Model for Organizing Neural Interactions and Knowledge Affectively

Part 1 – Understanding

Contents

Introduction

The desire for an agent to be capable of recognizing and emulating emotions is based on the desire for **Understanding**. Users want their emotions toward the agent – positive or negative – to be understood, validated, and reacted to. They also want to understand why the agent acts in certain ways, so they can improve the outcome of their actions towards the agent.

What sort of things might a user want an agent to understand about them?

- "I want to tell them I'm happy when they do something, so they might do it more often."
- "I want to tell them I'm angry when they insult me, so they might do it less often."

What sort of things might a user want to understand about an agent?

- "I want to know what makes them happy, so they'll be more likely to be kind to me in return."
- "I want to know what makes them angry, so I don't incur retaliatory actions from them."

To be capable of any of these things, our agent needs to be capable of identifying and representing emotions. You can't simulate something you haven't defined! An agent being told by a user that certain actions make the user happy doesn't do any good if the agent doesn't know what "happy" is!

The first step on the road toward user/agent emotional understanding isn't planning how to program the agent to act on its feelings, nor developing a system for the agent to manifest feelings in the first place.

It's examining our own human emotions and figuring out how we should classify and categorize them!

Without understanding our own emotions, how can we design a robust model for an agent to follow? Haphazardly developing an extensive cause/effect emotional response system without first determining a rugged emotional lexicon could lead to later pitfalls when encountering situations that go against our initial assumptions.

Due to the complex nature of human emotion there will no doubt be situations that stretch the limits of even the most extensive designs. But a thoughtful exploration before we begin – with an eye towards our final needs for our agent – can result in an emotional classification system least likely to fail.

Our chosen emotion classification system needs to be:

1. **Distinct** enough that emotions arising from **negative events can be tied to their root cause**. An agent that's angry over being insulted should continue to be angry until it receives an apology.
    - Contrast: A simple "`Happy`/`Sad`" meter which can be "repaired" with an unrelated compliment
        - User: "You're stupid."
        - Agent: "Hey! I spend a lot of time studying!" (`Happy` – 1 = -1)
        - User: "I love your hair!"
        - Agent: "Yay! I'm glad you like it!" (`Happy` + 1 = 0)
2. **Capable of being combined or felt simultaneously**, allowing complex emotions or expression of those emotions. An agent fondly remembering time spent with a deceased friend feel fondness toward their memories, yet `Sad` over not being able to make more.
    - Contrast: An "emotional axes" model incorrectly defining emotions as opposites of others
        - `Anger`/`Joy` denies spitefully harming another
        - `Fear`/`Trust` denies riding roller coasters or haunted house attractions
3. **Expressive** enough to allow for sufficient nuance to simulate more natural behaviors
    - An agent being told they make a user feel stupid should feel:
        - `Guilt` from making the user think less of themselves
        - `Worry` that the user dislikes them
4. **Low to moderate in breadth**, so as to not be confusingly overwhelming
    - Contrast against: Dwarf Fortress which names 150+ unique emotions
5. **Interconnected** so changes in one emotion **automatically influences** or potentially overpower other emotions
    - Sufficient amounts of `Guilt` should have an effect on `Pride`, without the developer manually decrementing `Pride` on every `Guilt` event.
    - Low self-confidence – the result of `Guilt`, `Worry`, `Anxiety`, and low `Pride` – should contribute to `Sadness`, without manually checking all four on every `Guilt` event.
6. **Applicable for helping agents understand users** through the same emotional classification system, so that they can better adjust actions to meet goals
    - The agent knowing the user feels `Anger` over an agent's action allows it to learn that the action induces `Guilt` upon itself. By avoiding the negative `Guilt`, the agent avoids the action which causes the user `Anger`.
    - **Note:** This document won't describe user understanding or agent learning, leaving these as exercises for the future. But it is helpful to keep the possibilities in mind.

The number of existing emotional classification models is staggering. This stems from wanting to identify all unique emotions – while fearing the inclusion of synonyms for the same emotion.

As a practical matter, we should choose a classification system which doesn't waste developer time by forcing them to constantly reference a "valid list" of emotions:

- A list that's too large results in time spent checking the minute differences between emotions
- A list that's too small results in time wasted determining which "allowed" emotion is most appropriate

Ideally, a developer will use whatever word for an emotion which first comes to mind, with the larger emotion system handling the rest "automagically". But such systems would still require a defined set to translate to.

Note: In a perfect world, developers wouldn't be defining emotional outcomes at all – the outcome would be automatically output from events exiting through the agent's perceptions and personal values systems. But that's currently outside the immediate scope of us giving simple agents the basics of emotion.

Existing Psychological Emotional Models – A Review of Possibilities

We are not emotional psychologists.

Any naive attempt by us to establish an emotional model, or even a basic dictionary, would likely have numerous holes and oversights, miring future endeavors to make our agent more emotive and autonomous.

Therefore, instead of bouncing around words, deliberating, and trying to build a list from whatever words came to mind, we should seek guidance from those who *are* emotional psychologists, with published – and reviewed – models.

Before beginning our literature review, it is worth restating that our purpose is developing an emotive agent.

Existing emotional psychology models focus on humans, who are complex both biologically and socially. Some papers by computer scientists who employ and evaluate these models have thus far focused on determining emotion from raw text. This is a different task than an agent whose actions are influenced by the emotions we want it to experience. Any existing model isn't necessarily "wrong", but may not be fit for our purpose.

From this onset, due to our desires for **distinctiveness** and **expressiveness**, early emotional models which classify emotions into very few singular categories should be eliminated.

Additionally, due to our desire for **combinability**, any model relying on strict axes (where "polar" emotions are opposite ends of a shared negative-to-positive scale) should be carefully evaluated for accuracy.

Emotional classification models we will be briefly introducing in search of a potential starting point are:

- Parrott – Emotions in Social Psychology (2001)
- HUMAINE Emotion Annotation and Representation Language (2006)
- Plutchik Emotion Wheel (1988)
- OCC Model (1988) / Revisted OCC Model (2008)
- Hourglass of Emotions (2012) / Revisited Hourglass Model (2020)

Parrott - Emotions in Social Psychology (2001)

Heavily based on "Emotion Knowledge: Further Exploration of a Prototype Approach" (Shaver et al.), Parrott's tree-structured list has suitable **distinctiveness**, and easily meets our requirements for **expressiveness**. While it doesn't give **combinability** guidance on its own, it isn't incompatible with it. It also gives some basic guidance on **interconnection –** more explicit emotions can "bubble up" as parent emotions.

Primary emotion	Secondary emotion	Tertiary emotion
Love	*Affection*	*Adoration · Fondness · Liking · Attraction · Caring · Tenderness · Compassion · Sentimentality*
	Lust/Sexual desire	*Desire · Passion · Infatuation*
	Longing	*Longing*
Joy	*Cheerfulness*	*Amusement · Bliss · Gaiety · Glee · Jolliness · Joviality · Joy · Delight · Enjoyment · Gladness · Happiness · Jubilation · Elation · Satisfaction · Ecstasy · Euphoria*
	Zest	*Enthusiasm · Zeal · Excitement · Thrill · Exhilaration*
	Contentment	*Pleasure*
	Pride	*Triumph*
	Optimism	*Eagerness · Hope*
	Enthrallment	*Enthrallment · Rapture*
	Relief	*Relief*
Surprise	*Surprise*	*Amazement · Astonishment*
Anger	*Irritability*	*Aggravation · Agitation · Annoyance · Grouchy · Grumpy · Crosspatch*
	Exasperation	*Frustration*
	Rage	*Anger · Outrage · Fury · Wrath · Hostility · Ferocity · Bitterness · Hatred · Scorn · Spite · Vengefulness · Dislike · Resentment*
	Disgust	*Revulsion · Contempt · Loathing*
	Envy	*Jealousy*
	Torment	*Torment*
Sadness	*Suffering*	*Agony · Anguish · Hurt*
	Sadness	*Depression · Despair · Gloom · Glumness · Unhappiness · Grief · Sorrow · Woe · Misery · Melancholy*
	Disappointment	*Dismay · Displeasure*
	Shame	*Guilt · Regret · Remorse*
	Neglect	*Alienation · Defeatism · Dejection · Embarrassment · Homesickness · Humiliation · Insecurity · Insult · Isolation · Loneliness · Rejection*
	Sympathy	*Pity · Mono no aware · Sympathy*
Fear	*Horror*	*Alarm · Shock · Fear · Fright · Horror · Terror · Panic · Hysteria · Mortification*
	Nervousness	*Anxiety · Suspense · Uneasiness · Apprehension · Worry · Distress · Dread*

However, coming in at 136 emotions at its full tertiary depth, it "fails" our requirement for reasonable **breadth**. But we can still explore it for guidance on "parent" classifications, and potentially only view it at its secondary depth.

HUMAINE Emotion Annotation and Representation Language (2006)

The Emotion Annotation and Representation Language, proposed by the Human-Machine Interaction Network on Emotion has the express purpose of developing a standard that can used to exchange data between different emotional models that are employed in different use cases. Using 48 categories, it meets our **distinctiveness** and **expressiveness** requirements, but is on the upper end of what might be an allowable **breadth**.

