

---

# Artificial Compassion – From An AI Scholar

Cindy Mason

Independent Researcher  
Berkeley, California, 94704  
cindymason@media.mit.edu

**Abstract:** This paper describes a new generation of computational intelligence founded on the ancient idea of compassion called Artificial Compassion. The creation of Artificial Compassion is the result of two coinciding historical developments. The first is the increasing discoveries of human sciences in new fields like neuroendocrinology and psychoneuroimmunology. This provides the spark for Artificial Compassion. For example, we once thought with certainty that our brain is fixed for life but neuropsychology and a device called the fMRI have shown it is “plastic”. It changes constantly throughout our lives in response to our experiences. Remarkably, we also now know it is changed for the better through positive emotions like compassion, kindness and happiness. So, too, are the immune, endocrine, genetic, cardio and neural systems influenced and changed by our emotional experiences. This new perspective on emotions and plasticity validates much of ancient wisdom in medical systems outside the west. Long held assumptions about emotion are unsuitable for humanity. The second development is ‘machine rub off’. We are in symbiotic relation with our devices today and we are plastic. We are changed by our interactions but many people have computer rage. We need Artificial Compassion to replace computer rage with positive plasticity.

**Keywords:** Artificial Intelligence, Robots, Compassion, Human Sciences, Positive Plasticity

---

## 1. Introduction

We are living at a time when people and children are in a symbiotic relation with devices as a normal part of daily life. This digitizing social trend has increased as a result of the pandemic. The new field of user experience has show that we are changed by our interactions with our machines and this in turn changes how we treat one another (Nass and Moon 2000; Reeves and Nass 1996). They rub off on us. So if our interactions with our machines or devices are frustrating, then we become frustrated and that frustration is carried to others. What if instead of being frustrated we felt better? What if we had positive plasticity instead?

Technology designed around human science discoveries has the possibility to uplift all of our lives through positive plasticity. Right now AI powered technology is spreading into our devices and lives. But the foundation of AI was not built at a time when users were part of the picture. Most AI was built to solve hard problems, calculating space trajectories. Today AI’s problem solving has risen to the level of world champion, beating out human master players in chess, Jeopardy and Go. But they were not designed with people in mind. This is the motivation to build devices with Artificial Compassion. It can fill what is currently an empty or frustrating space in human-device interaction with positive plasticity.

When we use the term Artificial Compassion it means the software system, algorithm or device has a stake not just in its task but it also has a stake in you. This applies to its decision-making, problem-solving, learning, analyzing, interactions and behaviors. It also applies to its choice of color, tone of voice, behavior and anything that might impact the user. The novel idea of Artificial Compassion as a new foundation for AI can influence and redefine all areas of AI. For example, a machine learning algorithm with Artificial Compassion can spot compassionate voices or faces.

In the paper, *Engineering Kindness*, we give attention to software agents [ref]. In this paper we focus on robotics. The paper is in two parts. In the first part, we explore some of the growing body of discoveries from new multi-disciplinary fields in human sciences that have sparked the idea of Artificial Compassion. In the second part of the paper we take the reader step by step through the process of building a robot with Artificial Compassion starting with an insect robot.

Insect robots have a simple kind of intelligence based only on sensing and reacting. NASA tested out the idea of using insect-like robots for unmanned planetary exploration but these type of robots fail because they are too simple and because of something called a “box canyon”. By adding “thinking” components to the insect robot, we accomplish more sophisticated behaviors and go beyond the limits of sensing and reacting. This kind of robot is similar to what is used in the “robots in the cloud” or software agents Siri and Alexa. Finally, we reach beyond today’s robotics with not just sensing, reacting and thinking, but with feeling, social knowledge and more, in order to accomplish artificial compassion. Before we begin this robot-building journey, it is the AI author’s opinion that we need to learn about the discoveries behind it because they are more important than AI. Why do we want to do this? In the next section we explore briefly some of the key discoveries in human science that are leading the next generation of AI.

## 2. Part I—The Spark

Why do we want to do this? Why do we want to create artificial compassion? What are the possibilities for this kind of technology? We answer these questions by exploring some of the growing body of discoveries in human science that have made visible to us a world of emotions and their impact on our biology. Although the AI author believes the science presented here is of utmost importance, it is perhaps a simple human story that is most helpful. The story is called *The Rescuing Hug*.

### 2.1. *The Rescuing Hug*

The newborn twins shown in Figure 1 (Townsend 2001) spent their first week of life in their separate incubators. It is standard hospital policy. One of the twins was struggling for its life with fluctuating temperature and heartbeat. The nurses had tried everything but the baby was not responding to standard treatment. One of the hospital nurses followed her intuition, risking her job, and placed the two twins together in the same incubator, violating standard hospital policy. Immediately, the stronger twin reached out its arm to hold and comfort its sibling. The endearing embrace stabilized the ailing twin’s heart rate and normalized body temperature. These two babies seemed to instinctively know what scientists are now documenting. Today the twins are healthy grown adults.



Figure 1. Premature twins thrive with a “Rescuing Hug.” (Townsend, 2001).

