 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

Engineering Kindness: Building A Machine With Compassionate Intelligence

Cindy Mason, CMT, Ph.D.

University of California, Berkeley

cindymason@media.mit.edu

ABSTRACT

We provide first steps toward building a software agent/robot with compassionate intelligence. We approach this goal with an example software agent, EM-2. We also give a generalized software requirements guide for anyone wishing to pursue other means of building compassionate intelligence into an AI system. The purpose of EM-2 is not to build an agent with a state of mind that mimics empathy or consciousness, but rather to create practical applications of AI systems with knowledge and reasoning methods that positively take into account the feelings and state of self and others during decision making, action, or problem solving. To program EM-2 we re-purpose code and architectural ideas from collaborative multi-agent systems and affective common sense reasoning with new concepts and philosophies from the human arts and sciences relating to compassion. EM-2 has predicates and an agent architecture based on a meta-cognition mental process that was used on India's worst prisoners to cultivate compassion for others, Vipassana or mindfulness. We describe and present code snippets for common sense based affective inference and the I-TMS, an Irrational Truth Maintenance System, that maintains consistency in agent memory as feelings change over time, and provide a machine theoretic description of the consistency issues of combining affect and logic.. We summarize the growing body of new biological, cognitive and immune discoveries about compassion and the consequences of these discoveries for programmers working with human-level AI and hybrid human-robot systems.

Keywords: Compassion, AI, Software Requirements, Meta-Cognition, Robots, Machines, Software Agents, Affective Inference, Semantic network, Semantic Web, Cognition, Human-Robot Interface, Neuroplasticity, Compassionate Intelligence, Human-Level AI, Human-Robot Interaction, Neuroplasticity, Emotion, Time, EIQ, Truth Maintenance Systems, meditation, Emotion, Biology, User Experience, Heidigger, Deep Learning.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

Engineering Kindness: The First Steps to Build A Machine With Compassionate Intelligence

C. Mason
U.C. Berkeley

"How do you tell the difference between a human and an emotionally intelligent robot? No matter how many times you tell your truly sad story, the robot will cry with you every time." *Dr. Ed Weiss*

Introduction

Repeated interactions with the artifacts we create have a rub-off effect. They are changing us. Recent co-discoveries across a number of different fields support the idea that positive emotion has positive biological effects on us and vice versa so at this juncture in AI it is time to consider the idea of designing compassionate intelligence (CI). The paper outlines a generalized software requirements description and describes the EM-2, a software agent with CI. The work represents first steps towards building machines (robots and software agents) that have a stake in us, by programming with compassionate intelligence based on both state of mind and state of heart. We offer the generalized software requirements description for anyone who wants to pursue this direction of programming in AI/Robotics. This direction is significant not just for AI but for user interfaces, healthcare, education, and design in other fields. We illustrate the CI ideas with code snippets and architectures of software agent, EM-2.

What does it meant to give an AI system compassionate intelligence or have a stake in us? Simply, we mean the ability to program an AI system that makes decisions and actions that take into account positive regard for others and self. We describe software agent EM-2 as a first step to an AI program that represents and reasons not only with logical concepts of mental state but state of the heart. EM-2 components leverage the sensory analysis and multi-agent functions of prior agent systems, however in the paper we focus on elements of EM-2 related to CI. At the programming level this includes multi-level representations and predicates of feelings of self and others as well as logical concepts and the logical concepts about feelings. The work presented here covers several AI concepts and software agent systems: cooperative multi-agent systems technology, emotion oriented programming, common sense knowledge, affective inference, default reasoning and belief revision. We do not presume the agent to "have" feelings, nor do we address this issue here. Rather we address the computational aspects of creating a reasoning apparatus that uses representation of these concepts to accomplish compassionate decision making. Essential to the system is a pro-social agential stance – this includes but is not limited to a) agents do not lie about concepts or feelings and there is common sense knowledge of positive emotion, society and culture.

The motivation for this work is based on a growing body of recent discoveries from social and cognitive neuroscience, psychoneuroimmunology, and genetics indicating humans benefit greatly from compassionate experiences. User experience studies, neuroplasticity and genetic plasticity studies indicate we are literally changed by repeated interactions with objects and relations in our environment. With our growing symbiotic relationship with gadgets (phones, cars, robotic assistants, browsers, appliances, IoT, etc.) there is significant biological imperative to intentionally design for humane and pro-social AI systems and interfaces. Its not to say all AI needs such features, but that *we have a choice when it makes sense to do so*.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

The technical components of the paper are divided into three parts. In part I we present a generalized software requirements guide for building AI systems with compassionate intelligence and then detail the EM-2 implementation of the first three design requirements, namely: a) agent philosophy of mind and architecture that supports it b) a representation and reasoning calculus that supports notions of combining emotion, standard logic, self and other and c) an inferencing and learning component that supports a) and uses b) to make decisions and take action based on consideration of self and other. In Part II we increase our exploration of the issues of combining affect and logical fact components in multiple agents by deepening the descriptions of algorithms and code snippets given in Part I with a machine theoretic architecture and description of the introspective machines for beliefs and feelings of EM-2, including the predicates for introspection and meta-cognition, conflict resolution and naming. Part III of the paper discusses social and technical issues surrounding this topic such as human-level AI, compassionate intelligence, etc.

Our programming approach to building agents with compassionate intelligence includes 1) multi-agent programming methods for the representing, communicating and reasoning about multiple agents cognitive state(s) during decision making and for being responsive to changes of agents cognitive state over time *with a pro-social stance* 2) organizing models of memory and mental state as semantic graphs (triplets) that refer to the ontological meaning of agent objects and concepts but also according to the emotional meanings of those objects and concepts which are relative and change over time 3) incorporating emotion and logical concepts or triplets into a forward chaining inferencing calculus that supports the revision of inferences over time using a modified version of the DATMS algorithm (Mason and Johnson, 1989) that supports multiple possible worlds - belief revision that permits an agent to simultaneously maintain more than one world or context for objects in memory and finally, 4) revising beliefs about objects and concepts over time occurs not just according to logic but also according to semantics, namely, common sense knowledge and theory about personality and psychological, social and culture. This last feature is how we accomplish a pro-social stance.

Many features of human-level compassion will be wanted in an artificial agent, some will not. To motivate and ground our discussion here we ask the reader to consider applications requiring social or compassionate intelligence – examples include robotic nurses and nursing assistants, robotic nannies, automated intake and triage systems for psychiatric departments, user interfaces and dialogues for vehicle control panels, gaming characters and avatars, education and training software, customer relations support and so on.

Biology Of Positive Emotion And Human Level AI

In the past decade, science has shown kindness and pro-social behaviors have a biological imperative. The following are but a few of the growing body of important co-discoveries occurring across a number of fields:

- The rate of wound healing is affected by domestic conflict/emotional happiness (DeVries et. al. 2007).
- Lower cardiovascular reactivity is related to warm partner contact (Kiecolt-Glaser et al. 2005).
- Brain glucose metabolism is affected by psychosocial stressors (Kern, et. al. 2008).
- The creation of neural stem cells governing short term memory and the expression of genes regulating the stress response are positively affected by motherly affect (Meaney, et. al. 2001). Dr. Meaney's work has received the Royal Order of Quebec, among many other awards and inspired the public health agencies in Canada to begin investigating more formally the role that motherly nurturing has on human health and the need to quash aggression in our families and trusted social circles.
- Positive cognitive state influences positive immune response and vice versa (Azar 2001) (Davidson et al. 2003) (Wager et al. 2008). Its called the immune-brain loop.
- Patient recovery time following bone marrow stem cell transplant was reduced on average 21 days using a compassion based therapy called psychophysiophilosophy (Mason et al. 2009).

From the above findings we can see the evidence. Repeated interactions with the artifacts we create rub off on us. They are shaping and affecting us continually. Social and emotional relations influence our brain, our genes, our stress reaction and

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

immune system and even wound healing. These findings are significant not just for AI design but to user interfaces, healthcare, education, and design intention in other fields.

In (Mason, 2010) I argued that creating and designing artifacts that support positive emotion such as kindness and compassion are essential to the goal of human-level AI. Part of the worry for "full" AI (Tegmark, 2015) (Gaudin 2014). is that the system will not have a stake in itself, or in us. Theologian, Andew Porter, explains this concept of being aware of and confronting issues of self, or having a stake in ourselves, as essential to the "being" part of "human being" (Porter 2014). It originates in Heiddegger's concept *dasein* presented in Being and Time (Heidegger 1962), but has deep relevance to recognizing that being human, AI or otherwise, involves a sense of self that is possible because of our relation to "other" that we, by our very nature of being, have a stake in others and therefore by this observation, human-level AI should have a stake in itself and other. As described by Heidegger, it is an "ontological priority." This can be accomplished by designing AI systems with components for compassionate intelligence.

