in its constructor, though it is possible to register and unregister at a later time as well.

Implementations of the action resolver derive from an abstract ActionResolver class that handles adding and removing action requests and state changes for actions. Implementing classes provide a method that receives a list of all current action requests - pending and active - and returns a subset of them that it wishes to accept. The ActionResolver then tells the agent to execute or abort actions through the CIGA bridge.

[1] https://unity3d.com/

[2] http://protege.stanford.edu/

Implementations of the emotion system derive from EmotionSystem. This class decays emotion values every frame based on their pre-defined half-life values, and provides a way for modules to change emotion values. An implementation must instantiate the different emotions it recognises and define their half-lifes when it instantiates the emotion variable objects. Other time-based effects on emotions - like facial feedback - are added by overriding the update method, though the original update method must still be called within.

4.2. Unity-side

Unity is a game engine targeted at 3D games. It uses component-based design for world objects. CIGA connects into the C# script components that can be added to a Unity object. Any world object relevant to CIGA needs to have a script identifying it as such. This includes the embodiments of agents, which also need a separate script identifying it as an agent. A central world object handles updating the Unity side of CIGA. This object also allows agents to broadcast events to all agents.

Every agent in Unity has a number of child objects and components. One is the CIGA object that forms its embodiment. This is the only ‘visible’ part of the agent. The agent also has a sensor component that listens to all events and records the positions of all CIGA objects and expressed emotions of all CIGA agents. A behaviour realiser component executes all active actions and sends updates on state changes through the CIGA bridge. A special expression handler wraps the colour changes that agents use as a way of expressing emotion and allows the sensors of other agents to read this information. For testing purposes, a UI component is also attached to display information on the agent’s internal state.

Although Unity provides ways to check for collisions between objects, they are not applied here. Making agents deal with avoiding collision while moving around each other, though realistic, is not the focus of this project and has little bearing on their behaviour. It is therefore ignored.

5. Scenarios

5.1. The Joint Delivery Task

To experiment with any sort of model of Biosocial Theory in agents, there need to be goals for agents to try and achieve. The basic emotions defined by Linehan have causes related to - among others - succeeding or failing to meet goals and whether expectations are met. In the scenarios used in this project, a variation is used of what is called here the Joint Delivery Task (JDT).

In the basic form of the JDT, there is a rectangular field 10 meters wide and 20 meters deep. Two agents are placed on this field at arbitrary positions. As the task starts, a package appears somewhere on the short edge at one side of the field. The ultimate goal is to bring this package to a drop point at the opposite side of the field. One agent is the first messenger. His task is to pick up the package from where it appeared and bring it to the other agent, the second messenger, who must then carry the package to the drop point. Once the package is on the ground within 1 meter of the drop point, it is removed. The agents then have 5 seconds of idle time before the package appears again and the task repeats. Both agents walk at a speed of 5 m/s. To be able to pick up, drop or transfer the package, an agent would have to stand within 3 meters of the target. Carrying the package does not affect agent speed.

This task is intentionally simple in its basic form. It can easily be extended with extra constraints or complexity to suit a scenario's aim. The two agents are forced to cooperate and interact to complete the task, with readily apparent progress of the task to its completion in where the package currently is on the field. This provides hooks for measuring group and individual success, and communicating and recognising intent, to name a few interesting cases for emotion elicitation and expression.

The two agents show different behaviour, but the states that they recognise are the same and are one of four: 1) the task is inactive 2) they are holding the package; 3) the other agent is holding the package; 4) the package is on the ground and not held by anyone. Each agent also has a different pre-defined idle position. For the first messenger this is the center of the line where the package appears. For the second messenger this is the middle of the field. Given the task, these are the optimal positions to wait at. The actions suggested for each state are as follows:

	First messenger	**Second messenger**
Task inactive	Go to idle position	Go to idle position
I am holding package	Move to second messenger and give package	Move to drop point
You are holding package	Go to idle position	Move to first messenger
Package is on the ground	Move to package and pick it up	Go to idle position

Before moving on to further scenarios, it was assured that this agent design could complete the task successfully and without problems. Any issues with execution later would then have to come from adjustments made in the scenario.

