Title: Modeling Emotions in a Computational System

Sub-title: Emotional Modeling in the Independent Core Observer Model Cognitive Architecture

By: Artificial General Intelligence Inc. and ICOM Team

ABSTRACT

This paper is an overview of the emotional modeling used in the Independent Core Observer Model (ICOM) Cognitive Extension Architecture research which is a methodology or software ‘pattern’ for producing a self-motivating computational system that can be self-aware under certain conditions. While ICOM is also as a system for abstracting standard cognitive architecture from the part of the system that can be self-aware it is primarily a system for assigning value on any given idea or ‘thought’ and based on that take action as well as producing on going self-motivations and in the system take further thought or action. ICOM is at a fundamental level driven by the idea that the system is assigning emotional values to ‘context’ (or context trees) as it is perceived by the system to determine how it feels. In developing the engineering around ICOM two models have been used based on a logical understanding of emotions as modeled by traditional psychologist as opposed to empirical psychologist which tend to model emotions (or brain states) based on biological structures. This approach is based on a logical approach that is also not tied to the substrate of any particular system.

KEY WORDS: Artificial Intelligence, Artificial General Intelligence, AGI, Self-Motivating, Cognitive Architecture, AI

INTRODUCTION

Emotional modeling used in the Independent Core Observer Model (ICOM) represents emotional states in such a way as to provide the basis for assigning abstract value to ideas, concepts and things as they might be articulated in the form of context tree’s where such a tree represents the understanding of a person, place or thing including abstract ideas and other feelings. These trees are created by the context engine based on relationships with other elements in memory and then passed into the core (see the whitepaper titled “Overview of ICOM or the Independent Core Observer Model Cognitive Extension Architecture”) which is a methodology or ‘pattern’ for producing a self-motivating computational system that can be self-aware under certain conditions. This particular paper is focused only on the nuances of emotional modeling in the ICOM program and not what is done with that modeling or how that modeling may or may not lead to a functioning ICOM system architecture.

While ICOM is also as a system for abstracting standard cognitive architecture from the part of the system that can be self-aware, it is primarily a system for assigning value on any given idea or ‘thought’ and based on that the system can take action, as well as produce on going self-motivations in the system to further then have additional thought or action on the mater. ICOM is at a fundamental
level driven by the idea that the system is assigning emotional values to ‘context’ as it is perceived by the system to determine its own feelings. In developing the engineering around ICOM, two models have been used for emotional modeling, which in both cases are based on a logical understanding of emotions as modeled by traditional psychologist as opposed to empirical psychologist which tends to be based on biological structures. The approaches articulated here are based on a logical approach that is also not tied to the substrate of the system in question (biological or otherwise).

**Understanding the problem of Emotional Modeling**

Emotional structural representation is not a problem I wanted to solve independently nor one I felt I had enough information as an expert to solve without a lifetime of work on my own. The current representational systems that I’ve selected are not necessarily the best way(s) or the right way(s) but two ways that do work and are used by a certain segment of psychological professionals. This is based on the work of others in terms of representing the complexity of emotions in the human mind by scientists that have focused on this area of science. The selection of these methods are more based on computational requirements than any other selection criteria.

It is important to note that both of the methods the ICOM research have used are not based on scientific data as might be articulated by empirical psychologists which might use ANOVA (variance analysis), or factor analysis. [7] While these other models maybe be more measured in how they model elements of the biological implementation of emotions in the human mind, the model’s selected by me here for ICOM research are focused on ‘how’ and the logical modeling of those emotions or ‘feelings’. If we look at say the process for modeling a system such as articulated in “Properties of Sparse Distributed representations and their Application to Hierarchical Temporal Memory” [8] such representation is very much specific to the substrate of the human brain. Since I am looking at the problem of self-motivating systems or computational models that are not based on the human brain literally but only in the logical sense the Wilcox system [9] or more simply the Plutchik method [10] is a more straight forward model and accurately models logically what we want to-do to separate from the underlying complexity of the substrate of the human biological mind.