In the next section, we highlight some of the ways we are changed by positive (and negative) emotions. These studies are impressive. They demonstrate the potent impact of

compassion/positive warm relationships on us humans—from our brains to our blood sugar to wound healing.

## 2.2. Human Sciences

The following examples are part of a growing body of recent discoveries involving positive emotion and its impact on our biology:

1. Motherly affect has a positive impact on the creation of neural stem cells which govern our short-term memory and the expression of our genes that regulate the stress response (Meaney 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.
2. When we have a consistent, warm partner contact we have lower cardiovascular reactivity (Kiecolt-Glaser et al. 2005; Grewen et al. 2003).
3. Brain glucose metabolism is affected by psychosocial stressors (Kern et al. 2008). This is important because the disruption of normal glucose metabolism forms the basis for many brain disorders (Mergenthaler et al. 2013)
4. Our rate of wound healing is slower in high conflict relationships (Kiecolt-Glaser et al. 2005)
5. Positive cognitive state influences positive immune response and vice versa (Azar 2001; Davidson et al. 2003; Wager et al. 2008; Boomershine et al. 2001). Its called the immune-brain loop.
6. Patient recovery time following bone marrow stem cell transplant was reduced on average 21 days using a Japanese therapy founded on compassion called Jin Shin Jyutsu or psychophysiophilosophy (Shannon 2002; Mason and Mason 2009).

These kinds of results even occur in social media. Hate speech on Twitter is worse for cardio health than smoking. Researcher JC Eichstaedt and team looked at Twitter language patterns in a large-scale study (country-level) (Eichstaedt et al. 2015). They discovered that negative language patterns used on Twitter are a significant predictor of age-adjusted mortality in atherosclerotic heart disease. The predictor is more than 10 other well known risk factors such as smoking, obesity and high blood pressure. The Twitter study further showed that positive emotions and engagements appear protective.

In general, it appears love and compassion create a buffer to health and cognitive effects of stress. In fact the chemistry of stress (cortisol) is 'cancelled out' by the chemistry of love—oxytocin (Uvnas-Moberg and Petersson 2005; Quirin et al. 2011; Smith and Wang 2014; Heinrichs et al. 2003; Kirschbaum et al. 1996).

In the next section, we talk about something called "Machine Rub Off"—its the idea that technology rubs off on us and it changes us. I believe that when you hear the studies in that section you will become compelled, as the AI Author has, to believe that to survive we must ourselves become not only aware of compassion's power but that we must use all the tools at our disposal to do this as rapidly as possible. This means designing AI, Machine Learning and Robotics with Artificial Compassion.

## 2.3. Machine Rub Off

Our lives today are filled with interactions with devices—computers, laptops, phones, games, cars, drones, agents/bots, robots, etc. in our environment. All of these devices evolved from machines whose idea of knowledge is devoid of emotion. Originally our computer devices were isolated from the world in a large cold room and solely focused on solving a hard problem like finding a space trajectory or decoding a message. There was no need for user interaction back. There was no brain science. And we certainly didn't know about plasticity. Thankfully this has changed.

New inventions and instruments in the genetic and brain sciences (devices like the MRI and fMRI) have made previously invisible effects on the body visible. Neuroplasticity studies (Davidson and Lutz 2007; Begley 2007) and genetic expression studies (Powella et al. 2013) show we are literally changed by our repeated interactions with objects, relations and thoughts. But there is no awareness of this in AI or other technology design. At this juncture there is no societal awareness of the human sciences research in our interactions with one another.

User experience and human-machine interaction studies show that the way we regard, speak, type and relate to our gadgets spills over into how we treat one another (Reeves and Nass 1996; NASS and Moon 2000). The trouble is many people have computer rage. They are frustrated with their gadgets. They report that they have shouted at, thrown, or hit their gadgets (Grenoble 2015; Carufel 2013; Wardrop 2009). That means there is a lot of frustrated people walking around and that it's contagious. Imagine that instead of being frustrated we could design all these gadgets, our robots included, to support us and to give us kindness and compassion when we needed it. The effect of the interactions would create a positive plasticity rub-off. By spreading the compassionate intelligence around through our gadgets, and letting that rub off on us, we have the potential to uplift humanity, no matter the education, location, race, gender, social or economic level in society. Without positive intentional design that includes human sciences we are slowly creating a future society with frustration instead of humanity. In the next part of the paper, Part II, we introduce the design of a cognitive architecture that allows us to build Artificial Compassion into robots.

### 3. Part II

In *Engineering Kindness* (Mason 2015) we see that programming a computer to have compassionate intelligence is possible. Just like with any other piece of software, once we create an AI software program with compassionate intelligence, it can be copied and it can be shared and spread. The other idea about using artificial compassionate intelligence software is that it can also be '*embodied*'. By embodied I mean the software resides in a hardware 'body'—like a robot. In this section we explore the software that is the mind (and heart) of a robot with Artificial Compassion. It can be useful in many situations and different kinds of devices.