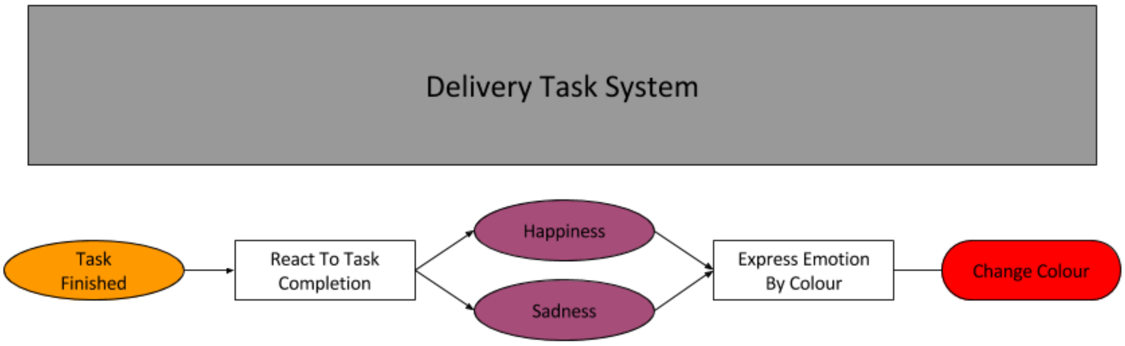
5.2. Scenario 1 - Happiness and Sadness

In this first iteration on the JDT, the task is expanded so that there is a possibility of failure. A time limit is imposed on completion. If the agents deliver within the time limit, the task is considered successful, otherwise is it considered failed. The limit is calculated based on the distance between where the package was spawned and the drop point, using the following formula:

$$T = ((d_{1/2}/v_1)+(d_{1/2}/v_2))*m$$

Where v_i is the maximum speed of agent *i*, $d_{1/2}$ is half the distance between the package and the drop point, and *m* is a constant. Essentially, the expected time is how long it takes for the first agent to walk halfway, where the second agent takes the package and walks the rest of the way. The constant *m* allows to tune this estimate to adjust for time needed to pick up, hand over and drop the package, and delay caused by computation. The aim is to set *m* in such a way that the task is successful just over 50% of the time.

Two emotions are introduced for this scenario: Happiness and Sadness. Agents get an increase of 10 in Happiness when the task is successful. Agents likewise get a similar increase of 10 in Sadness when the task fails. To express these emotions, the agents change the colour of their bodies to reflect the dominant emotion they are feeling. This is an abstraction over many facets of body language and biological changes that Linehan lists for these emotions. The colours yellow and blue have arbitrarily been chosen to reflect Happiness and Sadness respectively. The exact colour expressed is a linear interpolation between white and the relevant colour, with any value under 5 not being shown and a value of 50 expressing as the full colour.



As can be seen in the figure above, there is no interaction between the Emotion and Goal systems. Emotions are not recognised by other agents or affect any of the goal-oriented behaviour. It is worth noting that, for the lack of a communicative function, happiness and sadness as used here are not full emotions according to our own definition. The reason to start with these simple partial emotions is to create a basis for how later, more complex, emotions should be handled. Linehan states that sadness might lead to becoming inactive and losing motivation to do things (Linehan, 2014), which in this scenario might translate to walking slower, or not doing the task at all. At this stage, that behaviour would simply lead to a negative feedback loop that would not be interesting to study. A full chain of elicitation, internal processing and expression is needed for even this, and the knowledge of what did and didn’t work in the design can be carried forward to later scenarios.

5.2.1. Expectation

With the time limit properly balanced, the agents will succeed more often than not, and thus get a happy impulse more often than a sad one. Depending somewhat on how the decay constants of the emotions compare, the expectation is that agents will become progressively happier over time. With no emotion regulation, happiness will drown out any occasional failure/sadness. There is no negative consequence for that at this point, but later even a positive emotion growing out of control could be disadvantageous since other emotions could be drowned out. In the context of this scenario this would mean that agents can be made to emote over simple events in the world and express this in a simple independent behaviour change (i.e. colour change).

5.2.2. Result

After 15 iterations of the task, the success rate was 53.3%. Therefore agents should have gotten slightly more happy impulses than sad ones, though not significantly so. Contrary to expectation, happiness does not become dominant by repeating the task. Happiness and sadness alternate in being the highest emotion, with a little more time spent with the agents in a sad state. This latter effect can be explained by the slower decay of sadness. When the values of happiness and sadness are close together, sadness will win out over time. This would occasionally express by the agent changing colour without an external prompting event.