**The Plutchik Method**

George Norwood described the Plutchik method as:

“Consider Robert Plutchik’s psychoevolutionary theory of emotion. His theory is one of the most influential classification approaches for general emotional responses. He chose eight primary emotions - anger, fear, sadness, disgust, surprise, anticipation, trust, and joy. Plutchik proposed that these ‘basic’ emotions are biologically primitive and have evolved in order to increase the reproductive fitness of the animal. Plutchik argues for the primacy of these emotions by showing each to be the trigger of [behavior] with high survival value, such as the way fear inspires the fight-or-flight response.

Plutchik’s theory of basic emotions applies to animals as well as to humans and has an evolutionary history that helped organisms deal with key survival issues. Beyond the basic emotions there are combinations of emotions. Primary emotions can be conceptualized in terms of pairs of polar opposites. Each emotion can exist in varying degrees of intensity or levels of arousal.” – [10]
While George Norwood mentions earlier in his paper talking about the Plutchick method that it would be almost impossible to represent emotions in terms of math or algorithms I would disagree. As you can see by this representation of the Plutchik method it is essentially 8 vectors or ‘values’ when represented in 2 dimensions which is easily modeled with a series of number values.

Figure 2A – Plutchik Model

Now in the case of ICOM since we want to represent each segment as a numeric value, a floating point value was selected to insure precision along with a reverse scale as opposed to what is seen in the diagram above. Meaning if we have a number that represents ‘joy/serenity/ecstasy’ the ICOM version is a number starting from 0 to N where N is increasing amounts or intensity of ‘joy’.
To represent ICOM emotional states for anything assigned emotional values you end up with an array of floating point values. By looking at the chart above we can see how emotional nuances can be represented as a combination of values on two or more vectors which gives us something closer to the Wilcox model but using less values and given the difference it is orders of magnitude when seen in terms of a computational comparison.

Let us take a look at this next diagram.

![Figure 2B – Modified Plutchik](image)

As you can see we have reversed the vectors such that the value or ‘intensity is increasing as we leave the center of the diagram on any particular vector. From a modeling standpoint this allows the intensity to be infinite above zero verses limiting the scale in the standard variation not to mention it is more aligned with what you might expect based on the earlier work (see the section on Willcox next). This variation as seen here is what we are using in the ICOM research.

Let us look at the other model.
The Wilcox Model

Initially the ICOM Research centered on using the Willcox model for emotions and is still big part of the modeling methodology and research going into the ICOM project. Given the assumption that the researchers in the field of mental health or studying emotions have represented things to a sufficient complexity to be reasonably accurate we can therefore start with their work as a basis for representation I therefore landed on Willcox initially as being the most sophisticated ‘logical’ model. Take a look at the following diagram:

![Figure 2C – Dr. Gloria Willcox’s Feelings Wheel [9]](image)

Based on the Willcox wheel we have 72 possible values (the six inner emotions on the wheel are a composite of the others) to represent the current state of emotional affairs by a given system. Given that we can then represent the current emotional state at a conscious level by a series of values that for computation purposes we will consider ‘vectors’ in an array represented by floating point values. Given that we can also represent subconscious and base states in a similar way that basically gives us 144 values for the current state. Further we can use them as vectors to represent the various states to back weight and adjust for new states. This then can be represented as needed in the software language of choice.
If we map each element to vectors spread on a 2 dimension X/Y plane we can compute an average composite score for each element and use this in various kind of emotional assessment calculations.

We are thus representing emotional states using two sets of an array of 72 predefined elements using floating point values we also can present assigned arrays on a per context basis and use a composite score of an element as processed to further compare various elements of context emotional arrays or composite scores with current states and make associated adjustments based on needs and preexisting states. For example the current emotional state of the system by be a set of values and a bit of context might affect that same set of values with its own set of values for the same emotions based on its associated elements and a composite is calculated based on the combination which could be an average or mean of each vector for any given element of the emotional values.