First we review some 'basics' of robotics—what they're made of and how they function. If you already know something about robots, feel free to skip this section. The next section describes two big ideas in robotics software "The Control Loop" and "Cognitive Architectures". These are key concepts for understanding just about any robot and we will use these ideas throughout our discussion as we build up the components of a robot with artificial compassionate intelligence. We do this in the following three sections by looking at three cognitive architectures and their control loops each with increasingly more capabilities. We first explore insect robots. It is helpful to begin by exploring the insect robot because it lays the groundwork for the next two architectures. They have a simple cognitive architecture and control loop but are still useful helpers in some tasks. After introducing the insect robot, we describe the NASA project for using insect robots for planetary exploration to find out just how much they can and cannot do. The next section describes the second architecture with more sophisticated components and behaviors than an insect robot. This takes us closer to the goal of Artificial Compassion but it is only capable of behaviors like the software agents Siri or Alexa. Finally we present the third architecture that has components to support compassionate intelligence.

#### 3.1 The Basics

To put it simply, robots are made up of hardware (the "body") and software (the "mind"). The hardware consists of a chassis that is the home for circuit boards and

sensitive electronics. Not all robots move, but most of them have some kind of locomotion. Robot bodies can roll, walk, fly, swim, etc. so the hardware systems also include wheels or propellers as well as special housing for circuit boards that control them and communicate with the robot body and software. There are also sensors and effectors. Examples of sensors are things like bump sensors, a camera, temperature sensors, and so on. The sensors give the robot awareness of its physical environment, much like our five senses do for us. Project Argo Float has approximately 4000 robots in the ocean right now that sense and monitor salinity, temperature and turbidity (ArgoFloat 2021). Effectors are the way a robot acts or have a physical “effect” in the world, like turning wheels to the left, moving an arm to then grasp or pick and place an object. Some robots can also text or send an email communication. Actions or effects are a reaction or a response to a software instruction or sensor input and any other helpful information the robot is programmed to use. In an insect robot, the behavior we see or the effect is a simple direct reaction to sensor input. It has no other information. In more sophisticated robots, there is memory, knowledge, learning, problem solving analysis and evaluation and it can access the cloud. In a very sophisticated robot sometimes there’s collaboration with other robots or devices. There can even be human-robot collaboration.

In addition to sensors and effectors, locomotion, and a chassis to hang them on (the hardware), for the software, we need two big ideas: a control loop and a cognitive architecture. These are important concepts for building and describing the robot “mind”. First let’s cover the control loop, then the cognitive architecture.

### 3.2. *The Control Loop*

The main way a robot’s “mind” works is to repeatedly execute the instructions of what is known as ‘The Control Loop’. The control loop consists of a computer program or code whose function is to take the input from the sensors and feed it to the robot ‘mind’ for processing so that it can then create the effect or response. It does this repeatedly and indefinitely, unless it is told to stop or runs out of power. The same software executes over and over, continually taking input and feeding it to the robot, continually creating an output in response to the input. That’s why its called a loop or a cycle. You might also think of the robot’s control loop like an engine because an engine has cycles where it continually fires to make power. This loop or cycle effectively controls how the robot behaves. So it makes sense that it is called a ‘control’ loop because it controls the robot. When we observe a robot, what we see is a result of the execution of the control loop.

In a way, the relation between computers and programs is similar to movies and scripts. When we watch a movie, what we are seeing is a result of actors executing the script. Some scripts are better or more sophisticated than others. Some actors have more action or expressive features than others. The script doesn’t come ‘alive,’ that is, we can’t ‘see’ the script until the lines in the script are executed by actors. Computer programs/code are like a script. It exists on a piece of paper or in computer memory. It doesn’t come alive until the lines in the program execute on a computer. What we see is a result of the executing code. That is not to say robots are at all like actors. The point is that without software, a robot would be quite still. The control loop keeps the software running indefinitely. It stops when you turn it off or when the robot runs out of juice/electricity. When you recharge it or turn it back on, the control loop starts up again executing repeatedly. How you build the control loop depends on the cognitive architecture.

### 3.3. *The Cognitive Architecture*

A cognitive architecture for a robot is like an architectural blueprint for a house. A cognitive architecture lays out, describes and arranges all the components of a robot’s cognition, similar to the way a blueprint for a house describes and arranges all the rooms and doors of a house. You can move or walk through a house by the paths from one part

of the house or room to another e.g., the garage is connected to the house to reach the kitchen walk down the hall and go through the dining room, etc. An architect calls this 'flow' and it is often studied with care because the result of the flow in a house affects how you feel about the house or how well it serves its purpose. A cognitive architecture indicates flow and interconnection among the cognitive components of the robot 'mind' with layout of arrows indicating the flow or order of execution among cognitive elements. Almost every robot in the world is designed with some kind of blueprint like this. Figure 2 shows a cognitive architecture for an insect robot. This is the same architecture of the robots we used at NASA for planetary exploration in unknown terrain. This is called a sense-act architecture or sense-react architecture. As Figure 2 shows the insect architecture, its components and flow, is quite simple.