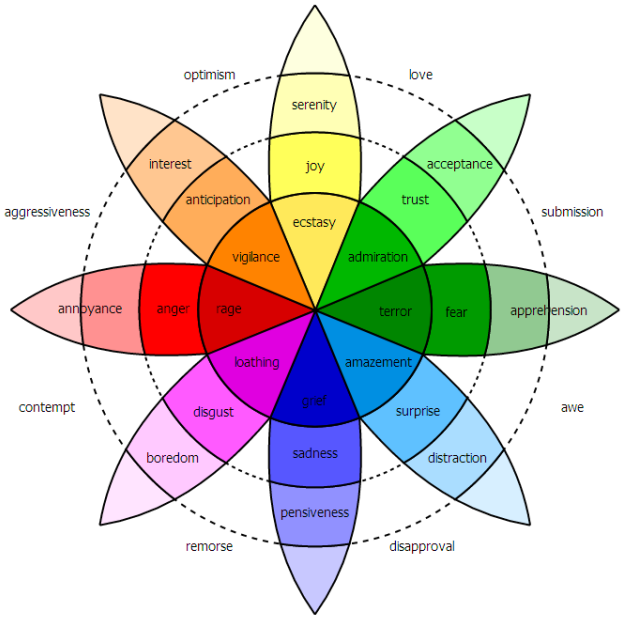
Positive (lively) Amusement Delight Elation Excitement Happiness Joy Pleasure	**Caring** Affection Empathy Friendliness Love	**Positive (thoughts)** Courage Hope Pride Satisfaction Trust	**Positive (quiet)** Calm Content Relaxed Relieved Serene	**Reactive** Interest Politeness Surprise
Negative (forceful) Anger Annoyance Contempt Disgust Irritation	**Negative (uncontrolled)** Anxiety Embarrassment Fear Helplessness Powerlessness Worry	**Negative (thoughts)** Doubt Envy Frustration Guilt Shame	**Negative (passive)** Boredom Despair Disappointment Hurt Sadness	**Agitation** Shock Stress Tension

Due to its design goal of being similar to a "Rosetta Stone" between different aspects of affective computing, it is worth our consideration in the context of a "general purpose" emotional model for our agent.

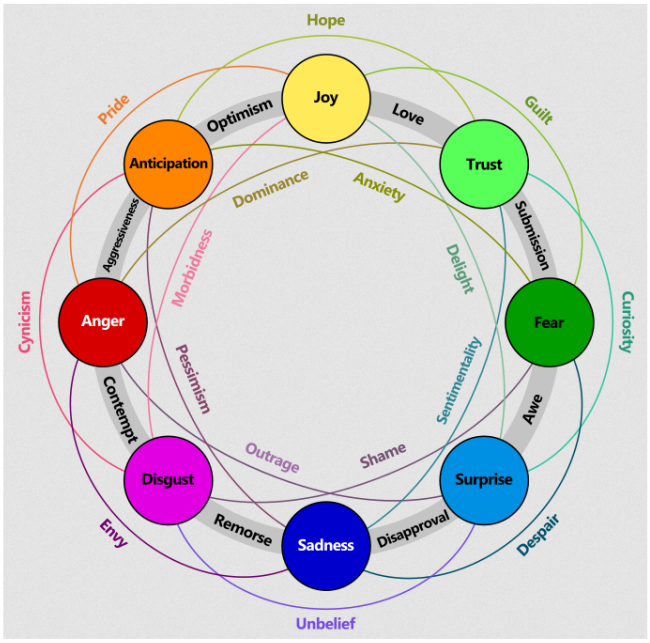
However, cursory examination of its final list reveals potential for confusion. What's the difference between `Shock` and `Surprise`? Between `Helplessness` and `Powerlessness`? EARL might serve as a good starting list, but its potential redundancies should be examined and pared down after further examination and comparison against other models.

Plutchik Emotion Wheel (1980)

"Plutchik's wheel", a very popular model in emotional psychology, has less default **expressiveness** than the aforementioned EARL, but still has suitable **distinctiveness**. We should give it additional consideration for it being "anchored" in **Combinability**, and also **Interconnection** when its "tertiary" emotions are examined.



Plutchik Core Emotion Wheel



Plutchik Wheel Emotion Combinations

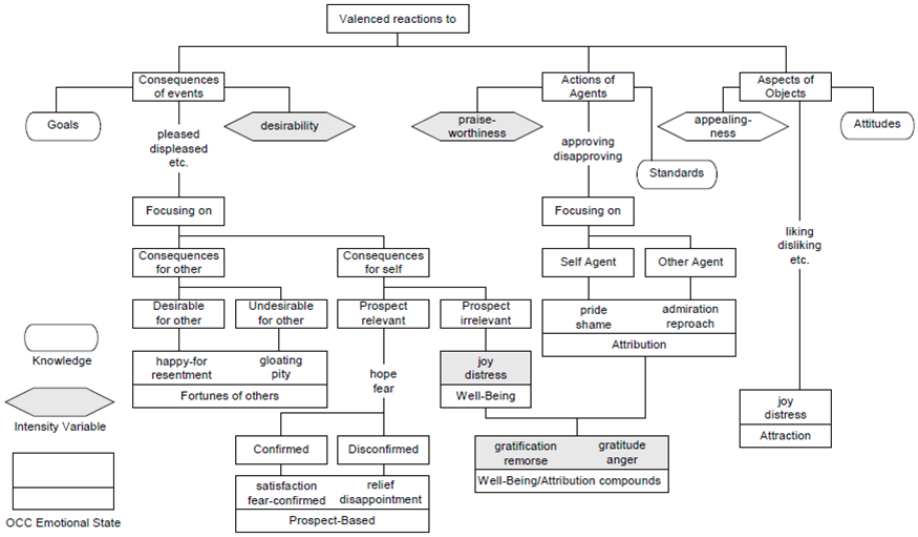
Despite its age it is still very commonly accepted, as illustrated by its appearence in EmoMix: Building an Emotional Lexicon for Compound Emotion Analysis (2019), which uses Plutchik's primary emotions and compound dyads to assign combinatoric emotion scores to individual words, which can be summed from a writing excerpt to evaluate the emotions of the author.

The work shown by "EmoMix" – emotional sentiment analysis – merits special consideration. A basic template for the capability to "measure" emotion from words may allow additional development avenues in the future when it comes to user/agent interaction. Intended dialogue from the agent could be with the combined with Facial Action Coding System to form agent facial emotes without manual definition by the dialogue author. Additionally, an agent might be able to sense the overall emotion from any freeform input from the user. This document won't speculate on how such systems would be implemented, but using an emotion classification system that's suitable for sentiment analysis would be a good fundamental step.

Finally, the "Plutchik Wheel" has a very suitable **breadth**. In open projects with multiple developers with varying skillsets, its more-compact classification is very appealing.

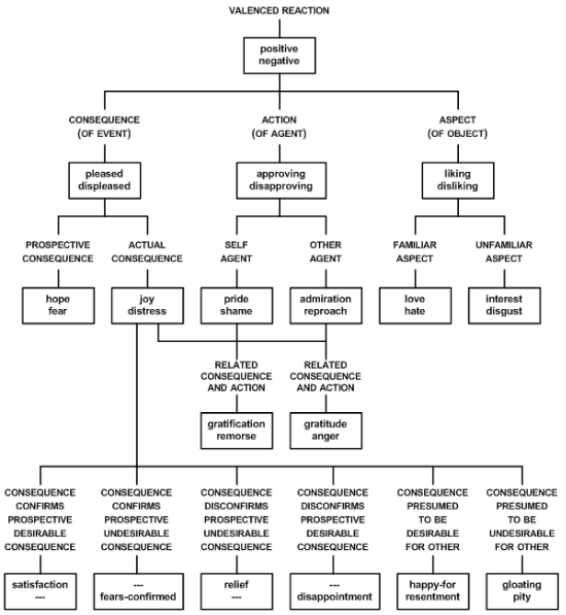
"OCC Model" - The Cognitive Structure of Emotions (1988)

The emotion model proposed by Ortony, Clore, and Collins examines how emotions are elicited, what affects their intensities, and how they might be influenced. Instead of exhaustive word lists or emotional combinations, it focuses more on the logical flow of emotional responses in context of the agent's relationship with another entity.



This is of special interest to us as developers, as it might lead to avenues of allowing the agent to define and differentiate *its own* emotional responses in response to events – based on its goals and values – as opposed to a developer dictating what emotion and intensity an agent should express. While such endeavors are outside our immediate scope of having an agent *with any* emotions, using an emotional classification that's "compatible" with the OCC Model might be beneficial in the future, as the agent could be further expanded without having to redefine and use a new list of "core emotions".