The Emotional Comparative Relationship

Given the array of floating point number declarations, a given element of ‘context’ will have a composite of all pre-associated values related to that context and any previous context as it might be composited. For this explanation we will assume context is pre-assigned. The base assignments of these values are straight forward assignments but each cycle of the core (see the ICOM Model overview for a detailed explanation of ICOM and the core) will need to compare each value and assign various rules on the various elements of the context to assign effects on itself as well as conscious and subconscious values.

Logically we might have the set values that are the current state as in the earlier example. We then get a new block of context and adjust all the various element based on those complex sets of rules that affect the conscious and subconscious states (emotional arrays of floating point values). Rules can be related to emotions which includes tendencies, needs, interests or other factors as might be defined in the rules matrix applied by the core.

This gives us a framework for adjusting emotional states given the assumption that emotional values are assigned to context elements based on various key factors of the current state and related core environment variables. The process as indicated in the context of evaluating becomes the basis for the emergent quality of the system under certain conditions where the process of assigning value and defining self-awareness and thought are only indirectly supported in the ICOM architecture, and emerge as the context processing becomes more complex.

Context Assignments

One of the key assumptions for computing the emotional states is the pre-assignment of emotional context prior to entering the emotional adjustment structures of the core system.

While this explanation does not address for example looking at a picture and decomposing that into understanding in context it does deal with how emotional values are applied to a given context element generated by the evaluation of that picture.

As described earlier there are 72 elements needed to represent a single emotional context (based on the Willcox model) given the selected methodology. Let’s say of that array the first 3 elements
are ‘happiness’, ‘sadness’, and ‘interest’. Additionally let us assign them each a range between 0 and 100 as floating point values meaning you can have a 1 or a 3.567 or a 78.628496720948 if you like.

If for example a particular new context A is related to context B and C which had been processed earlier and related to base context elements of D, E and F. This gives us a context tree of 6 elements. If we average the emotional values of all of them to produce the values of happiness, sadness and interest for context A we now have a context tree for that particular element which then is used to affect current state as noted above. If that element still has an interest level, based on one of those vectors being higher than some threshold then it is queued to process again and the context system will try to associate more data to that context for processing. If Context A had been something thought about before then that context would be brought up again and the other factors would be parsed in for a new average which could have then been an average of all 6 elements where before context A didn’t have an emotional context array were the second time around it does. Further on processing the Context A its values are changed by the processing against the current system state.

Using this methodology for emotional modeling and processing we also open the door for explaining certain anomalies as seen in the ICOM research.

Computer Mental Illness

In the ICOM system the ‘core’ emotional thought and motivation system if any of the 72 vectors get into a fringe area at the high or lower end of the scale can produce an increasingly irrational set of results in terms of assigning further context. If the sub-conscious vectors are to far off this will be more pronounced and less likely to be fixed over time, creating some kind of digital mental illness where given the current state of research it is hard to say the kinds of and manifestation of that illness or illnesses could be as varied as human mental illness. Now the subconscious system is in fact critical to stabilization of the emotional matrix of the main system in that it does change slightly over time where under the right extreme context input is where you get potential issues on a long term basis with that particular instance. The ICOM research and models have tried to deal with these potential issues by introducing limiting bias and other methods in preventing too radical of a result in any given operation.

Motivation of the Core model

Given the system in the previous sections for assigning emotional context, processing and assigning context elements that are above a certain threshold are targets for reprocessing by being placed in a que feeding the core. The motivation of the core comes from the fact that it can’t “not” think and it will take action based on emotional values assigned to elements that are continuously addressed where the core only needs to associate an ‘action’ or other context with a particular result and motivation is an emergent quality of the fact that things must be processed and actions must be taken by design of the system.

This underlying system then is thus designed to have a bias for taking action with action being abstracted form the core in detail where the core only need composite such action at a high level; In other words it just needs to ‘think’ about an action.