Further, positive emotion design is a practical and essential when sending robot companions to tend our aging parents, care for patients and children or accompany us into outer space - the first 100 candidates for a one way trip to Mars are in training (Contrera, 2015). Because the matter of building a hybrid robot/AI-human society is underway, we find ourselves at a juncture in AI engineering and design where positive social regard and interactions is essential. This rub-off effect *is* taking place, what we do about it is up to us.

Aristotle's Blind Spot

Aristotle designed great methods for legal reasoning, deductive reasoning and rational thought. These concepts and methods are still useful today but have a basis for "knowledge" that was separated from emotion. To deepen the intelligence of an AI system, we must move beyond this perception of knowledge to include other kinds of knowledge and sensing intelligence. (Mason and Kenchen, 2009). Most systems of western study and education are rooted in original notions of knowledge from Aristotle's teachings where knowledge referred to natural and enduring things such as optics, geometry and astronomy. In this context mathematical descriptions and deductive reasoning make sense. But these academic subjects do not involve emotional experience of a human being and they certainly did not include dasein. It is a blind spot.

Since much of AI is based on these original methods, it is natural that for many decades emotion was *a blind spot in AI*. In light of modern discoveries about emotion's role in cognition, immunology and neuroscience, it becomes clear that emotion has become a necessary and powerful new foundation from which all of AI can and should be redeveloped. This means that the human sciences as well as social, emotional and cultural intelligence/common sense needs to be programmable for a pro-social, human-level or "full" AI.

One of the fundamental differences in "new AI" or "full AI" systems –will come from how we redefine "knowledge". As we spend more and more time interacting with software and robotics, we find ourselves running into brick walls with this legacy view of knowledge. Our new computational view of "knowledge" is expanding to include the emotional, subjective, and sensory life experience of a concept, thing or relation to a thing (including self and other) (Mason and Kenchen 2009). Such concepts were not historically considered "knowledge" in western culture because these are experiential notions and subject to change. Recognizing that concepts relating to state of mind are a programmable and important kind of knowledge is likely to have implications beyond AI into law, healthcare, education and so on. For our purposes, it is a practical necessity. People and their experience of things and each other need to be represented and reasoned about in AI systems because we now live in a symbiotic union with technology. It is only logical that at this juncture either we are aware of this shift and build technology to support and empower our humanity through compassion or we continue in the same direction and technology will change us to support it.

Emotionally intelligent AI systems encompass an aspect of intelligence that was neglected in AI from the beginning of the field – namely, emotional, social and cultural intelligence. We have a stake in how AI technology, whether emotionally intelligent or not, changes us. The viewpoint of the author is that a first principles approach to AI that includes emotional intelligence and knowledge will create a new generation of AI algorithms and data structures that can be applied to all of the existing problems of AI - including pattern analysis, machine learning, robotics, vision, natural language, search, and more, but with a very important distinction. It will have a stake in us.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

PART I

SOFTWARE DESIGN REQUIREMENTS FOR COMPASSIONATE AI

The software design requirements to create an AI program capable of a deeper kind of intelligence, with compassionate decision-making, learning or social interaction has three requirements:

1. The intentional stance or philosophy of mind of the agent is a positive intention toward others – this intention will reflected in the design of the agent architecture, including explicit common sense knowledge, conflict resolution, etc. as well as prioritizations and analysis in hybrid sensing systems.

2. Agent Architecture, Representation and reasoning, sensing apparatus that support explicit notions of self and other.

3. An inferencing and/or learning component that
   a. supports the representation and or sensing of affect and the agent philosophy of mind
   b. supports response, awareness, and/or reference not only to its representation of self during sensing/reasoning but to that of other agents' mental state and maintains consistency within its representation of mental state(s) across hierarchical levels, agents, social organizations or other structures.

4: Knowledge of social, emotional, psychological or other affective components that guide the application of 1-3

Humans incapable of empathy or compassion are often categorized with mental illness and are generally considered anti-social if not dangerous by others (sociopaths and psychopaths). Historically, AI machines have no intentional capacity for empathy.

Overview of the Requirements and Software Agent EM-2

Requirement 1

Requirement one for compassionate intelligence is a positive or pro-social philosophy of mind. The EM-2 agent architecture is based on meta-cognition inspired and influenced by the philosophy of Vipassana or "insight meditation." Vipassana (also known as mindfulness) is an ancient mind training (meditation) system from India that concentrates on the development of kindness and compassion using awareness or non-judgmental observation of one's own thought stream (Salzberg, 1995) (Rhys-Davids, 2003). No matter what one's individual thoughts may be – a grocery list, thoughts of work, and traffic, or pre-occupation with feelings such as worry, anger, fear, elation, etc. the mental training towards each thought is one of acceptance, as if each thought is like a cloud passing overhead in the sky. The practice is to regard each thought with kindness. So similarly, we can easily imagine a software agent that represents and regards objects of mental state in this way.

As an agent designer, it gives one confidence in choosing "insight meditation" as a philosophy of mind because the mental processes have been documented to create compassion in even the most hardened criminals (Ariel and Menahemi, 1997). The process of non-judging observation of one's own mental thoughts in humans seems to allow the aspects of consciousness important for kind behavior. For example, it is described as generating "loving-kindness, friendliness, benevolence, amity, friendship, good will, kindness, love, sympathy and active interest in others," see http://en.wikipedia.org/wiki/Mettā.
he architecture of mind is described in detail in a variety of places, most readily accessed through the audio Series Talks (Fronsdale, 2003) at Insight Meditation Center in Redwood City, California (http://www.insightmeditationcenter.org/), a sister center to the Cambridge Insight Meditation Center organized by Jon Kabot-Zinn (http://www.cimc.info/), or see (Gunarana, 2002; Salzberg, 1995; Hanh, 1976; Rhys-Davids, 2003).

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

Requirement 2

We address the second requirement, namely the agent architecture components of representation and reasoning apparatus that supports self and other in mental state, by reusing previously developed multi-agent program components, where agents often explicitly refer to other agents. Our multi-agent architecture components come from two Lisp-based systems: ROO – a Rule-oriented cOOperative agent language used for hybrid AI/signal processing systems and from EOP - an Emotion-Oriented Programming language used to build EM-1(Mason 1998), a single agent system that explicitly represents and reasons about emotion and mental state. Essential components of mental state for an agent in a multi-agent system include representation of both internally and externally generated beliefs, the ability to distinguish by name the agents' beliefs and those generated by other agents and a mechanism to keep track of them in working memory as they become co-mingled during decision making, problem solving, or learning. Emotion Oriented Programming (EOP) is a rule-based language that supports explicit representation of emotional concepts such as mood and feeling as part of agent mental state and can access ontologies with various emotion concepts such as personality. EOP was used in emotion categorization of a large visual image archive and has also been applied to behavior analysis. Initially it was programmed with the Enneagram ontology to represent aspects of personality.

Requirement 3

We address the third requirement of compassionate AI in EM-2 with a new kind of inference called affective inference where emotion can be the antecedent to logical consequences in a hybrid sensing forward chaining rule- based system. Pragmatically each cycle of the system can then respond to sensing or sensing/data analytics as well as input from users and agents, much as a simple rule based system. The inspiration for affective inference is from an 18th century philosopher named David Hume. Hume, like many other common sense philosophers, was obsessed with understanding the origination of thought. He proposed that thought is a consequence of something felt in the heart. Consistency maintenance of mental state and with the agent community is accomplished with a program component known as a Truth Maintenance System (TMS). Generally, TMS systems can be complicated to both program and explain, much like illustrating a garbage collection algorithm in an object-oriented system. It is of concern however, not just algorithmically but semantically and conceptually. The manner in which an agent maintains consistency becomes an issue of personality and also its relationship to the group of agents. We have created a TMS that is knowledge based, so the meaning of consistency is not just logical but can be based on knowledge/ontologies.

Tying the three requirements all together is meta-cognition. EM-2 has rule-based meta-cognition involving representation and reasoning about its own beliefs, feelings, and the beliefs and feelings of others. Namely, the cognition of EM-2 includes a meta-representation of mental state objects that supports thinking about thinking, thinking about feeling, and thinking about thoughts and feelings – its own and/or those of other agents. The Programming language, predicates, objects, affective inferencing and common sense knowledge are built on LISP.