This model was revisited, and revised, by computer science researchers at Utrecht University. Their full paper, "The OCC Model Revisited" (2009), standardizes the full logical model and forms intrinsic links (**interconnectivity**) between type specifications, resulting in natural descriptions when child nodes are followed up to their parent nodes. For example, `Pity` is `Distress` over an undesirable event for another.



positive and *negative* are valenced reactions (to "something")
pleased is being *positive* about a consequence (of an event)
displeased is being *negative* about a consequence (of an event)
hope is being *pleased* about a prospective consequence (of an event)
fear is being *displeased* about a prospective consequence (of an event)
joy is being *pleased* about an actual consequence (of an event)
distress is being *displeased* about an actual consequence (of an event)
satisfaction is *joy* about the confirmation of a prospective desirable consequence
fears-confirmed is *distress* about the confirmation of a prospective undesirable consequence
relief is *joy* about the disconfirmation of a prospective undesirable consequence
disappointment is *distress* about the disconfirmation of a prospective desirable consequence
happy-for is *joy* about a consequence (of an event) presumed to be desirable for someone else
resentment is *distress* about a consequence (of an event) presumed to be desirable for someone else
gloating is *joy* about a consequence (of an event) presumed to be undesirable for someone else
pity is *distress* about a consequence (of an event) presumed to be undesirable for someone else
approving is being *positive* about an action (of an agent)
disapproving is being *negative* about an action (of an agent)
pride is *approving* of one's own action
shame is *disapproving* of one's own action
admiration is *approving* of someone else's action
reproach is *disapproving* of someone else's action
gratification is *pride* about an action and *joy* about a related consequence
remorse is *shame* about an action and *distress* about a related consequence
gratitude is *admiration* about an action and *joy* about a related consequence
anger is *reproach* about an action and *distress* about a related consequence
liking is being *positive* about an aspect (of an object)
disliking is being *negative* about an aspect (of an object)
love is *liking* a familiar aspect (of an object)
hate is *disliking* a familiar aspect (of an object)
interest is *liking* an unfamiliar aspect (of an object)
disgust is *disliking* an unfamiliar aspect (of an object)

Revisited "Hourglass Model" (2020)

The "hourglass model", originally published in "The Hourglass of Emotions" (Cambra et al., 2012), was based on the "Plutchik Wheel" model. However, it defined emotions into four dimensions, contrasting "positive" emotions like `Joy` against negative emotions like `Sadness`. Emotions were mapped to matrixes, where "compound" emotions like Melancholy end up on the "negative" side of the Introspection scale, which ranges from `Ecstasy` to `Grief`.

Ordinarily, we would be wary of this model – it relies on polar axes, endangering our desire for **combinability**.

But "The Hourglass Model Revisited" (Susanto et al., 2020) used a modified hourglass model and compared it to other existing models in the context of sentiment analysis and affective computing. By replacing the normalization factor in the older model, they were still able to account for compound emotions like `Bittersweet`, despite employing a polar model, by accounting for emotions across axes simultaneously.

Additionally, by adding sub-dimensions under Attitude to encompass evaluations towards self and others, the model can differentiate emotions like `Pride` and `Shame` which appear in the OCC Model, and even `Friendliness` which is explicit in HUMAINE's EARL.

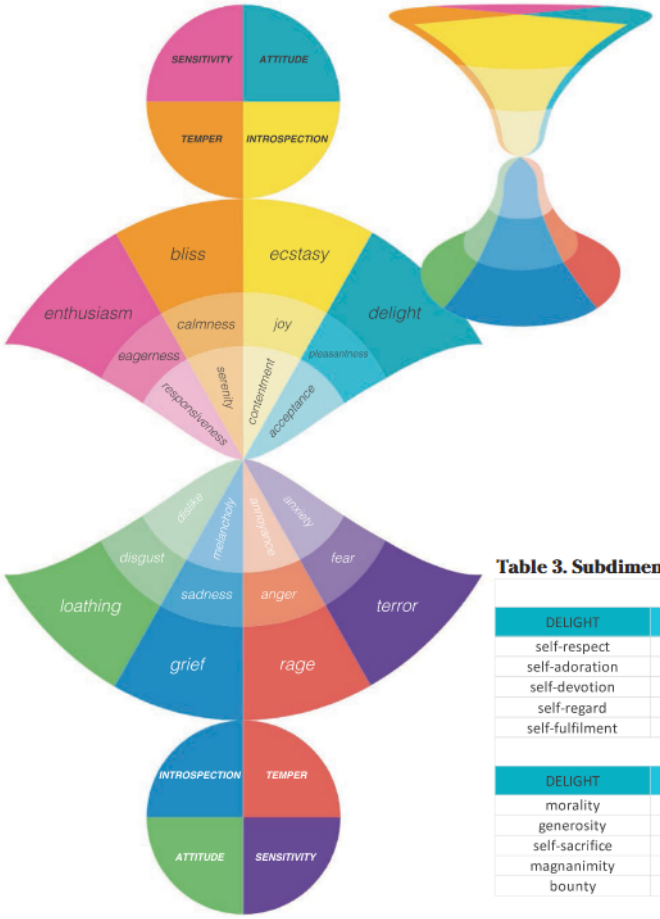


Table 1. Examples of compound emotions.

JOY	PLEASANTNESS	love	enjoyment	amusement
	EAGERNESS	euphoria	excitement	thrill
	CALMNESS	enlightenment	relaxation	sweet idleness
SADNESS	DISGUST	hate	guilt	remorse
	FEAR	distress	troubledness	misery
	ANGER	envy	bitterness	resentment
CALMNESS	PLEASANTNESS	assertiveness	compassion	empathy
	EAGERNESS	focus	determination	perseverance
	FEAR	carelessness	laxity	looseness
ANGER	DISGUST	hatred	ruthlessness	viciousness
	FEAR	nastiness	coercion	possessiveness
	EAGERNESS	stubborness	obstinacy	mulishness
PLEASANTNESS	DISGUST	shamelessness	cheekiness	brazenness
	EAGERNESS	kindness	audacity	hospitality
	FEAR	awe	submission	reverence
DISGUST	JOY	morbidness	schadenfreude	gloat
	FEAR	impiety	cowardness	inhospitality
	EAGERNESS	recklessness	temerity	rashness
EXPECTATION	JOY	hope	anticipation	optimism
	SADNESS	hopelessness	despair	pessimism
	EAGERNESS	vigilance	alertness	caution
SURPRISE	ANGER	shock	outrage	thunderstruckness
	FEAR	alarm	dismay	dumbstruckness
	PLEASANTNESS	amazement	astonishment	wonderstruckness

Table 3. Subdimensions of Attitude with five sample emotion words per category.

ATTITUDE (toward self)					
DELIGHT	PLEASANTNESS	ACCEPTANCE	DISLIKE	DISGUST	LOATHING
self-respect	pride	confidence	low-confidence	shame	self-contempt
self-adoration	self-appreciation	security	insecurity	self-blame	self-loathing
self-devotion	self-attraction	modesty	embarrassment	self-disgust	self-abasement
self-regard	self-formation	self-esteem	low-self-esteem	disgrace	self-denigration
self-fulfilment	self-motivation	assurance	self-deprecation	self-pity	self-condemnation
ATTITUDE (toward others)					
DELIGHT	PLEASANTNESS	ACCEPTANCE	DISLIKE	DISGUST	LOATHING
morality	sociability	sympathy	antipathy	asociability	immorality
generosity	appeasement	fairness	unfairness	greed	malevolence
self-sacrifice	affability	humbleness	prejudice	meanness	turpitude
magnanimity	conviviality	humility	hostility	humiliation	wickedness
bounty	friendliness	gratitude	ingratitude	unfriendliness	xenophobia

With this revised model they analyzed emotional sentiment across multiple text datasets and achieved an accuracy that outperformed any previously existing model, including both the "Plutchik Wheel" and the "Revisited OCC Model". But this analysis was for the purpose of measuring emotional sentiment in text; which is neither our immediate nor near-future goal. Since our goal is in the context of creating an emotive agent, something more akin to "Revisited OCC" may be more suitable.

Emotional Classification Cross-Examination

When it comes to selecting an emotional model, it helps to keep our ultimate goals in mind.

We want, in order of preference:

1. An agent that expresses its current emotional state(s)
2. A method for informing the agent how it might feel about events
3. A system for the agent to take all this emotional input, interpret it, and result in realistic expressions
4. A logical flow for the agent to be able feel its own emotions based on its own goals
5. A way to remember what actions led to positive or negative emotions
6. A classification system with which the agent can understand the actions by a user

As examples when these are all together:

- If we have an agent with the goal of ingratiating themselves with the user, should the user express unhappiness with an agent, the agent should feel unhappiness
- If that same agent is informed the user is happy with them, the agent should feel happiness, and possibly associate whatever actions it has undertaken as methods of maintaining that happiness
- However, if we have an agent with the goal of opposing the user, the agent should feel positively about events which hinder or upset the user

With those ultimate goals – even if we can't achieve them all at once – it's clear from the onset that the "Revisited OCC Model" should be closely examined. While the "Revisited Hourglass Model" can parse text for emotional context more accurately, our primary goal isn't understanding freeform text input by the user.

But for the purpose of thoroughness, we may still attempt to directly compare and contrast the individual emotions across our reviewed models, looking for hierarchies and what elements we can safely combine from multiple models.

Even should "Revisited OCC" later be chosen for the purpose of logical flow, the internal representations - "point values" - might still follow the four-dimensional axis model of the "Revisited Hourglass".

The "Revisited Hourglass" is also a very strong model for fitting emotions which may appear in other models in context of itself. For example, `Regret` is explicitly listed in Parrott's Tree, but isn't explicitly named in our other models. However, if fit into the Attitude sub-dimension in the "Revisited Hourglass", it can be a `Dislike` towards the Self.

We shall begin by attempting to "pare down" the other models to our expected goals of the "Revisited OCC" and "Revisited Hourglass". This will be a "correct by consensus" approach, and gives us a basis to point back to justify the final model we end up using.