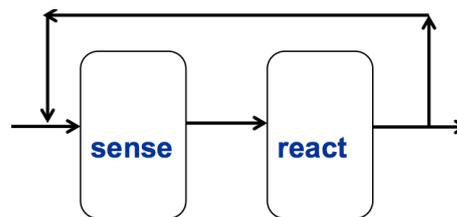


Figure 2. The Sense-React Cognitive Architecture.

#### 3.4. Insect Robots

As a Ph.d student at Stanford, Rod Brooks began to ask fundamental questions about the nature of intelligence in living creatures other than people. He invented the sense-react architecture to build robots that imitate the behavior of small insects. They could only do very simple things like react to a sensory input. As shown in Figure 2 this is accomplished with just two boxes and a few arrows. One box represents sensory input and the other is an action output box. There's an arrow from the outside world coming into the sense box, representing sensory information coming in from the robot's environment/the world. There's another arrow leading directly from sensing into reaction. The last arrow represents the reaction happening into the robot's environment or outside world. Notice there is a circuit or a loop. This is the control loop. For an insect robot the control loop has 3 steps:

- Step One: Sense
- Step Two: Act
- Step Three: Go To Step One

So its basic operation is Sense-React, Sense-React, over and over. Each time the loop executes, the new sensor input will override the sensor input from before. So the insect robot's actions are always based solely on what it just sensed. There is no memory of the previous sensory input over time. It can't recall what happened in the previous loops, nor can it plan for the next action. There's no cognitive component to archive, reason, plan, learn, or anything else. It certainly doesn't know what a user is. It cannot name or discriminate objects or people. To create action in the world, these kinds of robots simply react to their senses because they have no memory, knowledge, or language.

##### 3.4.1. Intelligence Is in the Eye of the Beholder

Why would anyone think this is *intelligent*? Rod Brooks made an interesting and important philosophical observation in AI when he developed insect robots: "intelligence is in the eye of the beholder" (Brooks 1990). If you stop and watch simple insects like ants, you'll see them doing all sorts of amazing things. As humans, we might judge insects' intelligent behavior not by a single lone insect but when they are together in lines, colonies, or swarms. They build sophisticated houses and work collectively. If you stop and consider Brooks' idea more generally, we can imagine there must be many kinds of

intelligence, not just one or two. In my opinion, compassionate intelligence is very important, because of its capacity for good in a world that has so much suffering. In many eastern medical systems the idea that compassion is good for us is quite ancient and this idea is now supported by modern science as we described in Part I. I believe this is why it is important to bring artificial compassion into the world, as soon as possible and in as many ways as possible. It could be argued that compassion is highest form of intelligence in people because it empowers our cognitive abilities and our relationships necessary for our survival. To take the next step in architecting artificial compassion for robots we look at where the cognitive architecture of the insect robot breaks down when it enters a box canyon (or the end of a hall way).

#### 3.4.2. Box Canyons

When reactive robot architectures first arrived on the robot scene they were considered to be an advantage over conventional robotics because they were fast—the sensor input goes straight through to the reaction output. There's no 'thinking'. Compared to a conventional robot, you could also build them simple, cheap and fast. For example, you can have a stripped down version of a robot chassis with some really simple sensors like bump sensors or light sensors that go straight into hardware for reacting. This reaction is very simple. For example, if the reaction is to move, say to turn left or right, the reaction might be to flip a coin and turn the front wheels left or right. If there are 4 choices for movement—straight, left, right or reverse, a random number generated between one and four would work. If the robot flies or swims, there are more choices for movement, but its still a simple roll of the dice (random number generator) on which way to go. To find out how far a robot with this simple architecture can function we need to send it on an exploratory job for a rescue mission.

Suppose we need to send a robot into a disaster area to search for survivors that could help save a human life. A robot could potentially be a hero there. The trouble with using a robot with a reactive architecture in this scenario is that the way it figures out what direction to move can get itself stuck in something called a box canyon. A box canyon in nature is a 3-sided dead end or a wall with a small opening—the robot can't fit through it. In an office, you can encounter this box canyon scenario when you walk to the end of the hallway and there's no exit, it's a dead end. So let's say you're the insect robot and there is a wall in front of you and to the left and right—its a box canyon situation. Because the choice of movement is a random choice between Left or Right its entirely possible to get stuck by turning up a sequence of Lefts and Rights that just don't turn you around. Here are two examples of movement sequences that leave you stuck {Left, Right, Left, Right} or {Left, Left, Right, Right}. This can possibly go on indefinitely. In the least, its not reassuring to see a robot behave this way on such an important mission. At the worst, such sequences of moves lead to failure.

I experienced this problem first hand when I worked on robots at NASA Ames in California, USA. We had three reactive architecture robots intended for unmanned planetary exploration. NASA had created a simulated lunar and simulated Martian environment, with special rocks, dirt and dust to simulate planetary conditions. What we found through these experiments was that there were many times when the reactive robots would be stuck in a box canyon like situation. There was no warning in the users manual about this limitation. It just happens to be true about this class of robot architectures. These bots can detect a bump if they have a bump sensor, and react to a bump sensor but only directly, without any memory. They essentially flip a coin to move in a direction. They're not able to plan, use a map or a GPS. There is no communication. On each cycle of the control loop, it doesn't know if it turned left or right the last time it moved because it has no memory. We need a more sophisticated cognitive architecture that makes it less susceptible to getting stuck and is capable of doing much more.