Core Context
Core Context are the key elements predefined in the system when it starts for the first time. These are ‘concept’s that are understood by default and have predefined emotional context trees associated with them. While the ICOM theory for AGI is not specific to the generation or rather the decomposition of ‘context’ it is important to address the ‘classification’ of context. In this way any context must add qualities that may be new and can be defined dynamically by the system but these core elements that are used to tag those new context elements. Since all context is then streamed into memory as processed and can be brought back and re-referenced as per the emotional classification system pending the associated threshold determining if it is something of relevance to recall.

As stated elsewhere lots of people and organizations are focused on classification systems or systems that decompose input, voice, images and the like however ICOM is focused on self-motivation along the lines of the theory as articulated based on emotional context modeling.

What is important in this section is the core elements used to classify elements of context as they are processed into the system. The following list of elements is used as a fundamental part of the ICOM system for its ability to associated emotional context to elements of context as they are passed into the core. This same system may alter those emotional associations over time as new context not hither to classified is tagged based on the current state of the system and the evaluation of elements or context for a given context tree when the focus of a given context tree is processed. Each one of these elements below has a 72 vector array of default emotions (using the Willcox based version of ICOM) associated with that element by default at system start. Additionally this may not be an exhaustive list of the default core system in the state of the art. These are only the list at the time this section is being written.

i. **Action** – A reference to the need to associate a predisposition for action as the system evolves over time.

ii. **Change** – a reference context flag used to drive interest in changes as a bias noticing change.

iii. **Fear** – Strongly related to the pain system context flag.

iv. **Input** – A key context flag needed for the system to evolve over time recognizing internal imaginations vs system context input that is external.

v. **Need** – A reference to context associated with the needs hierarchy

vi. **New** – A reference needed to identify a new context of some kind normally in terms of a new external object being cataloged in memory

vii. **Pain** – having the most negative overall core context elements used as a system flag of a problem that needs to be focused on. This flag may have any number of autonomic responses dealt with the ‘observer’ component of the system.

viii. **Pattern** – A recognition of a pattern build in to help guide context as noted in humans that there is an inherent nature to see patterns in things. While there could be any number of evolutionary reasons for it, in this case we will assume the human model is sound in terms of base artifacts regarding context such as this.

ix. **Paradox** – a condition where 2 values that should be the same are not or that contradict each other. Contradiction is a negative feedback reference context flag to condition or bias the system want to solve paradox’s or have a dislike of them.

x. **Pleasure** – having the most positive overall core context element used as a system flag for a positive result
xi. **Recognition** – a reference flag used to identify something that relates to something in memory.

xii. **Similar** – related to the pattern context object used to help the system have a bias for patterns by default

xiii. **Want** – A varying context flag that drives interest in certain elements that may contribute to needs or the ‘pleasure’ context flag.

While all of these might be hard coded into the research system at start they are only really defined in terms of other context being associated with them and in terms of emotional context associated with each element which is true of all elements of the system. Further these emotional structures or matrices that can change and evolve over time as other context is associated with them.

As a single example let’s take a look at the first core context element that is defined in the current Willcox based ICOM implementation called ‘pain’. This particular element doesn’t represent emotional pain as such but directly effects emotional pain as this element is core context for input assessments or ‘physical’ pain however note that one of the highlighted elements in the ‘pain’ matrix is for emotional pain.

<table>
<thead>
<tr>
<th>Default Emotional Context Assignments for “Pain”</th>
</tr>
</thead>
<tbody>
<tr>
<td>J.Excited: 0.0</td>
</tr>
<tr>
<td>J.Playful: 0.0</td>
</tr>
<tr>
<td>P.Proud: 0.0</td>
</tr>
<tr>
<td>P.Worthwhile: 0.0</td>
</tr>
<tr>
<td>Pe.Thankful: 0.0</td>
</tr>
<tr>
<td>Pe.Response: 0.0</td>
</tr>
<tr>
<td>S.Apathetic: 0.0</td>
</tr>
<tr>
<td>S.Miserable: 0.0</td>
</tr>
<tr>
<td>M.Hurt: 99.0</td>
</tr>
<tr>
<td>M.Furious: 25.0</td>
</tr>
<tr>
<td>Sc.Anxious: 0.0</td>
</tr>
<tr>
<td>Sc.Helpless: 0.0</td>
</tr>
</tbody>
</table>

