Assumptions and Pathway for Achieving Artificial General Intelligence

Ralph C. Ennis
Raleigh, NC USA
ralphennis@gmail.com

Abstract. The purpose of this paper is to posit core assumptions for adult human-level intelligence (HLI) and an assumptive pathway to achieve artificial general intelligence (AGI). These assumptions and pathway can form an evaluative means for constructing decision models and algorithms for AGI.

The core assumptions for HLI revolve around three constructs: 1) the nature of language, 2) sensibility of reason, and 3) ethical reasoning. Implicitly these constructs should be adaptive to cross languages, cultures and subject matter of the human experience.

The assumptive semi-sequential pathway to AGI involves: 1) process inputs, 2) learning, 3) evaluation, 4) imagination, 5) optimization, 6) resolving conflict, 7) solidifying rules of thumb, 8) selective memory recall, 9) sequential time markers, 9) prediction, 11) explanation.

From these core assumptions a mean of describing process (thought) decisions and output (behavior) decisions can be forthcoming. The assumptive pathway can guide formation of decision mapping structures and logical employments within those structures.

Keywords. Assumptions, pathways, artificial general intelligence.

1 Introduction

In 1950 Alan Turing established a test to evaluate the quest for human-level intelligence by machines or artificial general intelligence (AGI). That test involved a blind dialogue between an examiner, a human and a machine. When the examiner can discern no difference within the dialogue between the human, the machine and himself, then human-level artificial intelligence has been achieved. This simple test of “can’t tell the difference” has been a benchmark for achieving AGI. Over the past six plus decades since Turing proposed this test much progress has been made and yet this goal for adult-level human intelligence remains elusive.

Many assumptions are embedded within the Turing evaluative procedure. The foremost assumption involves determining the bare essentials for adult human-level intelligence. After which an assumptive pathway is required to achieve such intelligence before algorithms and code can be written. In most human goal-seek ventures, if the assumptions for solving a problem are both comprehensive and effective, the likelihood of achieving the goal is greater than with less effective or comprehensive assumptions.

The subject of this paper is the core assumptions for human-level intelligence and an assumptive pathway to achieving AGI. And the primal assumption is that if the core assumptions and the assumptive pathway are comprehensive and effective, then the probability of achieving AGI increases.

2 Core Assumptions for Human Level Intelligence

Three core assumptions of human-level intelligence (HLI) will be discussed: 1) the nature of language, 2) sensibility of reason, and 3) ethical reasoning. Stated otherwise, HLI is associated with language that makes sense and inherently involves ethical reasoning. Non-human intelligence may operate without this full set of assumptions. Though a different list could be devised, this set is posited as fundamental to human intelligence and should be included in any discussion of AGI.

2.1 The Nature of Language

The nature of language is complex across cultures, age, demographics and non-verbal expressions. Since Turing's test requires dialogue, for this discussion words will be viewed as the primary component of language for any AGI. Below are five assumptions regarding the nature of language.

Language, navigated through words, is symbolic, spatial-temporal, contextual and requires authorship. A "word" primarily represents, not itself, but something distant, apart from itself. That word is symbolic—it represents some spatial-temporal construct at a concrete level (often referred to as a sign) or a more abstract connotations (symbol).

> The philosopher Wittgenstein (1958) referred to the "spatial and temporal phenomenon of language" (p. 47). This spatial-temporal quality of language allows an author of words to transcend himself—perceive and transmit beyond himself. [Various spatial-temporal vantage points thus accounts for the differentiation between self and other awareness necessary for any discussion of commonly shared (perceive) reality as well as a sensibility for a variety of spiritualities].

A word is also contextual. The same word in two diverse contexts may carry different, though maybe similar, meanings. Furthermore, the diminishing presence from authorship to the receiving party accounts for much confusion in the transmission of meanings in language between humans.

For example, the simple words "I love you" conveys deep meanings to most who hear it or long to hear it. These words convey difference nuance meanings when spoken to a loving spouse as compared to a beloved child. Each word in this sentence communicates a spatial-temporal construct. "I" and "you" are concrete symbols of distinct persons while "love" is an abstract symbolic construct that has found meaning over time (to the author and hearers) through acts of love that resonate sensibility to them. Furthermore, inherent within these words is a sense of the author's presence in time. If "I love you" is spoken and received in the immediate present, the meanings is full and rich to the receiver. If however, it is spoken and received a year later without any direct author presence, the meaning of "I love you" may be quite different. If the author has subsequently abandoned the intended receiver, the receiver will undoubtedly understand these words quite differently.

Thus, if AGI code cannot convincingly convey words originating from itself and commonly understood symbolically, spatial-temporally, and in context, then Turing's test may not be fully satisfied.