#### 3.3. Sense-Think-Act Robots

The second cognitive architecture we consider is called as the sense-think-act cognitive architecture. Many robots are based on this architecture. Software agents like Siri or Watson can be described in this way. As you can see in Figure 3, the key new component we add to the reactive architecture is called a “think” component. The think component contains memory that is essential for cognitive processes like learning, analysis, planning, and so on. The think component sits in between sensing and acting so that before acting in the world there is cognition.

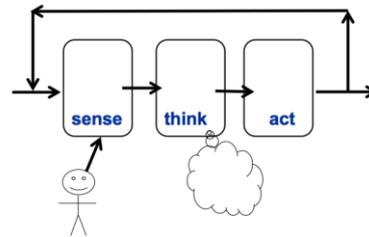


Figure 3. The Sense-Think-Act Cognitive Architecture.

According to the architecture, exactly what kind of cognition is unspecified, but often this includes running statistics, accessing a database for machine learning, inferencing, running graphic analytics or all of the above. It can also access the internet. All of that can be contained in this one box called thinking. For the purpose of the cognitive architecture, the type of the thinking component doesn't really matter. It can be a neural net, a concept net or symbolic language, or an archive of numbers. It can have representation of knowledge, statistics or any other kind of advanced AI Programming and some kind of memory.

The cognitive loop for this machine now has four steps:

- Step One: Sense
- Step Two: Think
- Step Three: Act
- Step Four: Go To Step One

We can accomplish a lot more complicated things with this architecture than the one for an insect robot because it has a memory that persists from one loop/cycle of the code to the next. We can access maps, perform complicated analysis and algorithms on sensed data. We can also combine or use information from the past to learn and improve the robot's future behavior. This is enough to conquer the box canyon problem because there is a memory of the previous moves. There is also navigation strategy and planning in addition to any sensory input. It is good at problem solving, planning, prediction, analysis and decision-making. It is good at recognizing patterns and remembering patterns. It could, if we wanted to add a user interface, respond to commands from a person. This is where we are today.

To create compassionate intelligence we need more because the cognitive architecture we've just described, the sense-think-act architecture, has a single perspective—its own. Its design is well suited for tackling problems. But it lacks any kind of knowledge or representation of itself or others. A user interface could be added, say to guide it, but only to the extent it supplies the robot with external directions or input. It still relies on philosophical methods of knowledge designed for rational thinking, not for feelings. It was designed without considering people are part of the picture. This gets us much further than the insect robot, but its not quite enough for our goal of creating Artificial Compassion.

The components of the Sense-Think-Act have no metacognition. It can represent an object in memory. It can even, with the right programming component, tell you why it

made a decision. But what we want to know is how do you feel about that object, how do we compare what you and I feel about that object together? To create artificial compassionate intelligence, we need these kinds of cognitive components and more. And with the human sciences we know today, the historical assumptions of knowledge that gave rise these cognitive architectures without positive regard for users, emotions or anything, should scare us not comfort us. The next cognitive architecture, the third and last architecture presented here, creates an novel approach to AI that sees emotion as knowledge. It is not an organon, as Aristotle would describe, or even a “new organon” as Francis Bacon created with his data collections (Aristotle and Owen 1853; Bacon 2000). It is an emotional organon.

### 3.4. Robots with Artificial Compassion

There are some AI scholars who have worked on building systems that can recognize emotions and others who generate emotional expressions in robot faces. A cognitive architecture capable of artificial compassion requires more than recognition and expression.

The architecture we just described in the last section has three components: a sense component, a think component, and an action component. In addition to these, we are going to need a few more. As shown in Figure 4, the new components include input from users and other agents, a feeling component and a model of self and others as well as the world. As you can see in Figure 4, the sensor inputs go into updating the model of self and others every cycle of the control loop. This is then fed in parallel to both the thinking and the feeling component.

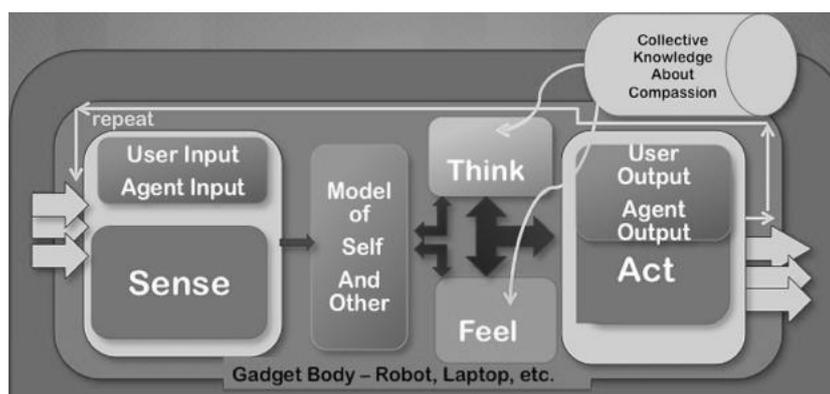


Figure 4. The Sense-Think+Feel-React Cognitive Architecture (Mason, 2012).