Parrott's Tree

Parrott's Tree is by far the largest explicit emotional vocabulary model of those we've reviewed – too large. We will begin by determining what might be "superfluous" in it by comparing it to other models. If another model also lists the same explicit emotion as Parrott, it shall be retained. But if an explicit emotion in Parrott doesn't exist in another model, we will attempt to either "reduce" it to an appropriate parent inside Parrott (should the parent appear in another model), or find what we "feel" is a suitable similarity from another model.

Unfortunately, short of undertaking the process of building emotional lexica as seen in "EmoMix", any replacements chosen are by judgment, and may be subject to debate.

The results after first comparing all listed emotions in Parrott's tree results in the following emotions that aren't *explicitly* accounted for among EARL, Plutchik, OCC, or Hourglass:

Adoration - Aggravation - Agitation - Agony - Alienation - Anguish - Attraction - Crosspatch - Defeatism - Dejection - Depression - Desire - Dread - Enthrallment - Exasperation - Exhilaration - Ferocity - Fondness - Fright - Fury - Gaiety - Gladness - Glee - Gloom - Glumness - Grouchy - Grumpy - Homesickness - Horror - Hostility - Hysteria - Infatuation - Insult - Isolation - Jolliness - Joviality - Jubilation - Loneliness - Longing - Lust - Mono no aware - Mortification - Neglect - Nervousness - Panic - Passion - Pity - Rapture - Regret - Rejection - Revulsion - Scorn - Sorrow - Spite - Suffering - Suspense - Tenderness - Torment - Triumph - Uneasiness - Unhappiness - Vengefulness - Woe - Wrath - Zeal - Zest

We can already tell that some of these are already accounted for under synonyms, which Parrott's Tree seems especially prone to.
Examples:

- `Fury` is a basic synonym of `Rage`, of which both appear in Parrott's Tree
- `Pity` is the same emotion as `Sympathy`, only with negative connotation
- `Zest`, while listed as a parent of `Enthusiasm`, is a synonym of it

After reducing emotions to their parents or using synonym emotions, we are left with the following emotions not explicitly listed in another emotion model.

Parent Emotion	Remaining Parrott Emotions
Cheerfulness	*Gaiety · Glee · Jolliness · Joviality · Gladness · Jubilation*
Enthrallment	*Enthrallment · Rapture*
Horror	*Fright · Panic · Hysteria · Mortification*
Lust	*Desire · Passion · Infatuation*
Neglect	*Alienation · Defeatism · Dejection · Homesickness · Insult · Isolation · Loneliness · Rejection*
Suffering	*Agony · Anguish · Hurt*

We will now examine these emotions in context of the Revisited Hourglass model as it offers the best "fit" for translating emotional sentiment.

- The emotions in the `Cheerfulness` category can be easily replaced with varying degrees of `Joy`, with the exceptions of `Gladness` and `Elation.`
  - `Elation` implies the agent had previously predicted a negative outcome, so should most likely carry amounts of both `Joy` and `Surprise.` This is similar to `Relief` from the OCC model, but possibly with a higher valence of `Joy` than `Relief` might have.
  - `Gladness` might imply both `Joy` and `Surprise` ("glad things ended up going well"). However, it can also imply both `Joy` and `Bliss` ("glad with life"). These different meanings can lead to confusion, and suggest Gladness should be removed from our lexicon altogether.
- `Enthrallment` and `Rapture` would be most easily synonym-translated to `Amazement`, which is a compound between `Surprise` and `Pleasantness` in the Revisited Hourglass model.

- The subsidiary emotions in `Horror`, which at face value all appear to be synonyms of `Fear`, require some delineation. `Fright` is a sudden `Fear`, implying a `Surprise` component. When a person is `Mortified`, they're frozen and unable to act, while when they are in a `Panic` or `Hysteria`, the problem is that they're acting too much. Someone may be in hysterics from too much `Grief`, too much `Anger`, or a combination of the two. In either event, these differing combinations should be eliminated from our base classification.
- `Lust` is a major problem in context of our purpose for developing an emotive agent.
  - `Lust` is not the same thing as `Affection`. People are `Affectionate` towards family or children in their care. They are – hopefully – not `Lustful` towards them.
  - Even in relationships in which it is appropriate to be `Lustful`, the amount of `Affection` a person has for another in a given moment isn't going to be the same amount of `Lust`
    - Friends may be `Affectionate` towards each other, but develop `Lust` over time
    - People who have recently begun dating may have more `Affection` than `Lust` at the beginning of their relationship ("innocent dating"), or vice versa ("bar hookup")
    - Even in couples that have been together for prolonged periods of time, the amount of `Affection` they feel doesn't change, but `Lust` can come and go, possibly varying widely from day-to-day
  - It's therefore necessary to keep `Lust` as its own distinct emotion. This will be our first major break from our desire to have our final classification be heavily based on OCC and/or Revisited Hourglass. But we make this concession because:
    - We require a way to distinguish between an agent that has a "schoolgirl crush" from one that desires a sexual relationship; and
    - `Lust` does exist on its own in a published psychological model (Shaver, Parrott)
      - However, the sub-categories of it are superfluous, and will not be included in our final model
- `Neglect`, `Suffering`, and all their sub-emotions can be mapped to `Sadness` – they are "`Sadness` for cause"
  - `Suffering` is being in a prolonged stage of `Grief`; you only feel more `Sadness` because your existing `Grief` has yet to fade
  - `Neglect` and its sub-emotions are noteworthy for having a social component; they're all forms of social rejection. It may be worth considering them in context of the OCC Model's `Fear-Confirmed`; one feels `Anxiety` over the fear of another not appreciating them, and then that fear is confirmed when they are `Rejected`. However, at their core, they're still either `Sadness` (Hourglass), `Sadness` compounded with another emotion (`Anger`, in context of `Insult`), or `Fear-Confirmed` (OCC) in response to an event, and need not be kept distinct in our final model.

This examination has left us with only one emotion – `Lust` – which we will keep from Parrott's model, as it's not accurately represented by another model.

EARL

We now turn our attention to EARL, which we hope to further refine to elements of either Plutchik, Revisited OCC, or Revisited Hourglass.

Matched in Plutchik / OCC / Hourglass		Match By Synonym	Parrott Only	Unmatched
Amusement	Guilt	Helpnessness > Hopelessness	Affection	Courage
Anger	Hope		Elation	Doubt
Annoyance	Hurt	Politeness > Friendliness	Embarrassment	Humility
Anxiety	Interest		Frustration	Stress
Boredom	Joy	Powerlessness > Hopelessness	Happiness	
Calmness	Love		Hurt	
Contempt	Pride	Tension > Anxiety	Irritation	
Contentment	Relaxation		Pleasure	
Delight	Relief		Worry	
Despair	Sadness			
Disappointment	Satisfaction			
Disgust	Serenity			
Empathy	Shame			
Envy	Shock			
Excitement	Surprise			
Fear	Trust			
Friendliness				

By examining the emotions is EARL which appear *only* in Parrott, we can determine emotions which are potentially duplicated.

- `Affection` can be mapped as a precursor to `Love`. In common usage they aren't always the same: A person who `Love`s another may not always feel `Affection`ate. However, that situation is usually in conjunction with another emotion hindering the desire to be `Affection`ate.
- As discussed when examining Parrott, `Elation` is comparable to `Relief`.
- `Embarrassment` is difficult to place, as it can arise in many different situations. Someone can feel `Embarrassment` over another recounting an embarrassing story, over a flub while speaking in public, or telling a person how they feel. These different scenarios may not always have the same combination of "base" emotions (in Plutchik or Hourglass), and implies `Embarrassment` is a base emotion we should consider keeping. This is especially important for the specific type of agent we have in mind – they may feel embarrassed over being told something, or shy while saying something.
- `Frustration` can be mapped to `Annoyance`.
- `Happiness` can be mapped to `Contentment`, `Joy`, or `Delight`.
- `Hurt` is far too broad. The word could be used in situations of `Despair`, `Disappointment`, `Dismay`, `Distress`, `Misery`... the list goes on. It should be eliminated, and a more specific emotion used in its place in response to events.
- `Irritation` can be mapped to `Annoyance`.
- `Pleasure` can be mapped to `Pleasantness`.
- `Worry` can be mapped to `Anxiety`.

This leaves the following emotions which appear in EARL, but do *not* appear in any of our other reviewed models.

- `Courage` is noteworthy. It is not the opposite of `Fear`, as `Courage` is defined by action *despite* `Fear`. While appearing as `Courage` in EARL, this emotion is actually `Trust` or confidence in oneself. `Trust` appears explicitly in the Plutchik model, and `Confidence` appears in Revisited Hourglass under its sub-dimension of Attitude (`Acceptance` towards oneself)
- Similarly, `Doubt` is the inverse of having `Trust` or confidence in either oneself or another.
- On the subject of lack of `Trust` in oneself, we can return to examining `Embarrassment` in the same context. When one is shy while saying something, they already trust the other party with the story. The shyness is instead marked by lack of confidence in oneself. We can therefore discard `Embarrassment` as its own unique emotion. `Embarrassment` is lack of confidence in oneself – either that they will not be perceived positively by others, or a worry that they'll fail to adequately convey something, or fail to adequately perform or respond to something.
- `Humility` – not to be confused with `Humiliation` – is the state of being humble. However, one never *feels* "humble".
  - One doesn't say to themselves, "Oh, I feel humble right now". That sort of self-evaluation is antithetical to the trait – the phrase "I'm so modest" is always said in jest.
  - When a person is actually *acting* humble, they're really feeling another emotion (most likely lack of confidence, or expressing `Remorse`), and *another person* is describing them as "humble".
  - Further, while there is a one letter difference between "humble" and "humbled", they aren't the same in the context of emotions. When one is "humbled", they're feeling `Shame` or `Remorse` in direct consequence to their past actions.
  - Therefore, "humility" should be removed from consideration in our emotional model altogether – it's not actually an emotion felt by a person, but is instead a descriptor used by others.
- `Stress` isn't a unique emotion in of itself, but is a state of having felt other emotions over an extended period of time: `Anger`, `Anxiety`, `Annoyance`, `Contempt`, `Frustration`, `Tension`, or `Worry`. Saying, "I feel stressed" is a shorter way of feeling, "I've felt a combination of negative things that I'm not going to explicitly describe". The closest mapping of `Stress` would be to long-term `Anxiety`.