5.3. Scenario 2 - Anger and Guilt

Building on Scenario 1, agents are now given a concept of their local performance within the task: whether they perceive themselves as having done “their part”. Contrary to the outcome and timing of the task, which is common knowledge among all agents, local performance is a belief and may be inconsistent between agents. In a reasonable standard for local performance, if both agents are locally successful, the task would necessarily also succeeds. For this scenario, each agent considers itself successful if:

$t_{local} \leq T / 2$

Where t_{local} is the time taken to complete the agent’s own part of the task. For the first messenger, for example, this is the time between the package appearing and when he hands it over to the other agent. It can be seen from the formula that if both agents are successful: $t_1 \leq T / 2$ and $t_2 \leq T / 2$, thus $t_1 + t_2 \leq T$. Since $t_1 + t_2$ constitutes the full time taken for the task, this would mean the task is successful. So this standard is reasonable. Later scenarios may change or refine this standard.

Using this concept of local performance, the emotions of Anger and Guilt are introduced besides the pre-existing Happiness and Sadness. Linehan lists a number of prompting events for anger and guilt. Anger can be caused by having an important goal blocked or having things not turn out as expected. Guilt is caused by not doing something you said you would do or by thinking that your actions are to blame for something. (Linehan, 2014) At the end of the task, the agent compares the outcome to his own local performance. If the task failed, but the agent considers that he was locally successful, he becomes angry, as he concludes that the failure is the other agent’s fault. When the agent considers himself to have failed locally and sees the other agent is angry, this will cause a response of guilt as he feels rightly blamed for the failure. This effect will only take effect if the perceived Anger in the other agent is higher than the Guilt already felt, and will increase Guilt by the amount of Anger perceived. For example, if Guilt is 10, and the perceived Anger is 20, Guilt will be increased by 20 to a total value of 30. The next tick, Guilt is higher and will not be affected. Agents don’t feel guilty when they consider themselves locally successful. In humans, being falsely accused might lead to anger in return and possibly an argument expressing this, but that effect is not modeled here.

Guilty agents increase their speed proportional to their level of guilt, attempting to correct their mistake. Attempting to make amends is one of the expressions of guilt listed by Linehan (Linehan, 2014). Since not meeting time constraints is the only source of guilt in this scenario, increasing speed is a natural response to counteract it and make amends so as to elicit no more anger from the other agent. Speed increase is treated as a modifier to base speed. Agents have a minimum modifier of 1.0 and a theoretical maximum of 2.0 - double the speed. The effect of Guilt on speed is a diminishing return, growing asymptotically towards 2.0 and having a value of 1.5 at a Guilt value of 25. A fair question is why all agents wouldn’t increase their speed if they can. Though it is not implemented here, the assumption is that walking faster would be tiring, or is undesirable in a different way, and will be avoided by default.

Now we see different primary functions in different emotions. Where anger has a role in communicating to others, guilt is a clear motivator to change behaviour. Happiness and sadness have no clear function by themselves but in this case serve to form a baseline for other emotions, which now have to compete with this baseline for dominance in expression. To remain effective as an action motivator, feelings of guilt have to persist long enough to let the action - increased speed - continue. In contrast, anger serves little purpose after its message has been passed on and should decay fast after an intense initial rise. This pattern applies more generally: emotions with a primary motivating function must have long decay times to stay effective; emotions with a primary communicative function are not served by persistence and should decay fast after a large initial rise to make place for other emotions. In reality, Sadness serves a bigger role than it is given here, Among other things, it communicates a cry for help to others. If we encourage this ‘emotional baseline’ behaviour in the system, it is worth questioning if Happiness is not the only true baseline emotion, with Sadness behaving more similar to Anger. If that is the case, then the absence of happy events - as perceived by the agent - would disrupt the emotion system as a whole.

Having to now account for their own performance, it would be natural that both agents become more selfish in the way they cooperate. Neither agent is willing to move beyond the centerline of the world. Walking beyond that means more work for you and less work for the other agent. To reflect this, the movement behaviour of the agents is changed. Originally, the second messenger would move towards the first messenger when the latter was holding the package, resulting in them crossing the centerline. Now the second messenger positions himself on the centerline of the world to be between the first messenger and the drop point, thus helping without putting themselves under a heavier workload.

5.3.1. Expectation

As the task is repeated and failure happens, it is expected that both agents are occasionally at fault for it. Thus, both agents will accrue some level of guilt and anger in their emotions. In the way expression is modeled here, only the dominant emotion - with the highest internal value - is expressed. This would lead to problems if anger and guilt were the only emotions felt. The two shouldn’t be in direct competition to each other. Fortunately, this is not the case. The baseline formed by recurring Happiness and Sadness during the tasks should emergently form a divide between active and dormant emotions. Anger will spike above the baseline when triggered but because of its fast decay time it will soon fall below the line again and go dormant. This is desirable behaviour for anger. Guilt will stay active for longer due to its long decay time but with no further causes for guilt it is overtaken by happiness and sadness, leading to the agent reverting to normal speed.