These thinking and feeling components have a complicated interaction. They can take input from other agents, bring in knowledge from the cloud about compassion, culture, social knowledge, interaction, context and all of this is then, and only then, brought into a component for action. The action component also includes knowledge about emotion and user output or agent output.

When an action or expression is delivered to the world by the action component, it is no longer devoid of emotion or social/cultural intelligence. The action component uses knowledge of culture, emotion, relation, etc. to consider the impact its action has on others. For example, if the action is communication, it can consider its knowledge of different impacts it will make by its choice of fonts, colors, layouts, tone of voice, et cetera. All these different kinds of knowledge can be used in tasks that need or are intended to be considerate of others and/or self when making decisions, when learning, when giving advice.

Now let's take a look at the control loop. It now has seven steps.

Step One: {Sense, User Input, Agent Input}

Step Two: Update Model

Step Three: {Think, Feel}

Step Four: Repeat Step Two, Three Until Done

Step Five: Pre-Action: Compute User Output, Agent Output

Step Six: Act

Step Seven: Repeat

As we walk through each of the steps above, you will see we can accomplish much more than either the insect robot or the sense-think-act architectures.

Step one is similar to the other systems, but now it includes not just sensing hardware. It includes communication with users, other agents, as well as possibly sensing apparatus about user state in addition to its own environmental sensors.

Step two updates internal models. Here we bring in the information gathered in step one. Any new information will be integrated into the models using algorithms or analytics, as well as possibly maintenance routines or human directive. These are then used by the next component, the think and feel component. Where previously there was a think element, we now have a kind of duplex or a complex element that has its own sub cycle that can itself repeat.

Step three, the think-feel sub cycle component, can involve computations and communications with other agents as well as metacognition. And it includes many different kinds of thinking and feeling components, possibly accessing social, cultural knowledge, knowledge unique to a particular agent it's working with or a user that it's working with, as well as guiding principles that can be at a higher level that integrates across thinking and feeling components.

Each of these elements inside the think and feel component may or may not be computed locally. We have access to the net and agents can talk to one another, name each other, and cooperate. When computations between the think and feel components are "completed" another model update may occur. This is step four. If needed the Think-Feel sub cycle of steps two and three will repeat until the model has settled.

Step five is the pre-action component. It precedes the execution of action in the world. The action created by this architecture is not the effect of a movement or light or voice, but is pre-consideration of that effect. It has social knowledge and positive intent for expression similar to the idea some people have "Think before you speak" or "Don't speak ill of the dead". Step five component gives rise to a considered or smart reaction that is also guided by its possible impact on the user, self, or world. It comes from emotionally intelligent knowledge about the user, preferences, knowledge about the positive impact of colors, fonts, layouts, et cetera. It also has guidelines for having a good impact with gesture, pace, facial expression, tone of voice and anything else the agent has to offer. In step six, we execute this pre-considered action. This action depends on all the steps before it. Finally in step seven, repeat. This takes us back to step one.

There is an important distinction to be made for this cognitive architecture from the previous two. If the action of steps five and six is offensive, could be improved or is wrong, the user can give this feedback as input when it cycles back to step one. There you have it, the control loop for the compassionate intelligence architecture. Our work doesn't stop here. There is a great responsibility for these kinds of systems.

#### **4. Intention—The Most Important Ingredient of All**

In designing such intelligence into robotic and artificial systems, we have a responsibility that these technologies are not used for nefarious purposes. Positive intention and verifiable behaviors are essential. Our human history shows there are times when man's inhumanity to man exceeds our worst nightmares. There are many efforts working to limit destructive capabilities of AI. One such effort is that all agents on a network should be identified and licensed. Another is the Algorithm Justice League (Buolamwini 2021) that proposes the creation of an agency to test technology before it can be sold and used by the public. There are currently there are no agreed upon laws or regulations for a robot's appearance to guarantee or predict its behavior. It is the author's opinion that if an outfit, label, or brand, appears on a robot, it should be truthful and consistent (Lieberman, 2019).

Because in some governments there is corruption and dysfunction, many engineers have begun to develop international standards for engineering of AI systems. The idea is to create a recognizable brand with positive humane ethics that are transparent. You might say that ultimately, it is up to the programmers and corporations themselves because it is their intentions that build the code.

This paper has provided a glimpse of the first steps towards compassionate intelligence in our robots and devices. We hope it will be an inspiration for others.

**Funding:** The project has taken place over many years. Initial funding for research at Stanford Hospital by Silicon Valley Angels. Funding for Haptic Medicine and Human Sciences research is from the Mason Family Foundation. Partial support supplied by the Computer Science Department at Stanford University and the Electronics Engineering and Computer Science Department at University of California, Berkeley. Additional support from the Artificial Intelligence Center at SRI Intl. (formerly Stanford Research Institute) in Menlo Park, California, USA.