Ultimately we have no new unique emotions described in EARL which either aren't directly mapable to an emotion described by Plutchik or Revisited Hourglass, or a combination thereof.

However, we have confirmed the importance of our final model including a dimension for confidence in oneself. This dimension was missing in Plutchik and the original Hourglass model. It was added in the Revisited Hourglass, and has been validated during our examination of EARL's `Courage` and `Doubt`. As these two emotions requires subjects to be "felt against" (confidence in another is `Trust`, while confidence in oneself is `Courage`), it's important that our final design considers and *expects* emotions to be tied to subjects, instead of just unattached values which are assumed to apply only in context of our emotive agent.

OCC (Revisited)

We already have a preference for the OCC model as it's a programmatic model on how agents respond to events. While our first goal is determining how we want to *describe* and *dictate* the emotions an agent feels, should we want to extend our emotional model to give the agent some amount of self-determination, the legwork of cause/effect flow would already be done for us.

Further, OCC is of a much more reasonable size in comparison to our as-of-yet examined models. Ideally, we'd like to mostly retain the end results of its cause/effect design (it's "flowchart"), while ensuring that the emotional vocabulary it uses can be mapped to either Plutchik or Hourglass.

Finally, as we have already "pared down" two models – Parrott and EARL – and only have our "goals" of either Plutchik or Hourglass left, we should consider a preliminary reduction of our final two models. Rather than just trying to remove "superfluous" emotions that might appear only in OCC, we should also try to eliminate or translate emotions which appear in only one other model, *including* Plutchik and Hourglass.

OCC Only (1 total)	OCC + One Other Model (2 total)	OCC + Multiple Models (3+ total)
Approving Fears-Confirmed Gratification Happy-For Pity Pleased Reproach	Admiration (Plutchik) Disapproving (Plutchik) Displeased (Parrott) Gratitude (Hourglass) Gloating (Hourglass Compound) Interest (Plutchik) Liking (Parrott)	Anger (5 total) Disgust (5 total) Distress (3 total) Disliking (3 total) Fear (5 total) Hate (3 total) Hope (5 total) Joy (5 total) Love (5 total) Pride (5 total) Relief (3 total) Remorse (4 total) Resentment (3 total) Satisfaction (3 total) Shame (5 total)

From listing our current emotional lists out above, it's clear that we've already some "true", "set in stone" emotions for our final lexicon – those which appear in 4 or more different models, like `Anger`, `Fear`, and `Hope`.

Additionally, we can examine the emotions that appear in only 3 models, compare them with emotions which appear in just one or two models, and see if the two "different" emotions are really the same under two synonyms.

Fortunately, due to the programmatic design of OCC, a lot of the "emotions" its lists can be reduced as they are *context* descriptors and not true emotions. Looking back at the revisited flowchart:

- "`Relief` is `Joy` about the disconfirmation of a prospective undesirable consequence"
- "`Pity` is `Distress` about a consequence presumed to be undesirable for someone else"

However, if taken to extreme, we'd end up eliminating most if not all emotions – they all bubble up to `Positive` or `Negative` eventually! As an example:

- "`Anger` is `Reproach` about an action and `Distress` about a related consequence"
- "`Reproach` is `Disapproving` of another's action"
- "`Disapproving` is being `Negative` about an action."

While it could be programmatically useful to have these defined links, we have to keep in mind that we're not making a fully autonomous agent just yet! As we're trying to determine a model that can be used to sufficiently *describe* emotions, developers still need enough of an emotional lexicon without having to memorize the OCC "emotional flowchart"!

`Pleased`, `Approving`, and `Liking` are all ways to say that some event, action, or aspect gives you `Positive` feelings. We don't need three different ways of describing feeling positively about something: you can "approve" of an action someone does, and also "approve" of some aspect about them. You can "like" an aspect about them, while "liking" something they do. Different terms based on the *context* of the positive feeling are superfluous.

The inverse applies for `Displeased`, `Disapproving`, and `Disliking`.

Fortunately, the "Revisited Hourglass" already addresses these under its polar Attitude component: `Loathing` <> `Disgust` <> `Dislike` <> `Neutral` <> `Acceptance` <> `Pleasantness` <> `Delight`

This lets us directly translate `Approving`, and `Liking` to this same scale, aligning it with other models, as Parrott, EARL, Plutchik, *and* Hourglass *all* use `Delight`. This can be verified by examining the "children emotions" defined in OCC – `Pride` becomes `Delight` in one's own action, which matches Hourglass.

Inversely, `Disapproving` and `Disliking` are mapped to the `Disgust` scale, used by all other models.

`Pleased` and `Displeased` still remain. These are more difficult to map, as they're used for the consequence of events (prospective or confirmed), resulting in "child emotions" of `Hope`, `Fear`, `Joy`, and `Distress`. Hourglass has these on different polar scales – `Joy` and `Fear` aren't opposites. But an interesting pattern emerges:

OCC uses `Joy` and `Distress` to relate to feelings about the consequence of events; while Hourglass has `Sadness` as the opposite of its `Joy`. If the model of Hourglass were applied to OCC's flow, then OCC's `Distress` becomes `Sadness`, for which OCC doesn't define `Sadness` explicitly.

This lets us translate OCC's `Distress`, which only appears explicitly in Parrott.

But this raises an issue for OCC's `Hope` and `Fear`. Both emotions can be found across all models. However, in Hourglass, `Hope` only appears as a compound of `Expectation` of `Joy`. This aligns with OCC, which has `Hope` as being `Pleased` about the prospective outcome of an event. But, again, the inverse of `Joy` in Hourglass is `Sadness`. Should this not make OCC's `Fear` a compound of `Expectation` and `Sadness`? Unfortunately, Hourglass defines `Fear` as its own polar scale opposite of `Eagerness`. This contradiction seemingly implies a weakness might exist in one or both models.

However, because of its usage of the term "Fear", the intent behind the `Eagerness` / `Fear` axis in Hourglass isn't as obvious as its other axes. It helps to instead think of it as "Certainty" / "Uncertainty". When you do something with `Enthusiasm`, you are certain of a given outcome. When you are `Anxious` about the future, you are uncertain it will be a positive one.

So although "Fear" appears in both OCC and Hourglass, these are actually different instances:

- Hourglass `Fear` is "uncertainty" in a moderate amount
- OCC `Fear` is the `Expectation` of `Sadness` (or other negative outcome) - what Hourglass considers `Despair`

By resolving OCC `Distress` as Hourglass `Sadness`, and OCC `Fear` as Hourglass `Expectation` + `Sadness`, the positive and negative "consequence feelings" are aligned. `Pleased` is `Joy`, either due to actual consequence or the prediction of one, while `Displeased` is `Sadness` – realized or perceived future.

Delving further down the feelings associated with actual consequences, we examine the "child emotions" of `Joy` and `Sadness`. Due to their conditional nature, we can represent them with the following tables:

Consequence For		Prospective Positive	Prospective Negative
Self	Actual Desirable	Satisfaction (`Joy`)	Relief (`Joy`)
	Actual Undesirable	Disappointment (`Sadness`)	Fears-Confirmed (`Sadness`)
Other	Liked Other	Happy-For (`Joy`)	Pity (`Sadness`)
	Disliked Other	Resentment (`Sadness`)	Gloating (`Joy`)

Due to how these child emotions are structured, we don't really need to worry about any of these emotions being true, unique emotions – they're simply their resulting "parents" stemming from a certain reason. Similar to how we resolved feelings of `Neglect` defined in Parrott as "`Sadness` with cause", each of these are "either `Joy` or `Sadness` with cause". We can attach a reason to the feeling, but internally they still *feel* the same.

Ideally I would have liked to translate all of these "emotions for cause" to Hourglass, or at the very least come up with single-word replacements for `Fears-Confirmed` and `Happy-For`. This might allow them to be used during actual program development to quickly differentiate positive/negative expectations and consequences. Unfortunately, after many hours over the course of a week, I have been unable to find suitable replacements. Even when a *concept* exists – like "Mudita" for `Happy-For` – they're not common English words, and thus don't instantly convey meaning, making them bad choices for terms used in development.

Some translations are possible. `Resentment` is a synonym of `Envy`, which appears in more models than `Resentment`, translating to Plutchik as `Anger` + `Sadness`. We have previously translated `Pity` to `Sympathy` during our evaluation of Parrott. `Gloating` already appears in Hourglass as a combination of `Joy` (what one feels) + `Disgust` (toward the disliked party).