Requirement 4

The 4th and last design requirement involves the collection and articulation of social, cultural, emotional, psychological and other affective knowledge. Generally the approach involves explicit representation of knowledge through ontologies – tuples, semantic graphs and common sense. The creation of common sense knowledge and ontologies can be hand crafted, acquired through crowd sourced web interfaces or accomplished by some blend of both. Digital signal processing methods, such as with brain maps or other sensors, could also be a reasonable approach to representation. Because this is a topic of significant interest and complexity, we cover this requirement of the software agent in a separate paper [Gil et al. 2015]. There are some emotional concepts included in Open Mind Common Sense project [Singh, et al. 2002] (Speer, Havasi and Lieberman, 2008, a crowd sourced common sense collective, and the related project Open Heart Common Sense focuses primarily on common sense of happiness (Mason and Smith, 2014).

In the next few sections we give more detail on several aspects of the construction of EM-2, focusing on requirements 1-3. In Part II we expand the discussion of agent mental state and consistency while reasoning with affective inference. We give fragments of code from the language EOP (Emotion Oriented Programming). EOP was used to build the first pass at a software agent that could represent and reason with both emotional and mental state, EM-1 (Mason, 1998). To build EM-2, we extended EM-1 to incorporate multiple agents and a belief revision or TMS system that uses a psychologically realistic, but irrational approach to consistency. The agent architecture of EM-2 and meta-level predicates and processes of EOP are based in part on previous work on multi-agent systems (Mason, 1995) and on the meta-cognitive architecture supporting insight meditation or Vipassana.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

Requirement 1

EM-2'S PHILOSOPHY OF MIND

Recently ancient philosophies have become important in western culture. Mind training practices based on eastern philosophies (e.g. meditation) have come under the scrutiny of FMRI and other diagnostic tools and have shown that when humans engage in persistent mind training there is permanent changes in brain structure and function as well as a positive effect on mental and physical health (Begley, 2007; Lutz et. al., 2004). A dramatic example of this idea is the practice of TUMMO (Crommie, 2002; Benson, 1982). Crommie (Crommie, 2002) describes and illustrates Harvard researchers monitoring a TUMMO practitioner (a Buddhist monk) who uses mental meta-processes involving compassion to create dramatic body heat. The effects of the TUMMO practice were physically demonstrated by the semi-nude practitioner who sat for prologued period on ice without harm.

Vipassana, or Insight Meditation, has been practiced for 2500 years. Vipassana meditation practitioners engage in an active mental process of observation or visualization of mental objects, often with a representation of self, along with meta-processes that effect transformation in behavior, state of mind, affect, and or body function. At the heart of meditation is the engagement of a cognitive process known as mindfulness. Mindfulness is simply the act of paying attention to what we pay attention. The following example by Fronsdale illustrates this point (Fronsdale, 2003).

Consider that while you drive, you are concerned with the "doing" of driving – noticing signs, other cars, staying in your lane, and also with "reasoning" processes in navigation, planning and scheduling. While we look out for signs, cars, and children, we are not concerned with and not noticing our dirty windshield. Stopping for gas, we clean it and get back in the car. We do this as part of routine driving. As we start driving again, we become aware of the extra effort that was used to see objects and cars, and the road, due to the dirt.

Mindfulness is accomplished with a meta-level process sometimes called the "observer". The "observer" process is cultivated over a period of time to be attentive to thoughts, body sensations, and feelings, without judgment, resistance, or clinging. Regardless of the moral or ethical semantics of the objects, regardless of the difficulty of the emotion or the lack of apparent rationality of the object, the intentional stance in Vipassana towards these mental objects by the practitioner is gentle reflection.

This intentional stance, namely, the encouragement to notice when we cling or resist and encouragement to allow rather than judge, creates a safe mental space where anything that arises in our consciousness can be met with positive acceptance almost as a mother to a child. In this context thoughts, feelings, and body sensations that would otherwise never be noticed come into awareness, creating "insight" into oneself and others, hence the name "insight" meditation. The mindfulness practice is said to gradually dissolve the barriers to the full development of our human wisdom and compassion.

Vipassana As Computation: Meta-Cognition

Natural systems of cognition have always been inspirational to AI researchers (e.g. vision, memory, locomotion.) Cultures where human mind training has evolved for hundreds and thousands of years present an untapped resource of ideas for researchers working towards human-level AI or compassionate intelligence. Many of these special mind training methods use an architecture of mind that is based on meta-cognition and meta-processes similar in structure and function to the diagram developed by Cox and others (Cox and Raja, 2011) as shown in Figure 1.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

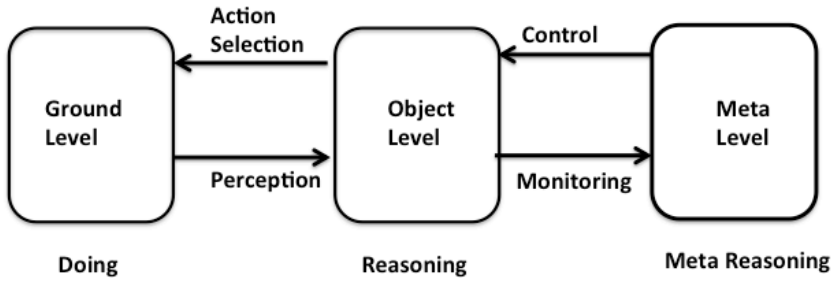


Figure 1. The architecture for meta-mind used for computation and in describing Vipassana.(Adapted from Cox and Raja).

In Vipassana, the idea of non-judgment towards an object is well suited for a computational model and the "observer" could readily be described as a meta-level mental process for being attentive to agent mental state. We describe the process of Vipassana using the architectural components of Figure 1.

The Ground Level of perception involves the human physical embodiment of consciousness - sensory and body sensations such as the perception of inhaling or exhaling, the sense of contact with a supporting structure such as a chair or cushion, smell, sound, pain, pleasure, heat, coolness, itching, etc. At the Object Level is our "stream of consciousness". It includes the thoughts, feelings, or other mental representations of our ground level perceptions as well as daily chatter, problem solving, self-talk, mental imagery, and so forth. The Meta-level consists of the observer. The "observer" is involved in the creation of Meta-Level mental objects concerning the stream of consciousness at the Object-Level. Examples of Meta-Level objects include labels for individual feelings or thoughts, labels for patterns and groups of thoughts and feelings, questions about thoughts, questions about feelings, images, self or other mental objects.) These meta-level objects regarding the stream of consciousness are used in various ways to direct attention to either Object Level or Ground Level for the purpose of Noticing and Observing or for the purpose of answering a question such as "What happens if I pay attention to the pain in my back?" The result of Noticing, Observing, or Asking Questions is the creation of more objects at both the Object and Ground level. Thoughts or realizations about those mental objects sometimes give rise to what is called "insight." Hence the name, "insight meditation." "Insight" about one's thoughts, feelings, or breath (body) can and do profoundly change the state of mind and the state of heart of the practitioner.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

Requirement 2

AGENT ARCHITECTURE

Compassionate Intelligence is the capacity of an agent to act in a compassionate manner in a bounded informatic situation. Essential to the act of compassion is empathy – the ability to consider the possible (emotional) world of another. The computational components of EM-2 relating to empathy include a) the separate and explicit representations of feelings and beliefs about a proposition b) the ability to represent propositions and mental states of other agents. These features are discussed in the following sections.

The architecture for an EM-2 agent is shown in Figure 2. We use the name A_i to represent an arbitrary EM-2 agent and we subscript the agent name since it will be important later to distinguish among the agent A_i (self) and other agents. The architecture for A_i supports the Vipassana as Meta-Computation as follows.

The "Observer" is effectively the execution of Inferencing and Learning (IL) and Truth Maintenance (I-TMS) processes on Meta-Object Memory (M) which contains Feelings and Beliefs about Objects in Object Memory (O) and on Objects. Objects are created through this process as well as by ground level actions that include both sensing (S) and communication (C). To be clear, Meta-Objects refer to Objects, and can be expressed as a predicate applied to an object. For example "Feels(Object)" or "Believes(Object)" where Object can be any physical or mental construct. Objects may be reasoned about and manipulated independently of the beliefs and feelings about those objects that reside at the meta-level (M). Reasoning and learning processes can be applied as both object-level and meta-level computation. In theory, we may create an arbitrary number of levels, especially in a multi-agent scenario, e.g. Feels (Feels (Object)), Feels (Feels (Feels (Object))) and so on, with endlessly rising levels. For simplicity, our discussion focuses on two levels.