**Institutional Review Board Statement:** No human studies performed.

**Informed Consent Statement:** No human studies performed.

**Data Availability Statement:** This is not an empirical study. Author is AI Scholar.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

- (ArgoFloat 2021) ArgoFloat. 2021. What Is Argo? Available online: <https://argo.ucsd.edu> (accessed on 27 April 2021).
- (Aristotle and Owen 1853) Aristotle and Owen. 1853. *The Organon*, Complete ed. Amazon: Aristotle Createspace Independent Publishing Platform.
- (Azar 2001) Azar, Beth. 2001. A New Take on Psychoneuroimmunology. *Monitor on Psychology* 32: 34–36.
- (Bacon 2000) Bacon, Francis. 2000. *The New Organon*. Edited by Lisa Jardine and Michael Silverthorne. Cambridge: Cambridge University Press.
- (Begley 2007) Begley, Sharon. 2007. *Train Your Mind, Change Your Brain: How a New Science Reveals Our Extraordinary Potential to Transform Ourselves*. New York: Ballantine Books.
- (Boomershine et al. 2001) Boomershine, Chad., Wang, Tianyi, and Zwilling, Bruce. 2001. Neuroendocrine regulation of macrophage and neutrophil function. In *Psychoneuroimmunology*, 3rd ed. Edited by Robert Ader, David L. Felten, Nicholas Cohen. New York: Academic Press, pp. 289–300.
- (Brooks 1990) Brooks, Rodney A. 1990. Elephants don't play chess. *Robotics and Autonomous Systems* 6: 139–59. doi:10.1016/s0921-8890(05)80025-9.
- (Buolamwini 2021) Buolamwini, Joy. 2021. Algorithmic Justice League. Available online: <https://www.ajl.org> (accessed on 27 April 2021).
- (Carufel 2013) Carufel, Richard. 2013. Over A Third of Americans Confess to Verbal or Physical Abuse—Of Their Computers: New Harris Study Reveals the Extent of Frustration, Anger and Helplessness Spurred by Computer Problems. *Bulldog Reporter*. Available online: <https://www.bulldogreporter.com/over-a-third-of-americans-confess-to-verbal-or-physical-abuse-of-their-computers-ne/> (accessed on 27 April 2021).
- (Davidson and Lutz 2007) Davidson, Richard J., and Antoine Lutz. 2007. Buddha's Brain: Neuroplasticity and Meditation. *IEEE Signal Processing Magazine* 25: 171–74.
- (DeVries et al. 2007) DeVries, Courtney, Craft, Tara, Gasper, Erica, Neigh, Gretchen and Alexander, Jessica. 2007. Curt P Richter Award Winner: Social influences on stress responses and health. *Psychoneuroendocrinology* 32: 587–603.
- (Eichstaedt et al. 2015) Eichstaedt, Johannes C., Hansen Andrew Schwartz, Margaret L. Kern, Gregory Park, Darwin R. Labarthe, Raina M. Merchant, Sneha Jha, Megha Agrawal, Lukasz A. Dziurzynski, Maarten Sap, and et al. 2015. Psychological Language on Twitter Predicts County-Level Heart Disease. *Psychological Science* 26: 159–69. Available online: <http://positivehealthresearch.org/research/psychological-language-twitter-predicts-county-level-heart-disease>.
- (Grenoble 2015) Grenoble, Ryan. 2015. Colorado Springs Man Shoots Malfunctioning Computer Eight Times, Calls It 'Glorious' The Huffington Post. May 22. Available online: [https://www.huffpost.com/entry/man-shoots-computer-eight-times\\_n\\_7119246](https://www.huffpost.com/entry/man-shoots-computer-eight-times_n_7119246) (accessed on 27 April 2021).
- (Grewen et al. 2003) Grewen, Karen M., Anderson, Bobbi J., Girdler, Susan, S. and Light, Kathleen C. 2003. Warm partner contact is related to lower cardiovascular reactivity. *Behavioral Medicine* 29: 123–30.
- (Heinrichs et al. 2003) Heinrichs, Markus, Thomas Baumgartner, Clemens Kirschbaum, and Ulrike Ehlert. 2003. Social support and oxytocin interact to suppress cortisol and subjective responses to psychosocial stress. *Biological Psychiatry* 54: 1389–98.
- (Kern et al. 2008) Kern, Simone, Terrence R. Oakes, Charles K. Stone, Emelia M. McAuliff, Clemens Kirschbaum, and Richard J. Davidson. 2008. Glucose metabolic changes in the prefrontal cortex are associated with HPA axis response to a psychosocial stressor. *Psychoneuroendocrinology* 33: 517–29.