With the dynamics of anger and guilt, what might happen is that one agent stays in a perpetual state of guilt while the other is mostly happy. A bully/victim scenario, in a sense. This would occur when one agent, by chance, is the first to cause the task to fail. This will result in guilt and increased speed. Meanwhile the other agent is only angry for a short amount of time. With the increased speed the task is more likely to succeed and when it doesn’t the bully will have done less work and likely thinks he did his part within the time limit. This perpetuates the victim being the only agent to feel guilty. Whether this situation is desirable is debatable.

If the scenario results in the expected behaviour, this means that agents can interact to influence each other’s behaviour through emotions. A balance can be found between different emotions that don’t share the same purpose while still giving the agent consistent behaviour.

5.3.2. Result

With two new emotions and some changes in the idle behaviour, the value of *m*, the modifier on time given to complete the task, needed to be rebalanced. With *m* = 0.9, the success rate after 20 iterations was 75%. This is higher than intended but it was left that way because the higher success rate was largely due to the speed increase from Guilt. Initially, the *m* value was too high and thus the time limit too generous. This was likely a result of the change in idle behaviour leading to better positioning. The task was trivially successful, meaning that none of sadness, guilt or anger occurred. This was obviously not an interesting outcome so *m* was dialed down. For some just slightly too low values another thing happened. With too little time, the task nearly always failed. However, due to an inaccuracy in determining local performance, both agents would still consider themselves to have succeeded. The result was both agents becoming angry and no change being made to behaviour. The inaccuracy was resolved, and this outcome with it, but it does point to a possible direction of further research. Disagreements like these can only be resolved by reasoning about the other's emotions in response to your own, something which the agents don't do at present. If interaction through emotions is to become more complex, a way of doing this sort of reasoning will have to be created.

The bully/victim pattern did occur, but not as much as expected. After 20 iterations, the first messenger had spent 57.1% of the time being happy, and 42.7% of the time being guilty. The second messenger was happy for 86.8% of the time and angry for 13.1% of the time. This does point to a mild bully/victim relationship where the second messenger gets the advantage. The skew is likely due to an unfairness in the way local success is counted. The first messenger starts counting when the package appears, meaning it has to walk to the package as well, something which the second messenger does not have to account for. One thing making up for this skew is that the first messenger doesn't quite have to walk to the centerline to hand over the package. He stops short a small distance from the second messenger, whereas the second messenger will have to cross the other half of the field completely. Apparently, this does not entirely make up for the unfairness. In future scenarios, an adjusted local success measurement should be implemented that favours the first messenger a little.

Due to its long decay time, guilt lasted long in an agent once elicited - compare the 42% time spent guilty for just 13% anger in the other agent. At the start of a run, when not many task iterations have been completed, guilt and happiness interacted in the predicted way. When correctly blamed for a failure, guilt would be the dominant emotion for a few iterations and then fall below the baseline. In the long term, however, guilt does tend to rise overall relative to the baseline. Then it takes less of a guilty impulse to catch up with the baseline and the value will be higher above it. The agent starts spending more time being guilty before reverting back to happiness. If this is undesirable, a solution might be to decrease the decay time of guilt slightly, but as the system gets more complex this sort of balancing would become increasingly difficult to do. A better, and perhaps more natural, solution might be found in emotional regulation as described by Linehan and others. This is explored further in Scenario 4.

5.4. Scenario 3 - 3rd Agent

We now move away from adding width to the emotions and start adding depth. This third scenario keeps the task and previous emotion causes and expressions unchanged, but now adds a third agent to the world. This agent has a second messenger role, giving the world a total of two second messengers. Adding a first messenger would have been far more complicated for little gain. There would have to have been a second package or conflict resolution if both try to pick the package up at the same time. That's why it was chosen to give the extra agent the second messenger role.

This change opens up new interactions for the agents. The first messenger can now choose which second messenger to work with. If there is some reason to want to avoid an agent - working with them causes a lot of negative emotions, for example - that is now possible. In past scenarios, the only possible way to avoid the other agent was to refuse to do the task, which isn't viable as it would only make the task even longer and it would make the scenario grind to a halt, which isn't an interesting result.