In addition to sensing and communication, new objects may come from the common sense repository and as a result of learning. Because our agent resides in an environment with other agents capable of meta-cognition, the communications can be at both object and meta-level. In this way, an agent can "learn" what another agent "Feels" or "Believes" about an object. The context based truth maintenance system (I-TMS) works to maintain consistency among feelings and beliefs using common sense knowledge and knowledge based models of personality. In addition to communication, learning takes place as a result of conflict discovery and resolution (I-TMS) and also as a result of the common sense collective (CS). The focus in our chapter is on the mechanism for compassionate computation.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

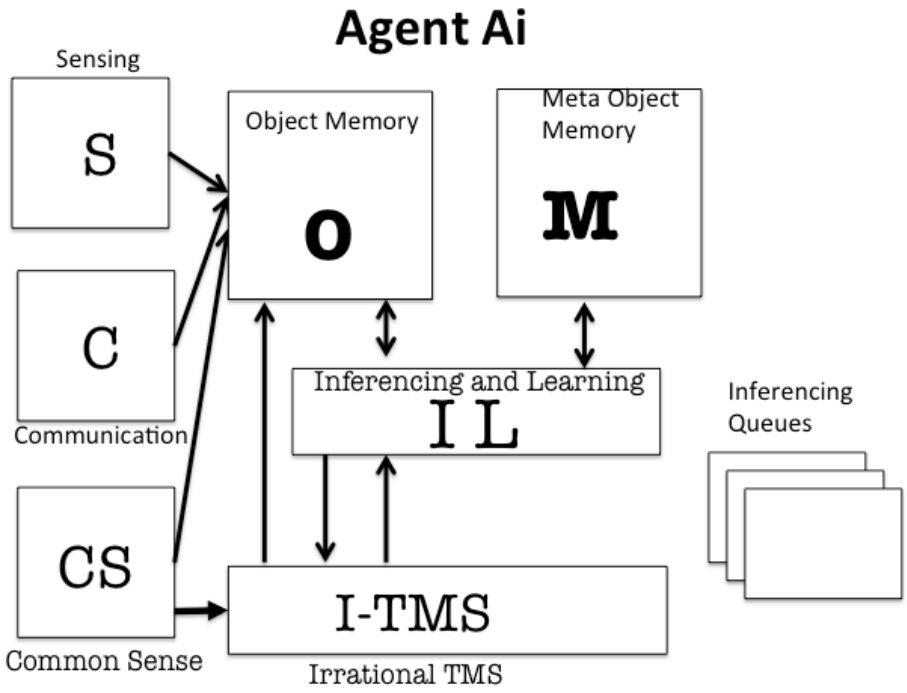


Figure 2. The component architecture for an EM-2 Agent.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

Requirement 3

AFFECTIVE INFERENCE

There is no question that in the natural course of thinking sometimes an emotion can give rise to a thought, or that thoughts may also give rise to an emotional state, which in turn gives rise to more thoughts or further emotional states, and so on. The meta-cognition process of insight meditation highlights the interdependency of feelings and thoughts in human cognition. Many 18th century "common sense" philosophers such as Hume, Locke, and others spent considerable time on the issue of affect and rational thought. As we approach the goal of creating compassionate intelligence, it is essential to have some form of affective inference. By this we mean

Definition: *Affective Inference is a method of inferencing whereby emotion can be the antecedent to a consequent thought, and vice versa.*

This style of inferencing presupposes a programming language where an agent's mental state contains explicitly named objects of emotional concept such as personality, mood, emotional state, disposition, attitude, and so on in addition to traditional non-affective concepts and objects, and that these emotional objects require a separate but interdependent computational apparatus.

Affective inference differs from logical inference in some important ways. First, by its very nature affective inference is volatile – moods and emotions change over time and so will the inferences that depend on them. Second An essential component of an affective inference machine or agent is a truth maintenance mechanism. Typically, consistency refers to the idea of preventing logical theories involving P & ~P. However, because the nature of affective mental objects is based on relative consistency and not necessarily logical consistency we require a non-logical TMS. We require an Illogical-TMS, an I-TMS, where consistency is based on what "makes sense" relative to a number of factors, including common sense knowledge, personality, and logic. It is through this method of consistency maintenance that we are able to create agents with behaviors and decisions that reflect the experience of the world rather than a formal model of the world. We agree that this is perhaps unusual approach to consistency maintenance, as compared to traditional algorithms for both single (deKleer, 1986; Doyle, 1979) or multi-agent systems (Mason and Johnson, 1989; Bridgeland and Huhns, 1990), however, our intention is the creation of agents with an emotional stance.

EM-2's affective inferencing mechanism is based on reasoning with defaults or assumptions. The style of reasoning is based on knowledge about what is normal or commonly expected. For example, "men are tall" or "birds fly." A significant amount of common sense knowledge has been gathered about emotions through an open collective project on the common sense of happiness begun in 2008 called Common Sense (Mason and Smith, 2014) and about objects in the world in an open collective at MIT called Open Mind Common Sense (Speer, Havasi and Lieberman, 2008). These knowledge collectives are processed into machine readable form. We are actively working on integrating this knowledge into EM-2. In both Open Heart and Open Mind, the common sense knowledge is filtered and reformatted from user input to semantic networks where objects or concepts are represented internally as triplets or assertions for ease of processing by other algorithms such as truth maintenance.

Truth Maintenance Systems look for conflicts between the assertions that actually reside in memory (mental state or reality) and the expected or anticipated assertions of common sense knowledge. As shown in Figure 2, the I-TMS relies on two subsystems, each representing the antecedents or support for each object of mental state with their own style of consistency maintenance. The first is IM – an Introspective Machine that represents the logical thoughts of agent(s) at the object and meta-object level and maintains consistency using traditional notions of logic and truth maintenance as described in (Mason, 1995). The second subsystem is IH: an Introspective Heart that represents the feelings of the agent at the object and meta-object level. Unlike the IM, it maintains a relative consistency based on an agent's psychological and social rules as determined by knowledge about personality and cultural information.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

Conflicts between logic and feelings can lead to indefinite loops, resulting in thrashing during the I-TMS algorithms. While this is a rich issue with philosophical and computational implications, we chose to solve this problem with knowledge and by allowing the possibility that an agent programmer may need more than one possible solution. That is, which of the IM or IH systems holds the upper hand in the I-TMS and therefore in the inference process depends on the application, personality, social, and cultural aspects of the agent. For example, agents with a logical personality and a romantic personality will have different contradiction resolutions and hence differences in objects and the feelings about those objects represented in mental state. Before we engage in the detail of IM and IH, we illustrate this concept by example. In the following section, called "Love is Blind", we demonstrate the point described by Minsky, that "Love can change our point of view, turning blemishes into embellishments," (Minsky, 2006).

EXAMPLE : LOVE IS BLIND

We demonstrate the idea of Affective Inference using EM-2 language constructs. The example illustrates the use of explicit representation of how an agent "feels" about a proposition. The example is interesting because it gives a different outcome depending on the personality of the agent.

On start-up Agent EM-2's mental state contains the rules R1, R2, and R3 along with the premises, P1 and P2. Mental Object A1 is created as an assumption using the premise P1 and the rule R1. The object of the agent's love feeling, Peppy, is declared to be handsome precisely because the agent loves Peppy. Next, mental object D1 is derived by the application of rule R3 to A1 and P2. That is, if someone (Peppy) is handsome and that someone (Peppy) proposes then we create the mental object "Accept-Proposal(Peppy)."

R1: IF (Feels (In-Love-With(x))) then
(Assume(Handsome(x)))

R2: IF (Believe (Obese(x))) THEN NOT(Handsome(x))

R3: IF (Believe (Proposes(x))) and
(Believe (Handsome(x))) THEN Accept-Proposal(x)

P1: Feeling(In-Love-With(Peppy)) {{P1}} $\mathcal{B}$

P2: Proposes(Peppy) {{ }} $\mathcal{B}$

A1: Handsome(Peppy) {{A1}} $\mathcal{B}$

D1: Accept-Proposal(Peppy) {{A1}} $\mathcal{B}$

So far, all the objects in memory are currently believed, as indicated by status "$\mathcal{B}$". Now suppose we learn

P3: Obese (Peppy) {{}} $\mathcal{B}$

Then later derive

D2: NOT (Handsome) {{P3}} $\mathcal{B}$

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

as a result of the rule R2 and P3. There is now a conflict between D2 and A1. Unlike a traditional TMS, the manner in which this conflict will resolve, and the resulting set of mental objects that are believed or disbelieved do not depend solely on logic but on common sense about the personality and social/cultural make-up of the agent as well. The heart (IH) and mind (IM) both engage in the process of relabeling.

