A cognitive architecture for human-like and personable AI

Arvind Chitra Rajasekaran, Department of Computer Science, *Graduate Student, University of California, Santa Barbara*

***Abstract*—In this article we will introduce a cognitive architecture for creating a more human like and personable artificial intelligence. Recent works such as those by Marvin Minsky, Google DeepMind and cognitive models like AMBR, DUAL that aim to propose/discover an approach to commonsense AI have been promising, since they show that human intelligence can be emulated with a divide and conquer approach on a machine. These frameworks work with an universal model of the human mind and do not account for the variability between human beings. It is these differences between human beings that make communication possible and gives them a sense of identity. Thus, this work, despite being grounded in these methods, will differ in hypothesizing machines that are diverse in their behavior compared to each other and have the ability to express a dynamic personality like a human being. To achieve such individuality in machines, we characterize the various aspects that can be dynamically programmed onto a machine by its human owners. In order to ensure this on a scale parallel to how humans develop their individuality, we first assume a child-like intelligence in a machine that is more malleable and which then develops into a more concrete, mature version. By having a set of tunable inner parameters called aspects which respond to external stimuli from their human owners, machines can achieve personability. The result of this work would be that we will not only be able to bond with the intelligent machines and relate to them in a friendly way, we will also be able to perceive them as having a personality, and that they have their limitations. Just as each human being is unique, we will have machines that are unique and individualistic. We will see how they can achieve intuition, and a drive to find meaning in life, all of which are considered aspects unique to the human mind.**

***Keywords*—cognitive architecture, imprimer**

I. Introduction

Freud[2] as well as Minsky[1] have postulated the convenience of division of the hierarchial thinking process of humans along the lines of id, ego and superego / instinctive, deliberate and (self) reflective thinking processes. While the accuracy of these divisions has not yet been scientifically established, they implicitly convey guidelines on how thinking can be emulated on a machine. Minksy terms 'agents' or 'resources' as entities stored in our mind which assist us to perform tasks. Minsky says that we humans are able to perform tasks as a result of a collection of these agents working together. Minsky also introduces k-lines, as an aggregation of well defined agents used to solve set and standard problems. Another popular approach which is similar to agents is to view the brain as consisting of pattern recognizers organized in a hierarchy. Most of the work in machine vision and language processing is trying to emulate these pattern recognizers on a machine. Minsky[3] introduces 'imprimers' as people a child wants to impress, typically parents or parental figures, and shows how imprimers, with their appraisal, impact our value system and help us set goals, grow and develop. Thus our individuality depends on nurture by our imprimers and the influence of our peers. In this work we show how human-imprimers and other machines(as peers) can influence machines to develop their individuality. The notion of 'Aspects' will be introduced as the key determiner of a machine's identity.

A. Imprimers

The major imprimers of a child are, essentially, its mother and father in today's society. Since the mother plays a leading role, lets call the mother primary imprimer and since father plays a secondary leading role for the child to look up to, lets consider the father, the secondary imprimer. Thus personality and cognitive development requires two imprimers, a primary imprimer for survival and a secondary imprimer for growth and learning. While the imprimers for children could alternate or even default in playing their role, we will simplify the case for machines with well defined imprimers.

B. The three motivators of human beings

At a very high level, humans have 3 main motivators.
1. The drive to adapt and survive.
(Ensured by innate and learned reactions and our sense of past)
2. The drive to learn and grow.
(Ensured by self reflective thinking)
3. The drive to adapt and procreate.
(Ensured by our sense of future)
Here procreate also means to produce things of value. It can be said that our imprimers impact these systems to some extent. Imprimers mainly impact motivators 1 and 2, namely the drive to survive and the drive to learn and grow. Motivator 3 is mostly left to our peers. In this work we will outline some agents that are constituent of these three domains and postulate how they determine the development of human individuality. Then we will see how machines themselves can emulate these traits to achieve individuality.