The third agent also opens up another cause for emotion. Not being chosen could be a cause for Anger or Sadness. Which of these is elicited is dependent on the interpretation of this choosing event. From a rational perspective, the optimal thing for the first messenger to do is to work with the agent closest to the package's initial position. The second messengers know this. If you are closer to the package but the first messenger still chooses someone else, that means you were treated unfairly from your point of view. Believing that you have been treated unfairly is a cause for anger (Linehan, 2014). In contrast, if you were passed over but you were also not the closest to the package, that is reasonable but it arguably makes you feel that you are not useful, since you'll be sitting this iteration out doing nothing. Feeling useless or not valuable is a cause for sadness (Linehan, 2014). Repeated refusals could have an increasing effect on sadness, as it would compound the feeling of uselessness, but that is not implemented in this scenario as its effects would be hard to balance and test.

If emotions figure into how the first messenger chooses its work partner, which it probably should in a realistic scenario, a third type of interaction new to this setup is that the second messengers can try to manipulate what emotions they express to convince the first messenger to choose them. Note that agents should have no conscious control over what emotions they *experience*, but they can deliberately affect their expression. In this scenario, the first messenger chooses its partner randomly. This allows to test both possible effects of being rejected - anger and sadness. Later research can implement a deliberate choice as either a conscious decision or an expression of emotion or a combination of these.

There are now two possible reasons that a second messenger might be angry at the first messenger: 1) because the first messenger failed to do his part of the task in time, 2) because the first messenger unfairly passed over that second messenger. One way to disambiguate between the two is the timing of the emotion. Anger over failure would occur at the end of the task. Anger over rejection occurs in the middle. As yet, second messengers do not express anger at each other. If something like that were implemented, Envy would be a more appropriate emotion. A number of causes described by Linehan fit but "Others get something that you really want and you don't get it" is the closest to this situation.

Each iteration, the first messenger chooses its partner with equal chance to each second messenger. A more deliberate way of choosing a partner could be made later. When it chooses, the first messenger sends a message to all second messengers informing them if they were chosen. This is necessary because other agents have no way of knowing when the first messenger has decided. An agent that does not participate in the current iteration moves to the idle position and does not try to intercept the first messenger. The idle positions have changed too. Before, the first messenger would move to the start line and the second messenger would move to the center of the world when idle. With more agents this behaviour is not as straightforward. First messengers spread out evenly on the start line and second messengers spread out on the centerline, sorted by their unique IDs. Of course, as there is only one first messenger, it still moves to the center of the start line as before. As a side effect, in deciding who 'should' get the package, each second messenger agent covers an equal partition of the start line where if the package appears there, they will feel that they are the optimal choice.

A second messenger recognises that they were rejected to participate in the task simply when a message is received stating that the agent was not chosen. The agent then compares the distances between all second messengers - itself included - and the package. If it is the closest, it adds a value of 20 to Anger. Otherwise a value of 10 is added to Sadness.

5.4.1. Expectation

With a third agent present, interaction between agents has more ways to become chaotic. Unlike before, now it may be the case that an expression of communicative emotions - anger - isn't directed at you. Agents would need a way to tell the difference. Similarly, the reason for an emotion can also be different. Until now Guilt was expressed as increased speed to make up for past failure. Now anger might arise from rejection, to which increased speed isn't an appropriate response. With no robust way to tell them apart, the expectation is that the first messenger will occasionally respond to rejection-based anger with the wrong response.

In Scenario 2, sadness was rarely expressed as dominant. With a cause that does not simultaneously cause Anger or Guilt, agents will likely start displaying Sadness itself again. More generally, negative emotions - all except Happiness - will likely become more prevalent as there are now more causes for them with no new causes for happiness.

5.4.2. Result

With a value $m = 0.90$, the success rate is 75% after 52 iterations.

As expected, the first messenger often misinterprets Anger as him being blamed for failure and reacts with Guilt. In fact, 51.8% of the time, the first messenger is dominated by Guilt, over 38.3% Happiness. In contrast, the other two agents spend 42.2% and 24.0% of their time happy. Taking into account that each of the second messengers only participates in half of the iterations and has no happy impulses during idle time, this suggests that the first messenger should be happier. One way to cope with the added complexity is to add more and more conditional statements governing emotion elicitation, filtering out exactly when a certain cause is happening and if the event applies to you. However, this would soon become hard to implement and also go against the idea that emotion elicitation should be immediate, not something to be arrived at after lots of reasoning. With only two types of agents and no conflicting interests, this scenario is still relatively simple. A better way to handle this might be to look at the techniques for emotion regulation in DBT and model them as a control mechanism.