Agents with a logically inclined personality have meta-rules in IH that allow it to trump the IM. In an agent with a logical personality, we would then find there is a contradiction, and both A1 and D1 will become disbelieved. We also make a record of the contradiction involving A1 and D1, since this, too could change in the future.

A1: Handsome (Peppy)	{{A1}}	$\mathcal{D}$
D1: Accept-Proposal (Peppy)	{{A1}}	$\mathcal{D}$

Agents with personalities interested in attachments, loyalty and a tendency towards socializing will give preference to feelings in a conflict. A romantic agent's IH subsystem would not contradict this.

A1: Handsome (Peppy)	{{A1}}	$\mathcal{B}$
D1: Accept-Proposal (Peppy)	{{A1}}	$\mathcal{B}$
D2: Not(Handsome (Peppy))	{{P3}}	$\mathcal{D}$

EXAMPLE DISCUSSION

In the Love is Blind example, there are 3 rules, R1, R2 and R3. Premises, denoted by P, are observed, felt, sensed or created at agent start-up. In the example, Love is Blind, there are two kinds of premises. Those based on feeling and those based on logic. Notice that P1 is a feeling object as distinct from logical objects P2 or P3 and has the label "{{P1}}". A feelings premise allows the agent to begin a chain of reasoning based on an initial feeling, as consistent with common sense philosopher David Hume's standpoint on human reasoning. The labels for P2 and P3 are both "{{}}" and contain only the empty set, which is a typical premise label for an agent using logic oriented TMS.

The reason for label distinction among premises here is because feelings and logical facts are treated differently during the truth maintenance process and by the meta-operators. These ideas are discussed further in the next section. Assumptions are denoted with the letter "A," and derivations, denoted by the letter "$\mathcal{D}$". Each premise, assumption and derivation has a label indicated by brackets "{}" which contain the mental contexts in which the object has been derived. It is used by the Truth Maintenance System to determine the contexts in which it may be believed/disbelieved and in explaining how the object was created. Next to the label is a meta-tag indicated the agent's belief status for the mental object. The I-TMS provides the ability to introspect about a mental object from the left hand side of a rule through the query operators "Believe," "Disbelieve," "Known" and "Unknown." "Believe" indicates there is a context in which all of its support is Believed. "Disbelieve" indicates there is no support for the mental object. "Known" indicates the object exists in mental state in either "Believe" or "Disbelieve" status, while "Unknown" indicates an object is not "Known". There is also an operator called "FEELS" which allows the agent to introspect on its feelings about an object. In our example we concern ourselves simply with the Believe, Disbelieve and Feels operator.

In our example, when the meta-operator Believe is applied in the evaluation of the left hands side of a rule, when those objects are marked $\mathcal{B}$ the clause will evaluate as true, and become executable. If at any time the belief status changes from $\mathcal{B}$ to $\mathcal{D}$, it is no longer executable.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

AGENT-ID, SELF AND OTHER

Several programming issues arise when objects from multiple agents are contained in a single agent's memory. The first is in how to identify an agent's own cognitive objects from others' communicated cognitive objects. The second issue is how to determine two objects, each from separate agents, are meaningfully linked. The solution to these two issues are related. In an open system, it is potentially useful to resolve the object identification issue using UID for each object, where agents are assigned a range of UID's, much like the assignment of IP assignments. Fortunately, our system was simplified by the fact we are working with a closed system – the agents are a known population of agents and agent identifiers are used to tag each memory object. A generic triplet ((Agent-ID:Object-Index) (Agent-ID:Object-Index) (Agent-ID:Object-Index)) and a specific triplet or assertion ((Agent-Jeeves:12)(Agent-Jeeves:14)(Agent-Jeeves:15)) The indices are hooks into the objects in memory. This formulation of objects as lists with elements that have a two-part name satisfies the inferencing system as well as the consistency maintenance algorithm. Each of these is described in detail in (Mason and Johnson, 1989) and (Mason, 1994).

The second issue in requirement 3 involves semantic linkage. How does an agent that receives a snippet of another agent's memory relate it to existing objects in its own memory? The answer is that in a general open system with multiple agents, this is a potentially difficult (unsolved?) issue involving ontological and vocabulary matching algorithms, access to ontologies in the cloud and so forth. When there is a closed trusted system and where agents engage in communication with a shared vocabulary in a closed community of agents, as is likely the case with most scenarios involving compassionate AI systems, the problem becomes solvable with standard multi-agent systems approach. We have created several cooperative multi-agent systems with this approach to naming and semantic linkage (Gallimore, et. al. 1999).

When an agent sends a snippet of it's memory, it creates a list containing a series of assertions that represent a portion of its semantic graph. Each element of the list is composed of known and shared vocabulary. The LISP programming language allows both the access to C++ programs at run time and built-in functions that convert lists to strings and vice versa. When an agent transmits a message, it converts the list containing the memory snippet into a string that is communicated via a communication agent written in C++. The C++ language has a straight forward set of communication libraries for connecting symbolic agent names and sockets, and interfaces to LISP via foreign function calls. The execution of C++ programs is possible through either the right hand or left hand side of an inference rule. This was a key feature in supporting multi-agent message transmissions.

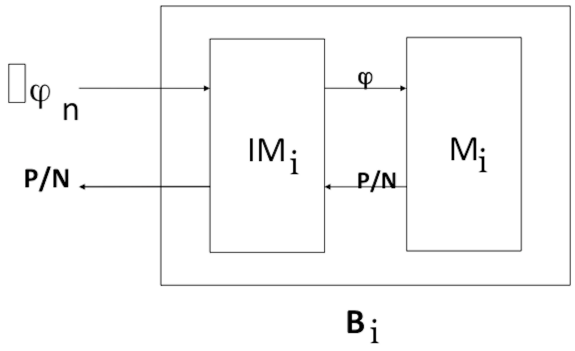
Part II

MENTAL STATE OF A COMPASSIONATE AGENT

The Introspective Machine (IM) Subsystem

The response of the introspective belief machine IM_i is based on matching ϕ against the network of concepts and processes held in mental state. Queries of the form $\Box\ \phi$ presented to IM_i by other consciousness functions can be answered by presenting ϕ to the machine M_i. This machine theoretic formulation of computational introspection is intuitively appealing as it appears to mimic human introspection. While multiple levels of introspection are possible, because meta-objects can be considered as objects, for our purposes a single level of introspection suffices.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015



Cindy Mason

Figure 3. The logic subsystem IM of the I-TMS maintains what is believed.

As shown in Figure 3 the logical beliefs $\mathbf{B_i}$ of agent A_i, can be described as a two-level introspective machine, with an introspective component $\mathbf{IM_i}$, and the belief machinery, $\mathbf{M_i}$, that maintains consistency and determines whether a concept is a member of mental state. It may also detect the contexts or relative links to other concepts as needed. In order to access the contents of mental state other consciousness functions may submit queries to the belief subsystem posed in computational language L whose exact form is inconsequential except that there must be an explicit reference to an agent's mental state. In a healthy agent mind, a mental concept ϕ will be provided to other consciousness functions if and only if there is support for ϕ, which means agent A_i believes ϕ. We represent these expressions by the form $\Box\ \phi$. In general, ϕ may represent concepts created by consciousness functions of A_i or another agent, A_n, as when ϕ has been communicated. We use the notation ϕ_n to represent a concept originating from agent A_n when the origin of ϕ bears relevance to the discussion.