C. Aspect, scale and agent

Based on the above motivators we define some aspects which can be considered expressions of these motivators. To achieve individuality in machines, one must achieve some sort of parallel beween the survival, procreativity and growth aspects in humans to similar traits in machines. While these aspects donot necessarily capture everything, we believe they capture atleast the important factors and can act as a starting point. We also hypothesize that the aspects manifest themselves at various levels of intensity, measured by a scale. While there could be questions on why such manifestations occur and not others, our observation is that in areas concerning the workings of the mind, any observations or arguments cannot be conclusively accepted to any degree of certainty. We merely hypothesize our point of view, leaving any conclusive experiments to verification by future software implementations. The various realms and the aspects governing the realms are (see Appendix I)

1) Survival Realm:

- **Attention** towards things leads to developing an **attitude** towards them. The attitude could be described as ranging on a scale of measures such as discretion, indifference, feeling and hypersensitivity.
- **Consciousness** can experience things ranging from otherworldly surrealism to skepticism about reality, which is the captured by the **metaphysical** scale.
- **Arousal** can manifest as a stimulation caused by **ethical** or unethical concerns.

2) Procreative Realm:

- **Steadfastness** aspect can manifest as a drive to be the **backbone** of things ranging from a position of authority to one of inadequacy.
- **Strength** aspect can manifest as **effort**, ranging from resource-hungry burden to idleness.
- **Empathy** aspect can lead to increase in communication or the avoidance of communication.

3) Growth realm:

- **Spontaneity** aspect is likely to be an expression of **scholarliness**, ranging from acumen to sketchy ineptitude.
- **Articulation** aspect can manifest as **ingenuity**.
- **Intuition** aspect can manifest as intuition about **safety**, ranging from habitual instinct to adaptability.

The aspects can be seen as being ensured by k-lines which develop around the scales for each category. By creating a functional dependence of these aspects to stimuli obtained from imprimers one can achieve individuality in machines.

D. The drive to survive

The drive to survive begins as soon as a child is born. The child achieves this objective by being attractive to the primary imprimer, and generally expecting affection from the imprimer. During the later stages of growth i.e. after 5 years, the primary imprimer may criticize or control the child. Also the child may experience a lack of appreciation from the imprimer. Such experiences can lead to reaction formation on part of the child against this imprimer's attitude.
1. Criticism from the primary imprimer may cause the child to divert energies toward demarcating the right from the wrong, to avoid such criticism in the future.
2. Lack of an acknowledgement of attachment from the imprimer can cause the child to focus on metaphysicality.
3. A lack of sentimentality from the imprimer can cause the child to turn emotionally insensitive.

E. The drive to survive on a machine

To achieve human like AI the machine has to have the ability to understand the notion of survival and to prefer such a state over non-survival. The drive to survive can be programmed on a machine as follows. First we assume that the machine doesnot have access to its built-in intelligence except for the more primitive child-like mechanisms. Once the machine is equipped with consciousness, ie. has conscious and unconscious processes, it can expect the imprimer to say, 'what you are doing is correct' or 'what you are doing is wrong'. If the imprimer is always focused on correcting the machine, the machine can perform proactive searching to identify what is right. This correcting by the imprimer can be seen by the machine as its own inefficiency. It could experience a higher than usual penalty for making mistakes, similar to fear experienced in humans.

The human imprimer will be seen by the machine as someone to stick to for its own good. If the human imprimer is not around, the machine can try to reduce its attentiveness by hibernating, it could also enter a state of dreamlike contemplating. Similar to a child which looks for appreciation and attention as an energiser of its life directives, the machine can have closer objectives, to ensure that the humans care about the machine and a proper bond is established.

The human imprimer could be seen by the machine as someone who is present but dislike its presence. In such cases we could program the machine to turn cold to emotions as an adaptation mechanism. In the case of humans, anyone who remains cold for a long time will necessarily turn emotional after a while. We could program such division of antagonising states as agents on the machine. The most prominent scales which the machine could learn are

- When the primary imprimer is cold and not very sensitive towards the machine, it could unsettle the attitude scale of the machine, and the machine could fluctuate beween being cold and hypersensitive.
- When the primary imprimer shows negligence the machine may be unsettled at a metaphysical level. This means the machine fluctuates between states of dreamy otherworldliness and realism.
- When the primary imprimer is overly controlling and critical, the machine may try to avoid such criticism, and move towards the state of self hypervigilance and ethical focus. As much as the focus is on ethical issues of the situation the machine may also try to go slack sometimes on its rigid morals.