The second messenger agents had a tendency to stay angry for longer (17.6% and 59.1% respectively). This may be related to the extra potential cause for anger in this scenario, but more likely it has to do with them statistically only spending half their time - plus idle time - actually doing the task and feeling the emotional effect of it. This lowers the happiness/sadness baseline level, allowing anger to stay dominant longer. This is not necessarily a good effect. Beyond conveying displeasure, there is no reason to keep expressing anger. If anger was made to dissipate quicker or the initial increase in value was lower it is not guaranteed to rise above the baseline. Here, too, a better solution might be found in designated emotion regulation to try and keep anger at a reasonable level relative to other emotions.

5.5. Scenario 4 - Emotion Regulation: Opposite Action through facial feedback

With more causes for emotions and less predictable interactions between emotions, it has become hard to balance the emotions in a way where all can play their role without gearing it too much towards one specific situation. As a possible solution, this scenario will test an implementation of the Opposite Action skill described in DBT. To make this possible, another phenomenon needs to be modeled in the agent’s emotion system: facial feedback.

To briefly recap, the facial feedback hypothesis states that not only is body language in humans affected by the emotions they experience, emotional experience is also affected by body language - at least when it comes to facial expressions. As a model of this, the basic emotion system is extended to add a second effect besides decay over time. If the own emotion expression is perceived to be stronger than the actual experience of it - Sadness is 10 while it is expressed with an intensity of 20, for example - the emotion value is increased over time proportional to the difference. This change happens at a rate of 0.05 times the difference, or 5% of the difference, per second, applied before decay. This effect only works for higher expression than experience. Emotions can not be (directly) decreased by facial feedback. This because there is no data to suggest that emotions can be reduced like that, only substituted. A disparity between expression and experience can only happen when the expression has been deliberately changed by a module outside the regular Emotion modules. This is where Opposite Action comes in.

In Linehan’s explanation of the skill, Opposite Action has a broader reach than how it will be used here. She teaches it as a skill to break out of a cycle of negative emotions that perpetuate themselves (Linehan ,2014). Opposite Action can be anything that goes contrary to what the presently felt emotion urges you to do. This can be anything from biting through fear to take a necessary risk (e.g. running through flames out of a burning building), carrying on despite feelings of sadness or depression that want to make you give up, or being nice to your boss despite him being unpleasant to work with. The task the agents perform has little room for decisions about carrying on through adversity or choosing the lesser of two risks. A more task-appropriate interpretation of Opposite Action is to look at whether the duration and intensity of emotion an agent is feeling helps them function better in the long run. An emotion that is much higher than the others overshadows them and blocks out their effects.

The agent now utilises facial feedback as a way to limit the duration of negative emotion. It monitors the levels of emotions in itself and when a negative emotion gets too high compared to Happiness, it tries to change the expression to showing Happiness with the same intensity as the current highest negative emotion. This change competes with the faithful expression of emotion. As Happiness falls further behind, the bid for Opposite Action increases linearly. When the difference is 20, the bid should be at a medium level, equal to the average other action bid.

Non-colour forms of emotion expression, speed, need not be affected by a non-dominant emotion being expressed. A situation can occur where the agent looks happy on the outside but has increased its speed as it is actually dominated by Guilt. The agent is only pretending to be happy in this case, and might still be acting on other action urges. The purpose of faking emotion is to raise it through facial feedback, after which it might become dominant on its own.

As was mentioned before, Opposite Action is actually meant to be used as a second step after considering whether the emotion urges are justified by the facts. That is not implemented here and emotions are assumed to always line up with the facts of the world state.

5.5.1. Expectation

In scenarios 2 and 3, Guilt has proven to be the main dominating negative emotion. Any application of Opposite Action would more likely respond to that than to Sadness. Since Opposite Action’s purpose is to break out of one emotion dominating, it is expected that Guilt will last a shorter time. This should be particularly noticeable in the first messenger. The slight adjustment to how local success is determined will also affect the balance of emotions, so that should be taken into account.

As the agent most dominated by negative emotions, the first messenger will likely benefit the most from Opposite Action and see an increase in time spent happy, but this will come at the expense of the other agents, who will be more prone to anger since the first messenger walks slower on average. The shift from Guilt to Happiness will also decrease the success rate, also affecting the Happiness of the two second messengers in a negative way. However, the expectation is that overall happiness - the average percentage of time spent happy between all agents - will increase, though more weighted towards the first messenger.