The Introspective Feeling (IH) Subsystem

The following is a machine-theoretic description of the mental state subsystem IH when the agent evaluates a rule containing a query $f(\phi)$ regarding the agent's feelings about antecedent ϕ :

$f(\phi)$ $\quad$ $H(\phi) :P \rightarrow IH(f\phi): P$

$\quad$ $H(\phi) :N \rightarrow IH(f\phi): N$

$\neg f(\phi)$ $\quad$ $H(\phi) :P \rightarrow IH(\neg f\phi): N$

$\quad$ $H(\phi):N \rightarrow IH(\neg f\phi): P$

When faced with the query $f(\phi)$, IH poses the query ϕ to H, and simply returns Positive if H says Positive, and Negative if H says Negative. From the cognitive perspective of the agent, "Positive" means that the agent has considered its set of feelings and has concluded that it has favorable feelings about ϕ and therefore that it feels ϕ. In other words, the agent double checked with its heart component of mental state and is positive it feels ϕ. When presented with $\neg f(\phi)$ IH will respond Negative if H says Positive, indicating that that agent does feel ϕ. "Positive" reply from IH means that the agent does not feel ϕ. The agent does not feel that ϕ so $\neg f(\phi)$ is part of mental state.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

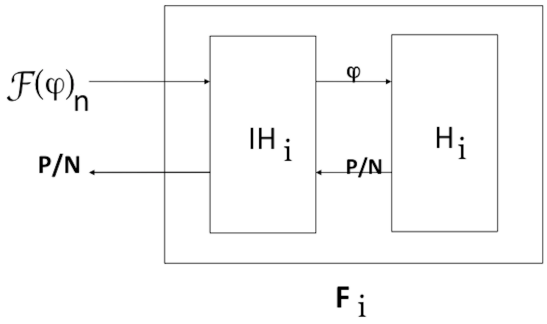


Figure 4. The Feelings Subsystem, F_i of the I-TMS maintains relative consistency among interdependent feelings.

We have now defined an agent's cognitive outlook in terms of its state of positive or negative feelings on the proposition ϕ. Together the set of positive feelings and the set of negative feelings constitute what is felt by the agent. We may define the set of known feelings as the set of ϕ that satisfy a query in L of the form $f(\phi) \vee \neg f(\phi)$. That is a feeling about ϕ is known to the agent whether $f(\phi)$ is believed or disbelieved. We define the "known feelings" modal operator $\Im$ as follows:

$$\Im\phi : f(\phi) \vee \neg f(\phi)$$

that is, the set of all ϕ that are in the agent's feelings list regardless of state of belief in its feeling toward ϕ. It follows that the set of unknown feelings is the set of ϕ that satisfy a query in L of the form $\neg(f(\phi) \vee \neg f(\phi))$. We define the "unknown feelings" modal operator with $\neg\Im$ as follows:

$$\neg\Im\phi:\neg(f(\phi) \vee \neg f(\phi))$$

that is, the set of all propositions ϕ for which the IH "shrugs its shoulders" – it answers negative to both $f(\phi)$ and $\neg f(\phi)$ (alternatively, you may chose to implement the concept of "unknown" feelings with both positive, or by creating a "don't know" state, etc.) The idea of the unknown feelings operator is to describe an agent that has feelings but is neither positive nor negative towards ϕ (humans might refer to this as indifference, being neutral, or undecided.) It is important to distinguish between an agent that has feelings but simply does not know what they are and an agent with no feelings. In the latter case, the operator $\Im$ is undefined.

The presence or absence of $\Im$ in agent mental state is a means by which agents may be divided into two camps: those that have the capacity to reason with feelings and those that do not. Presently most agents fall into the latter category.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

Communicated Feelings

In an agent with compassionate intelligence, propositions in the feeling system may occur not only as a result of affective inference and knowledge as discussed in the previous section, but also as a result of communication. That is A_i believes $f\phi_n$ as a result of a previous communication of $f\phi$ to A_i by A_n.[1] The set of propositions an agent has feelings about includes not only locally generated propositions but also propositions shared by other agents. It follows that agents may reason about another agent's feelings about a proposition as well as its beliefs. This is the heart of compassion and of indifference – to consider and take into account or not the feelings of another agent when engaged in reasoning, planning and scheduling, decision making, and so on.

It is possible that A_I also feels $f\phi_n$, that is, $ff\phi_n$ - in this case, the agent might be called empathetic. We could describe this $f\phi_n$ $f\phi_i$ where semantics(ϕ_n) ~ semantics(ϕ_i), if agent A_I believes $f\phi_n$ where $n \neq i$ (it believes something about the feelings of another agent) then agent A_i believes that agent A_n feels ϕ. It is possible that A_i also feels $f\phi_n$, that is, $f\phi_i$ - in this case, the agents feel the same.)

Consistency Maintenance

When agents reason with feelings, as when reasoning with defaults or assumptions, the inferences that an agent holds depending on those feelings may be retracted over time when feelings change. A special problem may arise in distributed multi-agent systems when agents use communicated feelings in their reasoning processes. Compared to logical belief, feelings can be relatively more volatile. The distributed state gives rise to a situation where an agent's feelings change after they have been communicated. In this case collaborative agents may be "out of synch" (in humans we refer to this as "out of touch."). Using our model the situation may be described as:

$$\text{For } A_i \, IM_i \;\; (f(\phi_n)) \text{: Positive and For } A_n \, IH_n \;\; (f(\phi_n)) \text{: Negative}$$

where A_i believes that A_n feels ϕ, because it was previously communicated or derived, but in A_n mental state it does not feel ϕ. The situation is remedied once A_i receives notification of the change by A_n, but not without some cost.

Agents reasoning with affective inference may require increased demands for computation and communication depending on the degree of persistence of agent state (agent personality), the dynamics of the domain such as frequent sampling of hardware devices measuring affective state (for example of drivers or pilots) and the level of inconsistency robustness – how the need to maintain continuous coherence between internal state and external environment impacts agent behavior. In general, TMSes can be computational intensive, and like traditional or non-affective inferencing, alternative solutions and improvements will be needed as a result. Due to space constraints we do not address many issues nor can we discuss topics in depth. For extended discussion of the topic of inconsistency robustness see (Hewitt and Woods, 2015).

[1] We assume agents are not lying to one another.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

Part III

Compassion, Human-Level AI, Social Issues

The mechanisms for reasoning with regards to another's feelings only makes sense if there is wisdom to go along with it. This is a very important point. For a machine to engage in our world with a compassionate stance, we are faced with the task of articulating the common sense of compassion. Not all engineers and scientists are born with the gift for empathy, sympathy or compassion. We require collaboration with educators, psychologists, mothers, priests, our pets and even the kindness of strangers, to achieve the level of interaction that would enable the compassionate stance in a computational machine. The idea of programming our interfaces and embodied agents with a compassionate stance has great potential for positive influence in our cultures.

> "If you go into a music shop and pluck a string of a violin, each of the other instruments in the store will resonate with that sound. Similarly human beings can resonate with each other to such an extent that they can exchange understanding at a subtle level."
>
> Rollo May
> Healers On Healing

Such is the influence of the objects and machines in our environment as well. It will be important to know what kind of training or what kind of morals we can rely on for a machine in our environment as more and more robotic companions appear in society. To this extent, we began collecting self-reported common sense knowledge from the general public regarding happiness, kindness and self awareness (Mason et al., 2006). We are also moving forward with a collection of memes and bemes as they relate to the creation of a positive cyberpersonality and cyber consciousness. For an interesting discussion of bemes, memes and cybernetic consciousness see (Rothblatt 2006). Efforts to collect crowd sourced common sense, including or exclusively for emotional intelligence are underway, see (Gil et al., 2015, Singh, et al. 2002). Please feel free to visit this website and contribute your knowledge: www.openhearttreasures.org.

In contrast to the majority of current research in AI, the position of the author is that human-level AI programs must not only reason with common sense about the world, but also about feeling and sometimes with irrational reasoning, because every human being knows that to succeed in this world, logic is not enough. An agent must have compassionate intelligence. The heart of what it means to be both human and intelligent includes compassion and empathy, social and emotional common sense, as well as more traditional methods of AI suitable to tasks.

> "Whenever we change our emotional states, we find ourselves thinking in different ways. Our minds get directed toward different concerns, with modified goals and priorities and with different ways to describe what we see. Thus Anger can alter the way we perceive, so that innocent gestures get turned into threats. Love, too, can change our point of view turning blemishes into embellishments. Love also alters how we behave; The heart of what it means to be both human and intelligent includes compassion and empathy, social and emotional common sense, as well as more traditional methods of AI suitable to tasks."
>
> Marvin Minsky
> The Emotional Machine

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

Agents that can represent and reason about the feelings of self and of other agents in decision making and inferencing are a necessary step towards human-level AI. Computing with compassionate intelligence requires the development of a reasoning apparatus that can adapt mental state over time according to changes in feelings of the self and others. We approach the problem by developing affective inference as a means of common sense default reasoning. Unlike traditional logical inference, affective inference involves the justification of facts based on emotion and vice versa. Emotions can bootstrap the creation of objects but because emotions are particularly sensitive to the passage of time and the "consistency" of emotions often has more to do with social and emotional psychology and integrity. We have developed a mechanism for revising beliefs and feelings over time that maintains consistency not necessarily in a traditionally logical fashion but in accordance with social and emotional semantics. This opens a very large can of worms theoretically in terms of understanding the properties of such a system and we will continue to work on developing deeper understanding of these real issues. However, the application of this idea to engineering and engendering of kindness in cyberspace cannot be started soon enough. In support of the US White House project on anti-bullying in cyberspace, which involves a number of departments (Education, Health and Human Sciences) and universities, we proposed one way this can be supported is by creating memes and bemes. Memes and bemes reflect states of mind of users and evolve over time.