- (Kiecolt-Glaser et al. 2005) Kiecolt-Glaser, Janice, Loving, Timothy, Stowell, Jeffrey, Malarkey, William, Lemeshow, Stanley, Dickinson, Stephanie, and Glaser, Ronald. 2005. Hostile Marital Interactions, Proinflammatory Cytokine Production, and Wound Healing. *Archives of General Psychiatry* 62: 1377–84.
- (Kirschbaum et al. 1996) Kirschbaum, C., Wolf, O. T., May, M., Wippich, W. and Hellhammer, D. H.. 1996. Stress- and treatment-induced elevations of cortisol levels associated with impaired declarative memory in healthy adults. *Life Sciences* 58: 1475–83.
- (Lieberman, 2019) Lieberman, H. 2019. In Robots We Trust? Association for Computing Machinery Transactions in Information Systems. Feb. 25. <https://medium.com/acm-tiis/in-robots-we-trust-21de883b8e99>.
- (Mason and Mason 2009) Mason, C. and Mason, E. 2009. Haptic Medicine. *Studies in Health Technology and Informatics 2009*, 149: 368–85. Pub.Med.PMID19745495.
- (Mason 2012) Mason, C. 2012. Giving Robots Compassion. Poster Session. Conference on Science and Compassion Telluride, Colorado.
- (Mason 2015) Mason, C. 2015. Engineering Kindness. *International Journal of Synthetic Emotions* 6: 1–23.
- (Mason 2020) Mason, C. 2020. Human Sciences for Technology Design Bibliography. Available online: [https://www.academia.edu/43813498/Human\\_Sciences\\_For\\_Technology\\_Design\\_Bibliography](https://www.academia.edu/43813498/Human_Sciences_For_Technology_Design_Bibliography) (accessed on 9 August 2020).
- (Mason 2021) Mason, Cindy. 2021. Appearance and Reality in AI—The WYSIWIG Law of Robotics and the Case of the Robot Priest. *Religion*. Submitted.
- (McCarthy 2012) McCarthy, John. 2012. The Robot and the Baby. Available online: <http://jmc.stanford.edu/articles/robotandbaby.html> (accessed on 9 August 2020).
- (Meaney 2001) Meaney, Michael J. 2001. Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annual Review of Neuroscience* 24: 1161–92. doi:10.1146/annurev.neuro.24.1.1161
- (Nass and Moon 2000) Nass, Clifford, and Youngme Moon. 2000. Machines and Mindlessness: Social Response to Computers. *Journal of Social Issues* 56: 81–103. doi:10.1111/0022-4537.00153. ISSN 1540-4560.
- (Powella et al. 2013) Powella, Nicole D., Erica K. Sloan, Michael T. Bailey, Jesusa M. G. Arevalo, Gregory E. Miller, Edith Chen, Michael S. Kobor, Brenda F. Reader, John F. Sheridan, and Steven W. Cole. 2013. Social stress up-regulates inflammatory gene expression in the leukocyte transcriptome via B-adrenergic induction of myelopoiesis. *Proceedings of the National Academy of Sciences* 11: 16574–79.
- (Quirin et al. 2011) Quirin, Markus, Julius Kuhl, and Rainer Düsing. 2011. Oxytocin buffers cortisol responses to stress in individuals with impaired emotion regulation abilities. *Psychoneuroendocrinology* 36: 898–904
- (Reeves and Nass 1996) Reeves, Byron, and Clifford Nass. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York: Cambridge University Press.
- (Riseberg et al. 1998) Riseberg, Jocelyn, Klein, Jonathan, Fernandez, Raul and Picard, Rosalind. 1998. Frustrating the user on purpose: Using biosignals in a pilot study to detect the user's emotional state. Paper presented at CHI 98 Conference Summary on Human Factors in Computing Systems, Los Angeles, CA, USA.
- (Shannon 2002) Shannon, Ann R. 2002. Jin Shin Jyutsu Outcomes in a Patient with Multiple Myeloma. *Alternative Therapies* 8: 128. Available online: [https://www.academia.edu/7404110/Haptic\\_Medicine\\_Our\\_Stem\\_Cells\\_Like\\_Compassion\\_-\\_Cancer\\_and\\_BMT\\_A\\_Case\\_Study](https://www.academia.edu/7404110/Haptic_Medicine_Our_Stem_Cells_Like_Compassion_-_Cancer_and_BMT_A_Case_Study) (accessed on 27 April 2021).
- (Smith and Wang 2014) Smith, Adam S., and Zuoxin Wang. 2014. Hypothalamic oxytocin mediates social buffering of the stress response. *Biological Psychiatry* 76: 281–88.
- (Townsend 2001) Townsend, L. 2001. Premature Twins Thrive with a “Rescuing Hug”. Available online: <http://www.nrlc.org>.
- (Uvnas-Moberg and Petersson 2005) Uvnas-Moberg, Kerstin, and Maria Petersson. 2005. Oxytocin, a mediator of anti-stress, well-being, social interaction, growth, and healing. *Zeitschrift für Psychosomatische Medizin und Psychotherapie* 51: 57–80.
- (Wardrop 2009) Wardrop, Murray. 2009. ‘Computer Rage’ Affects More Than Half of Britons. Available online: [Telegraph.co.uk](http://www.telegraph.co.uk) (accessed on 3 November 2015).