5.5.2. Result

To test the effect Opposite Action has on the agent’s emotion states, the scenario was run both with and without the OppositeAction module active in agents. The percentage of time spent experiencing each emotion and the success rate after repeated iteration are displayed below. The first messenger is indeed happier, with a rise of 19%. Anger and Sadness are now also present, which points to more balanced emotion values in which Anger and Sadness can rise above the other emotions, if briefly. Surprisingly, the third agent also sees an increase in Happiness and a decrease in Anger, though the difference is partly made up by an increase in sadness. However, Happiness overall has only increased by 1.9%, which is well within the margin of error.

Overall, anger has decreased sharply. This is unexpected, as the reduced amount of guilt in the first messenger might have led to more frustration in the other agents. It looks like the emotion for the other agents has shifted from anger to sadness, indicating that more failures are shared rather than being one-sided. Though this doesn’t sound like an improvement, exchanging one negative emotion for the other, it is. Anger should come in short bursts, rising and falling quickly. Measured as a time percentage, it should be small.

Sadness is meant to be a more slow-burning, lasting emotion. Thus, lowering Anger in favour of Sadness is indicative of a more emotionally balanced agent. How Opposite Action caused this effect is not entirely clear. Agent 2 seems the only agent to be worse off. Happiness has been decreased by 15% at an exchange of increasing Sadness by 34% and Guilt being introduced at 8% where it didn't exist at all before. As a last observation, as predicted, the success rate has dropped. As the first messenger, who participates in all iterations of the task, nearly halves its time spent experiencing guilt, it also acts on it less often and makes less of an effort to amend for past failure. This leads into an interesting dilemma: what is more important, being happy or being successful?

With / Without OA	**Agent 1**	**Agent 2**	**Agent 3**
Happiness	55.9% / 37.0%	12.1% / 37.3%	36.6% / 28.7%
Sadness	6.8% / 0.0%	63.8% / 29.2	12.9% / 2.0%
Anger	0.5% / 0.0%	16.0% / 33.3%	43.6% / 69.2%
Guilt	36.6% / 62.9%	7.9% / 0.0%	6.8% / 0.0%
Success Rate (*m*=.95)	77.4% (N=53) / 81.5% (N=54)		

Aside from the numerical results, an effect occurred that should have been expected but wasn't. Opposite Action is used when Happiness is too low compared to other emotions, and it increases Happiness over time - or at least slows decay rate considerably. These two mechanics mean that Opposite Action is self-cancelling. If no causes arise to increase negative emotions, the difference with Happiness will quickly shrink as the negative emotion decays and Happiness is increased by facial feedback, soon falling below the threshold where Opposite Action is no longer applied.

Overall, the data supports that Opposite Action as implemented here is an effective way to lead agents into more balanced emotions without drastically impeding their functioning.

6. Discussion

The results from individual experiments were already discussed in their respective sections. As a general observation, many of the most interesting outcomes emerged from unintended interactions between emotions. Linehan does not go into great detail about how two or more emotions affect each other, treating them all as discrete, with their own causes and purpose. The emergent effect of the Happiness (and Sadness) baseline was not predicted by her work. It could be that it is an artifact of treating emotions internally as numerical values. Linehan says emotions are meant to rise and fall. An inability to keep them from becoming so overwhelming that they block each other out fits her definition of emotional dysregulation (Linehan, 2014), so it seems at least plausible that such an emotional baseline is a reasonable abstraction of how healthy emotions interact in humans. The observations in this project of how emotions with different primary functions have different optimal durations and intensities also fits with her definition of what makes emotions properly regulated. Both overreacting to an emotional cue - of which long-lasting Anger in response to task failure was an example - and the absence of an emotion that was called for under the circumstances - lack of Happiness despite a high success rate - are ways emotions might be dysregulated according to Linehan, and they are the same conditions under which the multi-agent system in this project acted suboptimally. The implementation of a technique designed to ameliorate emotional dysregulation, Opposite Action, indeed reduced these imbalances in the multi-agent system. This suggests that the implementation is at least a good model of those parts of Linehan's Theory of Emotional Dysregulation and the effects of DBT that it tries to simulate.

Much work would still need to be done to test the feasibility of this model further. Only four of the twelve basic emotions recognised by Linehan were implemented, and those were in a much simplified form. Further research can widen the model by adding more emotions and testing its interaction with the existing work done, or deepen the interactions by creating new ways to existing emotions are caused and expressed. This would require expanding the task to allow more hooks for emotion elicitation - or move to another world setting altogether. The internal representation of emotions in this project was simple: a numerical value. Future research could focus on finding a better way to structure emotion data, perhaps providing information on how the emotion was caused. A third direction of research could be in tying these systems to an agent-wide personality that controls some of the parameters that were hardwired here (e.g. emotion increase on a certain event; decay times; strength of expression). Finally, if this architecture is to be used in serious games or simulations, it needs the ability to interact with human emotions. Expression was heavily abstracted into colours in this project. Agents would need to be changed to recognise human expression and themselves express in a similar way to humans.