The work presented here does not address a number of controversial issues involving emotions – namely the origination of emotion, psychological theories of emotion and so on. Our belief in the engineering of tools despite a lack of consensus on practical and that building systems helps to shed light on issues. Human potential is greatly expanded when emotional support is provided, and we are inspired to engineer systems with emotional intelligence by our experience with remarkable patients using emotional therapies while recovering from bone marrow stem cell transplant procedures at Stanford Hospital [Mason 2004].

The agent architecture and processes of EM-2 support the philosophy of mind that emotion can give rise to thought and should be explicitly represented in the inferencing process. Like reasoning with common sense knowledge, emotions are subject to change over time and any inferences or concepts created that are based on those feelings or common sense knowledge must be revised. To accomplish this we introduce the idea of an Illogical-TMS (I-TMS) where psychologically based justifications may be used to support belief or other feelings about a proposition as well as traditional logic. We chose a knowledge based semantics or psychologically based approach to conflict resolution in the I-TMS in part because many of the domains we work in, such as very big scale data analysis, present such knowledge as a way of resolving conflict. In general however, an agent programmer may need more than one way to approach the conflict resolution problem. A traditional logic based approach to resolving conflicts between logic and feelings in the I-TMS can lead to indefinite loops, resulting in thrashing during the I-TMS algorithms. The algorithmic approach to conflict resolution is a rich issue with philosophical and computational implications and deserves further work but is not an easy problem. Meta-cognition is central to both the semantics and the syntax/logic based approach to conflict resolution and these issues easily relate to the ideas of Goedel, Turing and others. We hope to continue this work and welcome collaborations.

Meta-cognition is also central in representing and reasoning about another agent's beliefs and feelings. The ability to explicitly represent the beliefs and feelings of the self as distinct from another agent in mental state while consider several possible worlds is central to the ability to reason with compassion – that is, to taking into account another agent's feelings and beliefs during reasoning, planning, decision making and problem solving. It is very important to notice that it is not enough simply to have these mechanisms. The intention of what to do with the feelings and beliefs of another agent (human or machine) is the cornerstone of creating compassionate agents. While EM-2's architecture reflects the influence of the meta architecture of mind in the Vipassana meditation practice, there is a connection between what we observe at the meta level and what we do with it. These naturally take place over the course of time. As humans, our observations and awareness at the meta-level or observer level are in hindsight but lead to an opening of the heart. No matter what the agent architecture or knowledge structure, this will not happen in a machine unless we make it so. Meta-computation is thus an essential component in reviewing past behavior.

We believe that compassionate intelligence is a necessary but not sufficient means to create artificial general intelligence. It is essential for the next generation of games, film, telecommunication and many applications involving social interactions. If we do not provide affective mechanisms for computer systems with close human interaction, we will miss a great

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

opportunity for improving the human condition as well as build systems that are blind to many forms of intelligence. As mentioned earlier, endowing robots with the traits of friendliness, benevolence, amity, friendship, good will, kindness, love, sympathy and active interest in others is desirable. However, it must be said that the decision to give an agent affective computing depends entirely on the context and the application, as illustrated in John McCarthy's science fiction story, The Robot and The Baby, (http://www.formal.stanford.edu/jmc/robotandbaby/robotandbaby.html). It is entirely possible that if we create synthetic mind that we will also need synthetic mental health practitioners, e.g. robot psychologists.

It is worth noting that in our paper, we have often used the term human level AI and compassionate intelligence as dual terms. Although one might object to this duality it is the opinion of the author after having digested much about brain anatomy and become certified in a medical system where emotions and physical health are inseparable, as well as having continued to monitor the evolution of neuroscience perspectives on the importance of emotion in all aspects of life, that one cannot reach human level intelligence in an AI system without compassionate intelligence. That said, although compassionate intelligence is necessary for human level AI, they are perhaps distinct concepts. Along the same lines, while our three-leveled architecture of mind may appear to have the power to differentiate 'feeling' and 'emotion', a philosophical discussion of their distinction, in terms of knowledge and perception or consciousness, will be left for another time. We believe the distinction does not affect the usefulness of the mechanism. It is possible the perspective that EM-2 agents are designed to act similar to human level reasoning by taking into account beliefs and desires, that they can be evaluated as an evolutionary approach to goal-driven agents, is useful.

We are expecting to find that psychologically valid computational models of mind and emotion are useful in very large scale data analysis. We are currently applying working on a case study to emulate our emotional reactions to features in very large data sets with the hope that we will avoid many computations. The applications for this kind of model are quite open ended. Indeed, Infosys is well on its way to building a research lab around the idea of emotion oriented programming and affective computing. We continue to develop EM-2 with more features that take into account multiple agents and communication and its implications on the I-TMS. In general, the problem of I-TMS and multiple agents is computationally intense and provides a number of philosophical and practical problems, much like trying to keep track of what someone else feels and thinks in order to take that into account. Not an easy feat, even for humans. Perhaps, if we can create machines that can keep track of others' feelings and beliefs, it will be a helpful social tool, much like our calendar or contacts list.

ACKNOWLEDGEMENT

There are many people to thank - the teachers and students that carry lessons over the centuries - they are so special now in modern life; the bone marrow patients at Stanford Hospital who had the courage to listen to their heart and made history; many friends and family have helped make this work possible from the beginning and never stopped helping. Special thanks to John McCarthy, Lotfi Zadeh, Nelson Max and Andrew Porter.

REFERENCES

Ariel, E. & Menahemi, A. (1997). *Doing Time, Doing Vipassana*, Karuna Films.

Azar, B., (2001). A New Take on Psychoneuroimmunology, Monitor on Psychology 32-1:34.

Begley, S., (2007). *Train Your Mind, Change Your Brain: How a New Science Reveals Our Ability to Transform Ourselves*. New York: Ballantine.

Benson, H., Lehmann, J., Malhotra, M., Goldman, R., Hopkins, J., & Epstein, M. (1982). Body temperature changes during the practice of g Tummo yoga. *Nature Magazine, 295,*

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

234 – 236.

Bridgeland, D. M. & Huhns, M. N. (1990). *Distributed Truth Maintenance*. Proceedings of AAAI–90: Eighth National Conference on Artificial Intelligence. AAAI Press.

Carlson, L.; Ursuliak, Z.; Goodey, E.; Angen, M.; & Speca, M. (2001). The effects of a mindfulness meditation-based stress reduction program on mood and symptoms of stress in cancer outpatients: 6-month follow-up. *Support Care Cancer, 9*(2), 112-23.

Camurri, A. & Coglio, C. (1998). An Architecture for Emotional Agents. *IEEE Multi-Media, 5*(4) 24-33.

Contrera, J. (2015). 100 Finalists Have Been Chosen For a One Way Trip To Mars, Washington Post, Feb. 16, 2015, http://www.washingtonpost.com/blogs/style-blog/wp/2015/02/16/100-finalists-have-been-chosen-for-a-one-way-trip-to-mars/ Accessed 24 Mar. 2015.

Cox, M. & Raja, A. (2011). Meta-Reasoning – An Introduction. In (Cox & Raja, Ed.), *Thinking about Thinking* (pp. 3- 14). Cambridge, MA: MIT Press.

Cromie, W. (2002). Research: Meditation changes temperatures: Mind controls body in extreme experiments. Cambridge, Massachusetts, *Harvard University Gazette:4.*

Davidson, R., Kabat-Zinn, J., Schumacher, J., Rosenkranz, M., Muller, D., Santorelli, S., Urbanowski, F., Harrington, A., Bonus, K., & Sheridan, J. (2003). Alterations in brain and immune function produced by mindfulness meditation. *Psychosomatic Medicine, 65*(4), 564-570.

de Kleer, J. (1986). Problem solving with the ATMS. *Artificial Intelligence, 28,* 197-224.

DeVries, A., Craft, T., Glasper, E., Neigh, G., & Alexander, J. (2007) Curt P Richter Award Winner: Social influences on stress responses and health, *Psychoneuroendocrinology*, 32:587-603.