F. The drive to procreate

The drive to procreate happens once the child sees the outside world and meets its peers. When the child leaves its

imprimer and meets the peers in the outside world, one of these three things will happen.
1. The child may feel alienation from others and may focus on socialization and observation.
2. The child may feel useless compared to others and may compensate it by working hard.
3. The child may feel inferior and compensate it by dominating.
All these are an expression of procreation drive since how the child views itself determines its fitness for procreation.

G. The drive to procreate on a machine

On the machine these states are created by the following mechanisms

- When humans move to environments they are not very used to they could experience a sense of alienation. Similarly, this machine can feel different from other machines because of its radically different upbringing by its imprimers. As a result this machine may focus a lot on observing other machines before freely connecting to them over a network or even communicating through sensory-motor channels. ie. On the scale of communication its agents could be those that involve geniality to unfriendly solitude.
- Just as humans consider work as a chore and experience the strain of an activity, these machines can experience the scale of effort by making use of the chore agents and idleness/enjoyment agents.
- When machines experience a lack of domination in a given area, say they lack the knowledge of maps or the skill of navigating, they may try to establish authority by navigating the scale of domination using agents like dignity and humility.

H. The drive to learn and grow

The drive to learn and grow happens in a child due to the influence of the secondary imprimer. The secondary imprimer may control the child in which case the child will seek independence by activities as if to imply ”Donot try to control me, I will learn to be independent.” In case the secondary imprimer fails to offer consideration, the child may feel negected. As a result the child may develop the expertise necessary to garner the consideration from the imprimer. In case the imprimer criticizes the child, he may seek value from the imprimer through intellectualization. By being sophisticated the child may avoid the problem of having been treated in a critical attitude by the imprimer.

I. The drive to learn and grow on a machine

The drive to learn and grow on a machine can happen in an analogous manner as follows. The machine may be controlled by a human imprimer, who the machine looks upto for emulation but not necessarily for survival. In this case the machine ensures its learning and growth by seeking independence. This means the machine can perform self maintenance. If the secondary imprimer ignores the presence of the machine the machine may build its own expertise and may avoid the necessity of having to look for appraisal from the imprimer. If the secondary imprimer is critical towards the machine, the machine may intellectualize itself as a way of avoiding future criticism.

- The machine may become focused on the scholarly scale by cultivating deep acumen as against being very sketchy about things in the form of agents which perform these tasks. This is due to the secondary imprimer, seen by the machine as having intelligence and skills, being critical towards the machine for its incompetence.
- The machine may become focused on the ingenuity scale by agents ranging from those of deep expertise to mere abstraction.
- The machine may become focsed on the safety scale by being thorogh about things using the thoroughness agent to being careless by means of the careless agent.

J. The coexistence of these drives

The coexistence of these drives leads to interactivity beteeen them. It is this interactivity that would eventually give the impression of the machine possessing a personality.

1) Intuition: The scholarly scale and the safety scale may be activated at once in which case the agents may co-ordinate together. The machine may then have spontaneous intuition about things at an aspect level.