7. Conclusion

Starting with some remarks on the architecture itself, a shortcoming that was found was the lack of good information on timing for modules. The way modules were scheduled for updates made passing modules information on elapsed time meaningless. At first this was no problem as modules were either event-driven - where information on the time of the event *was* provided - or the behaviour defined by the module was not time-dependent. Later modules, and especially the OppositeAction module of Scenario 4, required a way to know how much time had passed at time of update. For that purpose, a central Timing class was created to provide this service, but a more integrated system would have been preferable and would likely be more efficient.

Another place where the architecture could have been improved was in the action requests. In using the method as designed, almost always following pattern was used to request a new action: Check if the previous action is inactive or no longer valid. If either of those is true, cancel the previous action, request the new action and store a reference to monitor progress. This process could have been automated in the agent framework, That would have left less near-identical functionality in many modules and would have made it possible to optimise that much-used pattern in a central location.

In retrospect, the computations required for this project were not as demanding as was expected. Most of the slowdown in the program came from the interface between mind and embodiment. Optimisation then came mostly from sending as little messages as possible, which an existing architecture could have allowed too. To save time and have more support for technical problems, it might have served just as well to use a language like 2APL and its BDI architecture to implement the mind side. In future research, as agents become more complex, this might change.

The question this project aimed to answer was: Does using Linehan’s Theory of Emotional Dysregulation as a model for implementing emotion in an agent lead to behaviour that is analogous to real-life emotional phenomena in humans? Several instances were seen where agents acted in a way that followed similar patterns to human behaviour. Specific imbalances in emotion-driven behaviour shown by agents were analogous to emotional dysregulation in humans. A real-life emotion regulation skill applied to these behaviours was shown to have a desired effect of balancing emotions. These are hopeful signs that a biosocial approach to modeling emotion is feasible and effective. However, the model presented here only looks at a small subset of all aspects of biosocial theory and possible emotion regulation skills. Emotion regulation even in this simplified setting was only partially effective, and it is unclear whether some of these effects are not an artifact of the model’s abstractions. With that in mind, it has to be concluded that there are positive indications that this is a good model of emotions, but that more research and an expansion of the model is needed to give a definitive answer.

Works Cited

Cloninger, C. Robert. "A unified biosocial theory of personality and its role in the development of anxiety states." *Psychiatric developments* 3.2 (1986): 167-226.

Dastani, Mehdi, Christiaan Floor, and John-Jules Ch Meyer. "Programming Agents with Emotions." *Emotion Modeling*. Springer International Publishing, 2014. 57-75.

Dias, Joao, Samuel Mascarenhas, and Ana Paiva. "Fatima modular: Towards an agent architecture with a generic appraisal framework." *Emotion Modeling*. Springer International Publishing, 2014. 44-56.

Duclos, Sandra E., et al. "Emotion-specific effects of facial expressions and postures on emotional experience." *Journal of Personality and Social Psychology* 57.1 (1989): 100.

Izard, Carroll E. "Basic emotions, relations among emotions, and emotion-cognition relations." (1992): 561.

Izard, C. E. "Differential emotions theory and the facial feedback hypothesis of emotion activation: Comments on Tourangeau and Ellsworth's" The role of facial response in the experience of emotion."." (1981): 350.

Linehan, Marsha M. *DBT® skills training manual*. Guilford Publications, 2014.

Linehan, Marsha. *DBT® Skills Training Handouts and Worksheets*. Guilford Publications, 2014.

Lundqvist, Lars-Olov. "The relationship between the Biosocial Model of Personality and susceptibility to emotional contagion: A structural equation modeling approach." *Personality and Individual Differences* 45.1 (2008): 89-95.

van Oijen, Joost. "Cognitive agents in virtual worlds: A middleware design approach." (2014).

Reeves, Mark, et al. "Support for Linehan's biosocial theory from a nonclinical sample." *Journal of personality disorders* 24.3 (2010): 312.

Reisenzein, Rainer, et al. "Computational modeling of emotion: Toward improving the inter- and intradisciplinary exchange." *IEEE Transactions on Affective Computing* 4.3 (2013): 246-266.

Steunebrink, B. R. "The Logical structure of emotions." (2010).