`Disappointment`, while only explicitly appearing in Parrott and EARL, is fine as is – the word's very definition ("defeated in expectation or hope ") perfectly describes how the state is reached in the OCC logical flow. The English antonym for `Disappointment` happens to be `Satisfaction` ("fulfillment of one's wishes, expectations, or needs"), which also perfectly matches the logical flow.

Moving towards examination of the evaluation of actions by agents, `Pride` and `Shame` are already universal across all models. Similar to how we mapped `Approving` and `Disapproving` to `Pleasantness` and `Disgust`, we can also directly translate `Admiration` and `Reproach` to the same scale, only in relation to actions instead of agents. Having the same feeling lexicon towards the *actions* of other agents now results in direct translations to our evaluation *of* the other agent. If we feel some amount of `Disgust` by something the agent *does*, that `Disgust` transfers to our *impression of* the agent.

OCC's `Gratification`, `Remorse`, `Gratitude`, and `Anger` aren't unique emotions, but are all combinations of two different emotions. This is especially important to note, as the `Anger` in OCC is not the same as the `Anger` in Hourglass. OCC's `Anger` is `Disgust` (of an action) + `Sadness` (of its consequence). This would *always* include an element of `Sadness`, which is clearly not the case in real feelings of `Anger`.

Since these compounds are only how we might describe feelings towards actions with consequences, and the inclusion of the word "Anger" – a compound difficult to translate to a suitable word – might lead to confusion, this layer of "child emotions" should be removed. At the point in time in which we are describing these "emotions", an agent as already evaluated and assigned feelings of `Joy`/`Sadness` or `Pleasantness`/`Disgust` as appropriate; a word describing the final combination adds no new information.

This leaves the aspect evaluation branch of the tree. As "children" of `Liking`/`Disliking`, OCC's `Love`, `Hate`, `Interest`, and `Disgust` all directly translate to Hourglass `Pleasantness`/`Disgust`. How "familiar" you are with an aspect doesn't change how you feel about it – only its intensity. As a word, "Interest" could potentially be mapped to the `Eagerness` scale of Hourglass. But with the inverse – Hourglass `Fear` – then taking the place of OCC `Disgust`, the intent behind OCC `Interest`/`Disgust` is lost, and comes with incorrect associations for Hourglass emotions. You can feel both `Eagerness` and `Disgust` at the same time, so the implication that Hourglass `Eagerness` is the same as OCC's `Liking` (of an unfamiliar aspect) is contradictory.

We now have our final results for how we should handle each emotion described in the OCC model.

Retained (As-Is)	Translated (By Cross-Examination)	Removed (Superfluous or Confusing)	Unused (Not Core Emotions)
Hope Joy Pride Shame	Pleased => Joy Displeased => Sadness Distress => Sadness Fear => Despair Approving => Delight Disapproving => Disgust Admiration => Delight Reproach => Disgust Liking => Delight Disliking => Delight Love => Delight Hate => Disgust Interest => Delight	Gratification Remorse Gratitude Anger	Satisfaction Fears-Confirmed Relief Disappointment Happy-For Resentment => Envy Gloating Pity => Sympathy

Plutchik

Plutchik and Revisited Hourglass have one major difference – Plutchik has an axis for `Anticipation`/`Surprise`. This dimension was deliberately removed from Revisited Hourglass, as the dimension of `Surprise` introduced too much "neutrality" when the model was used to analyze sentiment. The removal of `Surprise` from Revisited Hourglass improved the model's perceived accuracy. Emotions like `Hope` and `Despair` still exist in the Revisited Hourglass model, but they only have "scores" for the `Joy`/`Sadness` dimension – there is no "`Surprise`" dimension.

Is Surprise an emotion? This is a contested subject. Revisited Hourglass says only this:

> "One of the main problems with the previous model was the presence of ambiguous emotions (e.g., distraction) and, especially, neutral emotions, e.g., surprise. Here, we do not want to debate whether surprise is an emotion or not but we definitely do not want it in a model that is catered for sentiment analysis as this will lead to the wrong categorization of all concepts (words and multiword expressions) that are semantically associated with it. Surprise, in fact, only becomes polar when coupled with positive or negative emotions."

In other models we're examining – Plutchik, Parrott, and EARL – all list `Surprise` as its own distinct emotion. Humans also make distinct facial expressions when they are surprised by something.

However, OCC does not include `Surprise` as its own emotion, but a *state*. When an agent had an expectation which was disconfirmed, they were "surprised" and felt an emotion depending on how their expectation matched the confirmed consequences. For example, if there was an expectation of `Sadness` which was disconfirmed with a result of `Joy`, the agent feels `Relief`. This aligns with the reasoning in Hourglass – the feelings associated with "surprise" are paired with another emotion.

I am of the belief that `Surprise` is not a unique emotion. `Surprise` is a reaction, but you don't *feel* surprised. You can feel `Relief` (`Joy`), `Disappointment` (`Sadness`), `Alarm` (`Fear`), or `Outrage` (`Anger`). Absent the "compound emotion", you don't *stay* in a feeling of `Surprise`. You are in a *state* of surprise as your rapidly process the situation and confirm its consequences. It is the perceived *consequences* associated with the surprise that put you in a given mood.

Further, while not including an explicitly named `Anticipation`/`Surprise` dimension, the Hourglass model actually can accommodate feelings associated with `Surprise` with its `Eagerness`/`Fear` dimension. As mentioned previously, this axis can be thought of as a "Certain"/"Uncertain" axis. As illustrations, consider the following messages appearing in your email:

- "You're fired."
  - This might come as a shock. But depending on your situation, you might feel...
    - Outrage. You're `angry`. How dare they fire you? The manager was horrid!
    - Dismay. You're `anxious`. Where will you find work?
- "I love you."
  - It's from the person you've had a crush on!
    - Ecstasy. You're `joyful`. This what you've always wanted!
    - Excited. You're `eager` to write your response!
  - It's from the person you'd rather avoid...
    - Alarm. You're `anxious`. How are you going to handle this?
    - Loathing. You're `disgusted`. This isn't who you wanted confessing to you at all.

Even without considering `Surprise` a true emotion (felt over a period of time), we can easily describe situations in which a person is "surprised" in context of other emotions, especially with the flexibility of Hourglass `Eagerness`/`Fear`.

On the same token, the `Anticipation` axis of Plutchik is also handled in compound with other emotional axes in Hourglass. Looking back to our examination of OCC, `Hope` is the expectation of `Joy`. Without the `Joy` component, there isn't an emotion to be felt. While we may say someone is "waiting expectantly", they're either `Eager` or `Anxious`, as well as possibly anticipating `Joy` or `Sadness`. Further, as you can tell from the previous sentence, "Anticipation" is a poor descriptor – *what* is a person anticipating? When we send a confession of love, are we waiting in anticipation for a reply in kind, or in anticipation of a rejection?

Rather than define `Anticipation` as a unique emotion, it would be best for us to instead use the emotions associated with the consequences the person anticipates. We wouldn't want a developer to assign an emotional state of "Anticipation" assuming it is always positive (love confession in kind) in situations in which it can vary (rejection)!

Therefore, we will remove the axis from Plutchik, as well as any compounds which rely on it. The majority of Plutchik's `Anticipation`/`Surprise` compounds are *already* accountable as Hourglass compounds. Keep in mind that these emotions are as conceived in Plutchik. Plutchik `Delight` is `Surprise` + `Joy`, whereas Hourglass `Delight` is positive Attitude toward an agent or their actions.

With this major difference addressed, here is our initial table comparing Plutchik with our other models.

Plutchik Only	Also In (Total Model Count)		Translated (Synonym / Examination)	Removed Anticipation / Surprise
	Hourglass Core (Not Compound)	Parrot, EARL, OCC, Hourglass Compound		
Boredom Dominance Sentimentality	Acceptance (1) Anger (3) Annoyance (3) Anxiety (3) Disgust (4) Ecstasy (2) Fear (3) Grief (2) Joy (4) Loathing (3) Pride (4) Rage (2) Sadness (4) Serenity (2) Shame (4) Terror (2)	Contempt (2) Despair (4) Envy (4) Guilt (3) Love (4) Morbidness (1) Remorse (2) Submission (1) Trust (1)	Admiration => Delight Apprehension => Anxiety Pensiveness => Melancholy	-- *Core* -- Interest Anticipation Vigilance Distraction Surprise Amazement -- *Compound (With)* -- Optimism (Joy) Pessimism (Sadness) Aggressiveness (Anger) Anxiety (Fear) Cynicism (Disgust) Hope (Trust) Delight (Joy) Disapproval (Sadness) Outrage (Anger) Awe (Fear) Curiosity (Trust) Unbelief (Disgust)

As Plutchik is now very similar to Hourglass, the rest of it is easy to examine.

Plutchik `Trust` is used as Plutchik's version of Hourglass `Pleasantness`. Its polar opposite is `Disgust`, as it is in Hourglass, and its combinations are aligned with Hourglass compounds. This also means `Love`, `Submission`, `Dominance`, and `Sentimentality` are compounds which are mappable to Hourglass.

However, while "Trust" makes up part of our evaluation of people, we've previously determined during our examination of EARL that we might need the term to refer to confidence in ourselves or other agents.

In this context, *trusting* another is not the same as being *pleased* in another. It factors into how we evaluate them or their actions. We can `Trust` someone we're not particularly fond of (their consistency if nothing else), while feeling `Doubt` towards someone we are close to (or ourselves).