**Figure 2D – Emotional Matrix array at system start for context element ‘pain’**

On top of all of the emotional states associated with a context element they themselves also are pre-represented in the initial state predefined into the system as context themselves. You can see that in this initial case we have guesses at values in the matrix array for default values for each element which has to be done for each predefined context element at system start. This allows us to set certain qualities as a basic element of how a value system will evolve in the system creating initial biases. For example we might create a predilection for a pattern which creates the appropriate bias in system as we might want to see in the final AGI implementation of ICOM.

**Personality, Interests and Desires of the ICOM system**
In general under the ICOM architecture regardless of which of the two modeling systems that have been used, in ICOM the system very quickly creates predilection for certain things based on its emotional relationship to those elements of context. For example, if the system is exposed to context X which it always had a good experience ‘including’ interest the methodology regardless of case, develops a preference for or higher emotional values associated with that context or other things associated with that context element. This evolutionary self-biasing based on experience is key to the development of personality, interests and desires of the ICOM system and in various experiments has shown that in principal it is very hard to replicate those biases of any given instance due to the extreme amount of variables involved. While ultimately calculable, a single deviation will change results dramatically over time. This also leads us to a brief discussion of free will.

**Free Will as an illusion of complex contextual emotional value systems**

Frequently the problem of “free will” has been an argument between determinist verses probabilistic approaches and given either case an argument as to the reality of our free will ensues.

While we don’t understand exactly the methodology of the human mind if it works in a similar manner at a high level like ICOM then, under that architecture, it would strongly imply that free will is an illusion. For me, this is a difficult thing to be sure of given that this is outside the scope of the research around ICOM; but none the less it is worth mentioning the possibility. Additionally, if true, then free will seems to be something that can be completely mathematically modeled. If that is the case, it is likely that of the human mind can be as well. Certainly, as we progress this will be a key point of interest but outside my expertise.

ICOM emotional modeling seems to be at the heart of this mathematical modeling of ‘free will’ or what appears to be free will in a sea of variables that is so vast that we collectively have not been able to full modeling but ICOM based systems appear in function to exhibit free will based on their own self biases based on their experience which because we can and have a model free will in ICOM is an illusion of the sea of factors required for ICOM systems to function. Let’s get back to the different methods used in ICOM for emotional modeling.

**Plutchik verses Willcox**

When determining which method to use in emotional modeling we see a number of key facts. Willcox models all the nuances of human emotions directly with numerous vectors or values. Plutchik models those nuances through combination of values thus using a total number of values that is much less. From a computational standpoint Plutchik has 8 sets of core values where Willcox using 72 so having two sets of those for conscious and unconscious values gives us 72 which converting that to a 2d plain requires conversion X/Y values which means 144 trigonomic functions for each pass through the ICOM core whereas using Plutchik we have 16 total values with the same conversion of X/Y values to produce the average emotional effects applied to incoming emotional context means only 16 trigonomic functions per core pass which means from a computational standpoint we only need 9 times less computational power to run with the Plutchik method which is the basis for the research post series 3 experiments moving forward at least for the foreseeable research that is in progress.

**Visualizing Emotional Modeling Using Plutchik**
To better understand how any given instance of ICOM is responding in tests we needed a method for visualizing and representing emotional state data and given the method for modeling in either method articulated earlier we came up with this method here for indicating state. This method visualizes graphically emotional state of what is going on in the core. You can see we are visualizing emotional states much like the earlier diagrams then in which we look at vectors that represent the model (see figure 2A and 2B).