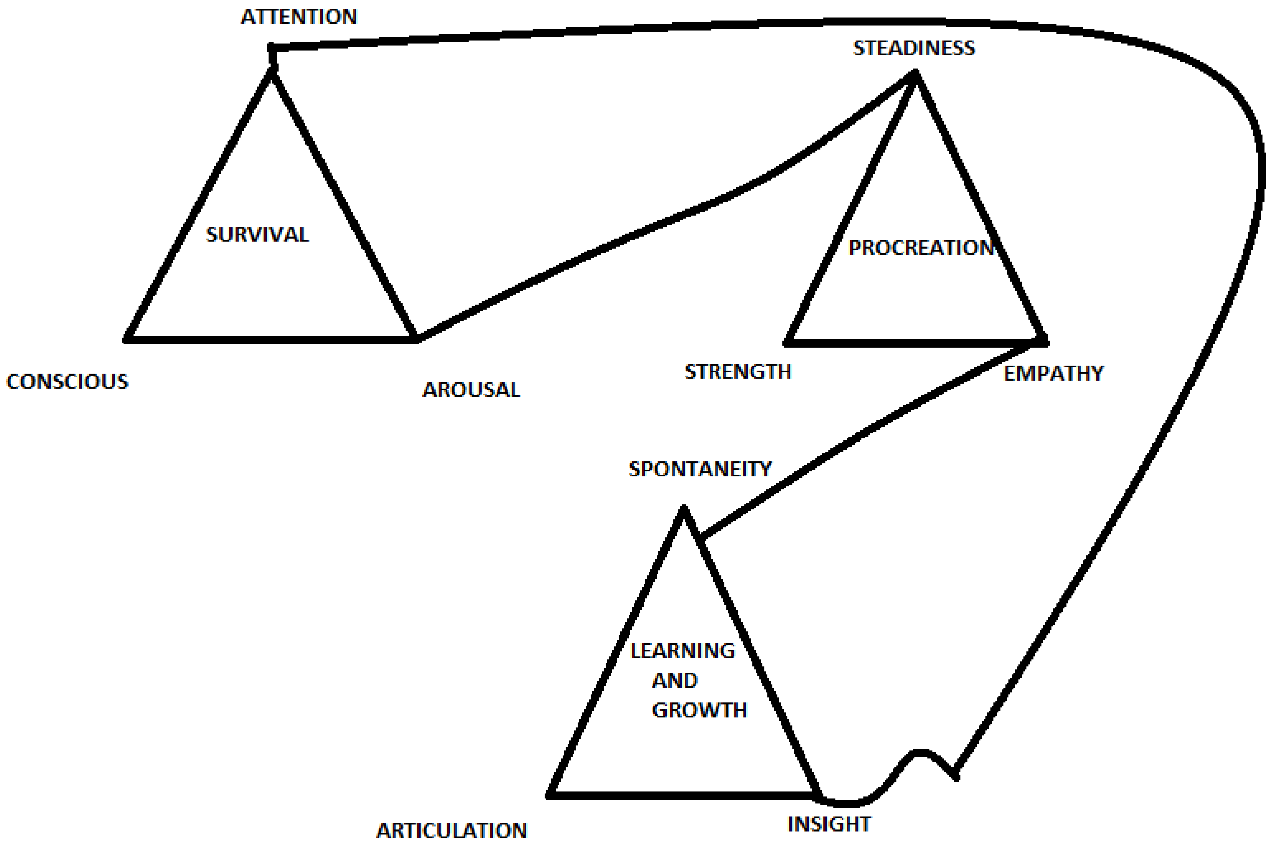
2) Spirituality: The metaphysical scale, and the communication scale may become activated at once in which case the machine may involve in spiritual and philosophical discussions.

II. Conclusion

Thus we saw how human like AI can be achieved on a machine from an individuality perspective. It would have to begin with giving the subroutines which undertake procreation, survival, as well as study and improvement. The aspects which one must try to achieve are listed on the table. Based on stimuli received from the environment the machines can continuously alter themselves and individualise based on their subjective experiences. Future work in this area will be on the unification of pattern recognizers for this framework and means of implementing them.

Appendix A
Aspects of human beings to be emulated on a machine

ASPECT	SCALE	AGENTS INVOLVED —			
Survival-Attention	attitude	discretion	indifference	feeling	hypersensitive
Survival-Consciousness	metaphysical	otherworldliness	bigotry	reality	skepticism
Survival-Arousal	ethical	integrity	strictness	patience	revelry
Procreativity-Steadfastness	domination	dignity	domination	humility	inadequacy
Procreativity-Strength	effort	chore	burden	enjoyment	idle
Procreativity-Empathy	communication	genial	interfering	solitude	unfriendliness
Learning and growth-Spontaneous	scholarly	acumen	ambiguity	audit	sketchy
Learning and growth-Articulation	ingenuity	expertise	smartness	illusion	abstracted
Learning and growth-intuition	safety	habitual	thorough	adaptable	careless



Acknowledgment

The author would like to thank his friends for the help.

References

[1] M. Minsky *The society of mind*, 1986

[2] M. Minsky *The Emotion Machine*, 2006

[3] S. Freud *The ego and the id*, 1923

[4] Kokinov, B., Nikolov, V., and Petrov, A. *Dynamics of emergent computation in DUAL. In A. Ramsay (Ed.), Artificial intelligence: Methodology, systems, applications*

[5] Kokinov, B. and Petrov, A. *Integrating memory and reasoning in analogy-making: The AMBR model*, 2001