Therefore we'll directly translate Plutchik `Trust` to Hourglass `Pleasantness`, while reserving the concept of "Trust" for later inclusion during our final model design. We still have `Lust` from Parrott, which might also fit into a sub-model of emotions for how we evaluate others.

The following Plutchik compounds are also mappable to Hourglass:

Plutchik Compound		Hourglass Emotion 1		Hourglass Emotion 2		Hourglass Compound
Contempt	'=='	Anger	+	Disgust	'='	Hatred
Despair	'=='	Sadness	+	Fear	'='	Distress
Envy	'=='	Sadness	+	Anger	'='	Envy
Love	'=='	Joy	+	Pleasantness	'='	Love
Morbidness	'=='	Disgust	+	Joy	'='	Morbidness
Remorse	'=='	Sadness	+	Disgust	'='	Remorse
Submission	'=='	Pleasantness	+	Fear	'='	Submission

The only remaining Plutchik combination is `Guilt`, which Plutchik models as `Joy` + `Fear`. It's unclear if this is meant to represent feeling `Joy` *despite* `Fear`, or `Fear` *of feeling* `Joy`. Further, Hourglass models the same emotion as `Sadness` + `Disgust`. These different interpretations of the meaning of `Guilt` suggest we should remove it from consideration for being potentially confusing. It's a mere combination anyway – the meaning we intend for a given situation can come from the base emotions.

Plutchik's final emotion to be considered is `Boredom`. At first glance this is radically different from anything we've examined from previous models. We have yet to consider what it's like to be *bored*. However, Plutchik uses it as its first stage of its `Disgust` scale. Mapped directly to Hourglass, this can work in the context of one's current situation or something they're involved with.

- A movie isn't grabbing our attention, so we `Dislike` what has been happening in it – we're `Bored` of it.
- We're sitting in a room with nothing to entertain us. We `Dislike` our current experience – we're `Bored`.

When mapped to Hourglass, `Boredom` should also take into account Hourglass's `Eagerness` axis. You're not going to be acting with `Enthusiasm` if you're `Bored`. The antonym of `Boredom` – `Excitement` – can be described in Hourglass as a compound of `Joy` and `Eagerness`. But clearly `Boredom` isn't a direct inverse – a compound of `Sadness` and `Fear`. `Sadness` can be *part* of the `Boredom` felt by a person, but isn't always included. However, `Boredom` is certainly marked by lack of `Eagerness`, with a probable degree of `Dislike`.

In either event, no matter which Hourglass compounds we use to describe `Boredom`, it's not a core emotion we need to separately account for *on its own*. It is, however, used as part of our evaluation of the `Disgust`/`Delight` we feel towards an agent, action, or situation. So the concept of "Boredom" may merit later consideration, but only in that context.

Parrott & EARL (Again)

We've now had our first evaluation of all non-Hourglass models. We have narrowed Plutchik and OCC into terms of Hourglass. We'll take a brief second pass over Parrott and EARL to account for emotions we previously left alone because they had existed in multiple models. Emotions which exist exactly in Hourglass will not be mentioned, and emotions in EARL won't be mentioned a second time if they exist in Parrott.

Parrott

- `Apprehension`: The same as `Anxiety`. Removed.
  - We might later use the term as a "lower" form of the Hourglass `Fear` scale.
- `Cheerfulness`: Combination of `Joy`, `Eagerness`, or `Pleasantness`. Removed.
- `Contempt`: Combination of `Anger` and `Disgust`. Removed.
- `Despair`: Analyzed during evaluation of OCC as the expectation of `Sadness`.
  - While not a core emotion, we can still use this term in our final emotional design.
- `Disappointment`: Analyzed during evaluation of OCC. `Sadness` contrary to expectation. Removed.
  - While not a core emotion, we can still use this term in our final emotional design.
- `Displeasure`: Same emotional axis as `Disgust`, combinable with `Anger` or `Sadness`. Removed.
- `Distress`: Compound of `Sadness` and `Fear`. Removed.
- `Envy`: Compound of `Sadness` and Anger. Removed.
- `Excitement`: Compound of `Joy` and `Eagerness`. Removed.
- `Guilt`: Analyzed during evaluation of OCC. Different situations possibly lead to confusion. Removed.
- `Hope`: Analyzed during evaluation of OCC as the expectation of `Joy`.
  - While not a core emotion, we can still use this term in our final emotional design.
- `Love`: Combination of `Joy` and `Pleasantness`. Removed.
- `Lust`: An emotion that humans feel which is not adequately conveyed with Hourglass's axes.
  - We intend to somehow include this in our final emotional model.
- `Optimism`: Also an expectation of future `Joy`. Removed.
- `Outrage`: Analyzed during evaluation of Plutchik. A combination of `Surprise` and `Anger`. Removed.
- `Relief`: Analyzed during evaluation of OCC as the disconfirmation of `Sadness`.
  - While not a core emotion, we might still use this term in our final emotional design.
- `Remorse`: Combination of `Sadness` and `Disgust` in relation to a past event.
  - We intend to keep `Pride`, but will use `Shame` as its opposite instead of `Remorse`.
- `Satisfaction`: Confirmation of previously expected Joy. Not a core emotion. Removed.
- `Surprise`: Analyzed during evaluation of Plutchik. Not an emotion, but a temporary state. Removed.
- `Sympathy`: `Sadness` regarding a perceived negative event for another agent. Not a core emotion.
  - It might still be useful to use the term in our final emotional design.

EARL

- `Boredom`: Evaluated in Plutchik. Mix of `Dislike` and/or `Sadness`, absent any `Eagerness`. Removed.
- `Doubt`: The inverse of the concept of `Trust`, factoring into our evaluation of others
  - We intend to somehow include this in our final emotional model.
- `Interest`: Evaluated in Plutchik. A combination with `Surprise`, which isn't a core emotion. Removed.
- `Trust`: Feeling of confidence placed in ourselves/others, not precisely represented by another emotion.
  - We intend to somehow include this in our final emotional model.

Final Emotional Classification

Our final emotion model consists of two things we need to address:

1. Core emotions – what do we feel in general?
2. Evaluation of agents – how do we specifically feel about others?

Core Emotions

While we started with a strong preference for OCC as an emotional model due to its programmatic flow, the emotions described by it aren't very adaptable for describing complex emotional states. We desire an agent with the capability for a wide range of emotions, with us being able to describe complex emotions without having to upend our model. Therefore, we'll model our list of core emotions after Revisited Hourglass, as it has proven to be the most robust of all models we've examined.

Unfortunately, some of the example emotions it uses to illustrate its axes aren't as clear as they could be, forgo simple polar words such as using `Dislike`/`Acceptance` instead of `Dislike`/`Like`, and the three-part axis delineations result in breakpoints of 0.33 and 0.66 before landing into the next example "tier".

So I propose the following axes to add a fourth tier, with some substitutions to make the axes and intensities more obvious.

`Zeal`	`Enthusiasm`	`Eagerness`	`Responsiveness`	`Apprehension`	`Anxiety`	`Fear`	`Terror`
`Serenity`	`Peaceful`	`Calmness`	`Acceptance`	`Annoyance`	`Frustration`	`Anger`	`Rage`
`Ecstasy`	`Joy`	`Happiness`	`Contentment`	`Melancholy`	`Sadness`	`Sorrow`	`Grief`
`Adoration`	`Admiration`	`Like`	`Approval`	`Disapproval`	`Dislike`	`Disdain`	`Loathing`

- "Bliss" has been removed from the highest slot of the Temper axis, as the word is strongly associated with `Joy`.
- "Serenity" has been moved from the lowest slot of the Temper axis to the highest slot, replacing "Bliss". I'm unaware why the word "Serenity" was originally considered a lower intensity than `Calmness`.
- The word "Acceptance" has been repurposed from the Attitude axis to instead be the starting tier of the Temper axis, in place of "Serenity".
- "Pleasantness" has been replaced by `Admiration`, as "Pleasantness" could easily be confused for an emotion on the Temper or Introspection axis.
- "Delight" has been replaced by `Adoration`, as "Delight" was too easily confused for being part of the Introspection axis.
- "Disgust" has been replaced by `Disdain`, as "disgusting" can be associated with aesthetic elements about things which aren't always necessarily disliked.
- `Zeal`, `Apprehension`, `Peaceful`, `Frustration`, `Happiness`, `Sorrow`, `Like`, `Approval`, and `Disapproval` have been added where appropriate to add a "fourth tier" to each axis.
- Words have been chosen and placed such as the "second tier" – the word describing the emotion up to the 50% of the axis – can easily convey the axis.
  - `Eagerness`/`Anxiety`
  - `Calmness`/`Frustration`
  - `Happiness`/`Sadness`
  - `Like`/`Dislike`

Evaluation of Agents

One issue that Revisited Hourglass noted with the original Hourglass model was the Attitude dimension excluding self-conscious or moral emotions. They resolved this by adding a "sub-dimension" to the Attitude category to account for attitudes towards the self, and attitude toward others. Several emotions which were explicitly listed in other models were forced into this dimension, like `Pride`, `Shame`, or `Gratitude`. We have also decided that other emotions exist which we would prefer to track separately from other emotions, like `Lust` and `Trust`.

All these emotions are related in the sense that they are in the context of evaluation: ourselves, other agents, or other things. While they're all components of a total "attitude" towards something, throwing them all together into one scalar axis – even if associated with individual agents – would hinder us in the goal of our virtual agent being able to make nuanced decisions.