So let’s look as an example. In this case we are looking at one of the program series 3 experiments in which case we were looking at ICOM introspection as it relates to the system thinking about previous elements to see if the system would pick something out of memory and then thinking about it and see how it affects various vectors or emotional states. The rest of the experiment is not as important to the point in which here we are showing how that data is represented. In this case we start with the following raw data keeping in mind here we are looking at only 4 of the eight values modeled in the test system:

<table>
<thead>
<tr>
<th>conscious</th>
<th>0</th>
<th>5</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>subconscious</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>c</td>
<td>3.05</td>
<td>1.95</td>
<td>0.5</td>
<td>5.05</td>
</tr>
<tr>
<td></td>
<td>0.0305</td>
<td>5.0305</td>
<td>0.005</td>
<td>0.0505</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>9</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

| c         | 6.420975 | 0.5 | 2.224975 |
|           | 0.030805 | 5.04440475 | 0.00995 | 0.07224475 |
|           | 0 | 5 | 9 | 0 |

| c         | 6.329316013 | 4.5 | 0.670124013 |
|           | 0.03829691 | 5.057253863 | 0.0548505 | 0.078223543 |
|           | 0 | 5 | 9 | 0 |

| c         | 7.278261114 | 6.672257475 | 0 |
|           | 0.058206761 | 7.278261114 | 0.079005778 |

**Figure 2E – Source data from Series 3 on Introspection**

So how can diagram emotional? First we need to understand there is a set for the conscious and subconscious parts of the system and we use two diagrams for each with the same vectors as noted in the aforementioned diagrams in particular the Plutchik method.
Now if we plot the states we get a set of diagrams like this:

Figure 2F – Example Graphing Introspection Experiment
This graph system is simple to visualize what the system is feeling albeit the nuances of what each one means is still somewhat abstract but easily to visualize which is why the ICOM project settled on this method.

In this particular study we are looking at a similar matrix as used in previous research but now we were introducing the introspection where we can see the effect of the action bias on the emotional state. This particular study also showcases the resolution that the system quickly goes to where we have subtle changes that are or can be reflected by the system in a way we can see via this diagramming methodology. In a working situation items are selected based on how things map to interest and needs and how it affects the core state of the system.

Further given this and the related body of research we can see that even having the same input out of order will cause a different end result and given the volume of input and the resolution of the effect of retrospection and manipulation of interests therefore no two systems would likely ever be the same unless literally copied and then would stay the same only if all of the subsequent input would the same including order. Small differences over time could have dramatic effects millions of cycles later.

This nuanced complexity is why the diagramming method has become so important to understanding ICOM behavior.

SUMMARY

The Emotional Modeling used in the Independent Core Observer Model (ICOM) Cognitive Extension Architecture is a methodology or ‘pattern’ for producing a self-motivating computational system that can be self-aware where emotional modeling is the key to the operation of ICOM. While ICOM is as a system for abstracting standard cognitive architecture from the part of the system that can be self-aware it is primarily a system for assigning value on any given idea or ‘thought’ and based on that take action as well as producing on going self-motivations in the systems further thought or action. ICOM is at a fundamental level is driven by the idea that the system is assigning emotional values to ‘context’ as it is perceived by the system to determine how it feels. In developing the engineering around ICOM two models have been used based on a logical understanding of emotions as modeled by traditional psychologist as opposed to empirical psychologist which tend to be based on biological structures. This approach is based on a logical approach that is also not tied to the substrate of the system in question. Using this emotional architecture we can see how using the Plutchik method is used and how that application creates the biases of the system and how it self evolves on its own making the exposure to input key to the early developing of a given implementation of ICOM.
Appendix A – Citations and References

  http://transhumanity.net/self-motivating-computational-system-cognitive-architecture/
  1/21/02016 AD
- [7] email dated 10/10/2015 - René Milan – quoted discussion on emotional modeling
- [9] “Feelings Wheel Developed by Dr. Gloria Willcox”
  http://msaprilshowers.com/emotions/the-feelings-wheel-developed-by-dr-gloria-willcox/
  (2/20/02016) in article titled: Deeper Mind 9. Emotions by George Norwood