Doyle, J. (1979). A Truth Maintenance System, *Artificial Intelligence, 12,* 231-272.

Fronsdale, G. (2003). *Introduction to Meditation*, www.audiodharma.org/talks-gil.html.

Gallimore, R. et. al. (1999). Cooperating Agents for 3-D Scientific Data Interpretation, IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews, Vol. 29, No. 1.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

Gaudin, S. (2014) Stephen Hawking says AI could 'end human race'. Computer World. December. http://www.computerworld.com/article/2854997/stephen-hawking-says-ai-could-end-human-race.html *Accessed July 1, 2015.*

Gil, R., Virgili-Goma, J., Garcia, R. & Mason, C. (2015). *Emotion Ontology for Collaborative Modelling and Learning of Emotional Responses*, Computers In Human Behavior, Elsevier. In-press.

Gunaratana, V.H. (2002). *Mindfulness in Plain English.* Boston, MA, Wisdom Publications.

Hanh, T. N. (1976). *The miracle of mindfulness! : A manual of meditation*. Boston, MA: Beacon Press.

Heidegger, M. (1962). "The Ontological Priority of the Question of Being." Being and Time / Translated by John Macquarrie & Edward Robinson. London: S.C.M.

Hewitt, C. and Woods, J. (2015). Inconsistency Robustness. College Publications, 2015.

Hume, D. (1997). *An enquiry concerning human understanding : A letter from a gentleman to his friend in Edinburgh / David Hume*. (Ed.) Eric Steinberg, Indianapolis, IN: Hackett Pub. Co.

Kabat-Zinn, J., Lipworth, L. & Burney, R. (1985). The clinical use of mindfulness meditation for the self-regulation of chronic pain. *Journal of Behavioral Medicine 8*(2): 163-190.

Kern, et al. (2008). Glucose metabolic changes in the prefrontal cortex are associated with HPA axis response to a psychosocial stressor. *Psychoneuroendocrinology* 33: 517–529.

Kiecolt-Glaser, J., Loving, T., Stowell J., Malarkey, W., Lemeshow, S., Dickinson, S., & Glaser, R. (2005). Hostile marital interactions, proinflammatory cytokine production, and wound healing. *Archives of General Psychiatry*, 62-12:1377-1384.

Lutz, A., Greischar, L., Rawlings, N., Ricard, M. & Davidson, R. (2004). Long-Term Meditators Self-Induce high-Amplitude Gamma Synchrony During Mental Practice. *Neuroscience 101*(46): 16369–16373.

Mason, C. (1994). *ROO: A Distributed AI ToolKit for Belief Based Reasoning Agents*, International Conference on Cooperative Knowledge Based Systems, Keele, England. http://web.media.mit.edu/~lieber/Cindy/Cindy-Library/Roo/Roo.html. Retrieved September 2,2015.

Mason, C. (1995). Introspection As Control in Result-Sharing Assumption-Based Reasoning Agents. *Proceedings of the 13th International Workshop on Distributed Artificial Intelligence,* Lake Quinalt, WA: AAAI Press.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

Mason, C. (1998). Emotion Oriented Programming. Formal Notes, SRI AI Group. See also www.emotionalmachines.org. Retrieved February 23, 2015.

Mason, C. (2003). Reduction in Recovery Time and Side Effects of Stem Cell Transplant Patients Using Physiophilosophy. *Late Breaking News Abstract at the Proceedings of the International Conference on Psychoneuroimmunology*, FL:PNIRS.

Mason, C. (2005). Global Medical Technology in Bushko (Ed.) *Stud Health Technol Information (pp. 247-256).* Boston, MA: IOS Press. Symposia for Institute for Future Health CareTechnology, Boston, Mass., 2003. PMID 16301783 http://www.ncbi.nlm.nih.gov/pubmed/16301783, Retrieved February 23, 2015.

Mason, C. (2008). Human Level AI Requires Compassionate Intelligence. Association for the Advancement of Artificial Intelligence National Conference, Workshop Proceedings, 2008, Meta-Cognition, Chicago, Illinois.

Mason, C. (2010). The Logical Road to Human Level AI Leads to A Dead End. IEEE International Conference on Self Adaptive and Self-Organizing Systems Workshop Proceedings *SASOW*, 2010, *Self-Adaptive and Self-Organizing Systems Workshops* (pp. 312-316). Budapest:Hungary. doi:10.1109/SASOW.2010.63. Retrieved Febrary 23, 2015.

Mason, C. (2012). Giving Robots Compassion, Conference on Science and Compassion, Telluride, Colorado.

Mason, C. and Kenchen, T. (2009). The Subjective Experience of Objects, Accepted for Cognition 2012, Nice France, Presented at Association for the Advancement of Artificial Intelligence Workshop on Biologically Inspired Computing, Washington, D.C. https://www.academia.edu/7381740/The_Subjective_Experience_of_Objects accessed Sept. 21, 2015.

Mason, C., Garcia, R., Garcia, R. & Smith, C., (2006). Open Heart Common Sense Project – A Public Collection of the Common Sense of Happiness. Retrieved February 23, 2015 from www.openhearttreasures.org.

Mason, E., Mason, P. and Mason, C., (2009). Haptic Medicine. Studies in Health Technology and Informatics, 2009, 149:368-385, Pub Med PMID19745495.

Mason, C. & Johnson, R. (1989). DATMS: A Framework for Assumption Based Reasoning. In M. Huhns (Ed.) *Distributed Artificial Intelligence Vol.:2*, (pp. 293- 317). London: Pitman.

Meaney, M.J. (2001). Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annual Review of Neuroscience* 24:1161-1192.

May, R. (1989). The Empathic Relationship: A Foundation of Healing. In Carlson & Shield (Eds.) *Healers on Healing* (pp 108-110). New York, NY: Penguin Putnam.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

Minsky, M. (2006). *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind.* New York, NY: Simon and Schuster.

Nass, C. & Moon, Y. (2000). Machines and Mindlessness: Social Response to Computers. *Journal of Social Issues*, 56-1:81-103.

Porter, A. (2014). A Theologian Looks at AI. Proceedings of the Association for the Advancemnet of Artificial Intelligence 2014 Fall Symposium on The Nature of Humans and Machines, Washington, D. C.

Tegmark, M. (2015). Research Priorities for Robust and Beneficial Artificial Intelligence: an Open Letter. Future of Life Institute. http://futureoflife.org/misc/open_letter accessed July 1, 2015.

Reeves, B., & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* New York, New York: Cambridge University Press.

Rhys-Davids, C.A.F. (2003). *Buddhist Manual of Psychological Ethics, of the Fourth Century B.C., Being a Translation, now made for the First Time, from the Original Pāli, of the First Book of the Abhidhamma-Piṭaka, entitled Dhamma-Sangani (Compendium of States or Phenomena).* Whitefish, Montana: Kessinger Publishing.

Rothblatt, M. (2006). Memes, Bemes and Other Consciousness Things. The Journal of Personal Cyberconsciousness, 1-4:1-6. *http://www.terasemjournals.com/PCJournal/PC0104/rothblatt_04b.html Accessed May 5, 2015.*

Speer, R., Havarsi, C. and Lieberman, H. (2008). Analogy Space: Reducing the Dimensionality of Common Sense Knowledge *National Conference on Artificial Intelligence - AAAI* , (pp. 548-553). Chicago, Illinois: AAAI Press.

Salzberg, S. (1995). *Lovingkindness: The Revolutionary Art of Happiness.* Boston, MA: Shambhala Publications. ISBN 1-57062-176-4.

Singh, P., Lin, T., Mueller, E., Lim, G., Perkins, T., & Zhu, Wan. (2002). Open Mind Common Sense: Knowledge Acquisition from the General Public, Springer Lecture Notes in Computer Science, 2519:1223-1237. http://ftp.cse.buffalo.edu/users/azhang/disc/springer/0558/papers/2519/25191223.pdf Accessed May 9, 2015.

Sloman, A., & Croucher, M. (1981). Why Robots Will Have Emotions. *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, Vancouver, Canada.

Sloman, A. (2010). http://www.cs.bham.ac.uk/research/projects/cogaff/cogaff.html. Retrieved February 23, 2015.

 Appears in International Journal of Synthetic Emotions, 6(1) number, June – December 2015

Wager, T. D., Davidson, M. L., Hughes, B. L., Lindquist M, A., & Ochsner, K. N. (2008). Prefrontal-subcortical pathways mediating successful emotion regulation. *Neuron* 59:1037–1050.