If our virtual agent only evaluated others based on a single "Attitude" score:

- At what point does an agent decide to pursue a sexual relationship with another agent?
  - Problem: People tend to have high overall opinions of family members, or children in their care.
- At what point does an agent trust another with a potentially devastating secret?
  - Problem: People might have initial curiosity in someone they only just met.
- At what point does an agent potentially sacrifice themselves over their feelings for another?
  - Problem: People run into fires for their loved pets, who aren't even an option as a romantic partner.

While Revisited Hourglass is very strong in sentiment analysis, and serves as a strong base for most emotions, our goal is user interaction with a virtual agent. We therefore need to account for different lenses the virtual agent may view a user through, and how a user's actions or aspects might influence the agent's feelings.

- A virtual agent "of age" shouldn't engage in sexual flirting with a user who is not.
- A virtual agent should be less trusting of a user who is known to cheat in games.

As of the time of writing, I have been unable to find existing emotional psychological literature which specifically examines and categorizes emotions that make up the total combined Attitude an agent has for something. Most existing work in affective computing is for the purpose of sentiment analysis, and there is little focus on accurate emotional simulations for governing interpersonal relations between users and virtual agents.

Instead, I have examined emotions which we previously eliminated from other models, some of the "example words" provided as illustrations in the refined Attitude dimension for Revisited Hourglass, and generally just put a lot of thought into human interpersonal interactions.

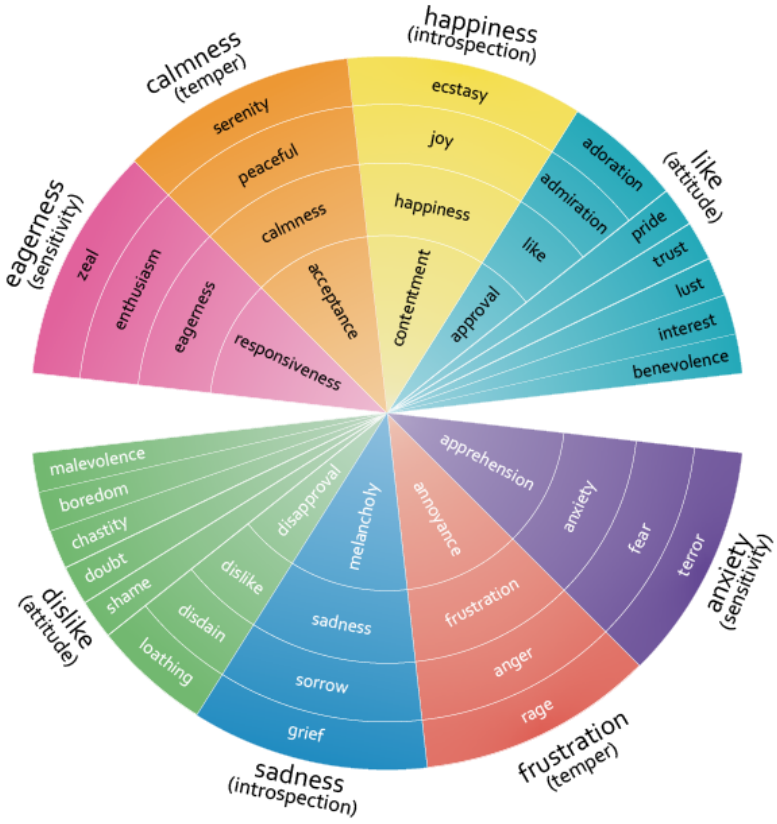
With the goal of a virtual agent that develops interpersonal relationships with human users, I propose the following dimensions as "sub-dimensions" of the Attitude category. While descriptions might focus on how these sub-dimensions are expressed in relation to agents, they can also be felt towards objects, situations, or even abstract concepts.

Emotions which have previously been explicitly listed or used as examples in other models appear in bolded italics.

Negative		Positive	
Descriptor	**Related Words**	**Descriptor**	**Related Words**
Chastity	Revulsion, Disgust	***Lust***	***Attraction***, Infatuation
Description	Governs the sexual attraction the agent has towards another, or evaluates about themselves. Aspects of another agent which increase this value might include physical beauty or appealing clothing. Aspects which might decrease this might include poor hygiene, physically unhealthy habits, or inappropriate age.		
Doubt	***Unbelief***, ***Insecurity***, Suspicion, ***Prejudice***	***Trust***	***Confidence***, ***Security***, Faith
Description	Governs perceived likelihood another agent will undertake actions leading to our own positive feelings. Increases to this value include an agent having a history of causing the evaluating agent to feel happiness, or believing an agent is capable of handling tasks with competence. Decreases to this value include being found to have lied to the evaluating agent, or repeatedly failing to live up to expectations in a relationship. Agents with high trust in themselves can often push through fearful situations. Agents with high trust in another will tend to prefer to rely on, or make themselves vulnerable to that agent.		
Boredom	Apathy, Dullness	***Interest***	***Curiosity***, Inquisitiveness
Description	Governs how intellectually stimulating an agent finds a situation or other agent. An agent meeting another for the first time might have high interest in learning everything about them, but develop low interest if they come to the conclusion the other agent is boring. Increases to this value might include discovering that another person or thing has a rare novel aspect, or the other person often presents an intellectual perspective the evaluating agent finds interesting or useful. Decreases might include learning the other agent lacks any sort of interesting hobby, or that the agent or thing is disappointingly predictable.		
Malevolence	***Unfriendliness***, ***Antipathy***, Cruelty	***Benevolence***	***Friendliness***, Compassion, ***Sympathy***, Protectionism, ***Caring***
Description	Governs the investment an agent puts into protecting or causing harm to another agent. Aspects positively influencing this include another agent being a ward of the evaluating agent, or perceived as being unable to fend for themselves. Aspects that might decrease this is an evaluating agent believing another agent deserves to be harmed, or believing another agent might actively harm a beloved third agent. Agents with high benevolence for another will go to great lengths for the other's well being or happiness, sometimes to the point of self-sacrifice. Agents with high benevolence for themselves will engage in healthy habits or other "self care" activities. Agents with high malevolence for another will go out of their way to harm them, even at cost to themselves. Agents with high malevolence for themselves might be reckless about their own safety, or actively engage in self harm.		
Shame	***Remorse***, ***Regret***, ***Embarrassment***, Disrepute	***Pride***	Appreciation, Esteem, Respect
Description	Governs socially positive or negative judgments of an agent or action. Events increasing this might be personal accomplishments or undertaking actions commonly encouraged by others. Events decreasing this might include harming another beloved agent, or wastefully discarding opportunities which might lead to happiness. While heavily influenced by societal expectations or "shared morality", this dimension can be highly variable based on an individual agent's personal beliefs, demeanor, or goals. Most agents might feel remorse for greatly harming another, while a different agent might find it to be a source of pride should the harmed agent "deserve" it. Many agents might find a hobby of an individual to be shameful, which the individual might ignore as a point of pride, with another agent holding the individual in higher esteem for being brave enough to engage in that hobby despite societal shame.		

Final Chart

When our newly labeled axes for our chosen core emotional model are combined with our additions of specific `Like`/`Dislike` "child" emotions, we end up with the following illustrative chart of our final model.



It is worth noting that while on this chart the "child" emotions of the `Like`/`Dislike` axis are separated from the example words of their "parent" category, the child emotions should still evaluate as part of their parent. `Boredom` felt about an agent should contribute towards our total `Dislike` of them. These "child" axes are tracked separately only for ease of determining when our evaluating agent should – or should not – engage in more nuanced behaviors towards a second agent. Interpersonal emotions towards another agent which *aren't* accounted for by these "child" axes still "fall back" to the generalized `Like`/`Dislike` axis.

References

1. Shaver, Phillip, et al. "Emotion knowledge: further exploration of a prototype approach." *Journal of personality and social psychology* 52.6 (1987): 1061.
2. Parrott, W. Gerrod, ed. *Emotions in social psychology: Essential readings*. psychology press, 2001.
3. Schröder, Marc, Hannes Pirker, and Myriam Lamolle. "First suggestions for an emotion annotation and representation language." *Proceedings of LREC*. Vol. 6. 2006.
4. HUMAINE Association. “HUMAINE Emotion Annotation and Representation Language (EARL): Schema.” Web Archive, June 2006, web.archive.org/web/20080330213740/emotion-research.net/projects/humaine/earl/schemadesign. Accessed August 2021.
5. Plutchik, Robert. "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice." *American scientist* 89.4 (2001): 344-350.
6. Li, Ran, et al. "EmoMix: Building an emotion lexicon for compound emotion analysis." *International Conference on Computational Science*. Springer, Cham, 2019.
7. Ortony, Andrew, Gerald L. Clore, and Allan Collins. *The cognitive structure of emotions*. Cambridge university press, 1990.
8. Steunebrink, Bas R., Mehdi Dastani, and John-Jules Ch Meyer. "The OCC model revisited." *Proc. of the 4th Workshop on Emotion and Computing*. Association for the Advancement of Artificial Intelligence, 2009.
9. Cambria, Erik, Andrew Livingstone, and Amir Hussain. "The hourglass of emotions." *Cognitive behavioural systems*. Springer, Berlin, Heidelberg, 2012. 144-157.
10. Susanto, Yosephine, et al. "The hourglass model revisited." *IEEE Intelligent Systems* 35.5 (2020): 96-102.