Language is embedded within diverse emotional constructs across cultures. Every healthy human being experiences emotions. Nevertheless there is no uniformity of emotional words that apply across all cultures.

For instance, in the Japanese construct of emotionality "*amae*" is a powerful emotion. The Japanese psychiatrist Takeo Doi (1981) unpacks this emotion for Westerners by stating, "The Japanese term *amae* refers, initially, to the feelings that all normal infants at the breast harbor toward the mother—dependence, the desire to be passively loved, the unwillingness to be separated from the warm mother-child circle and cast into a world of objective 'reality' " (p. 7). He went on to say, "… all the many Japanese words dealing with human relations reflect some aspect of the *amae* mentality. This does not mean, of course, that the average man is clearly aware of *amae* …" (p. 33). In the English language there is no direct translation for *amae*.

The complexity of language is displayed in the fact that all language has an emotional component in the originating and receiving for words—and that emotionality must be accounted for in AGI. Sometimes that emotionality may seem to be non-existent but it is better perceived as muted or laying in wait to spring into action—nonetheless, emotions are always present in language. Stated another way, the regions of the brain responsible for emotional processing never go entirely dormant. [Even mathematical symbols must pass through the grid of a students' emotionality as he/she struggles to solve problems.]

Thus, Turing implicitly requires that the examiner notice an appropriate handling of the emotional world between the dialogue of human and machine. For full AGI to be achieved the code must account for a multiplicity of nuanced emotions across cultural contexts.

Human language is bodily encased. Language is experienced and transmitted in and through the body that innately perceives "attractions-aversions." Without logical awareness, a child responses to stimuli in a manner that resembles the reaction of the simplest of life forms to outside influences. Each move toward or away from stimuli it innately perceives as beneficial or threatening. This surviving-to-thriving reaction is often translated into a language of "pleasure" and "pain" or, at higher abstract levels, into "attractions" and "aversions" or "harmony" and "dissonance."

Achieving AGI does not necessarily require code embodied by sensory "flesh." Nevertheless, AGI must account for "pain and pleasure" at a primal level. Without such primal responses, Turing's examiner may eventually perceive a flaw in the machine that doesn't account for HLI encased in bodies that perceive beneficial and threatening stimuli.

Words are best processed for learning as they attach to images and meanings attached through analogy. People process and remember images far better than words (Grady, 1997). The symbols that words reflect are often birthed through images. Images are powerful. They drive much human communication and learning. By comparing images, humans use pattern recognition to associate words, form abstractions, and learn through analogy. Gentner, et al. (2006) states, "The proposal that comparison processes can promote language learning is based on research in analogy and similarity." And Marvin Minksy (2007) believes only through the pathway of analogy will AGI be achieved.

For instance, in the mind of a child, the image of mother's face becomes deeply associated with the word "mommy." In time, the pattern of woman-with-baby recognized in living images (or artistic displays) is generalized to the abstract of "Mother." And by analogy "Mother" can extend to any female animal with offspring or even to Mother Earth as a birther and nourisher of life.

The fluidity of learning required for AGI may best be negotiated through image associations and secondarily through word associations until analogies are formulated that facilitate abstractions. The Turing examiner will look for such learning abilities within the dialogue between human and machine.

Words are not discrete and no exact definition of terms is required as a starting point for AGI—rather a process for dynamic adjustment of words is required. As Jacque Derrida, the late French postmodern philosopher, has stated, "It is at the price of this war of language against itself that the sense and question of its origin will be thinkable … Language preserves the difference that preserves language."

This convolution of words is what the twentieth century philosopher Edwin Wittgenstein referred to as "Language is a labyrinth of paths. You approach from one side and know your way about; you approach the same place from another side and no longer know your way about" (Philosophical Investigations 203). Or one might say, language is not formed by discrete, immovable categories (words) but rather by flows of embraced constructs through continuums of intersecting and interacting pathways.

Fixed definitions of words will insure that AGI will not be achieved. However, a radical relativism approach to language also dooms the quest for AGI. Common sense (semi-ambiguous) meanings inherent in a relational usage of words is a better approach both for transmission of meanings and for learning—and will appease the Turing examiner.

2.2 Sensibility of Reason

In reasoning with a child (or across cultures) is becomes quickly apparent that sensibility is fluid. Or stated another way, two people can arrive at the same or different conclusion by very different pathways—and yet sensibility is achieved for both people.

The discussion on sensibility usually starts with rules of formal logic. Computer code usually starts and stops there. However, human-level intelligence is present long before formal rules are acquired or followed. What makes sense to one three year-old may not make sense to another three year-old and yet both are seeking "sensibility"—trying to make sense of their worlds.

One possibility is that sensibility is the play of dissonance and harmony with the energy frequencies within the brain. Minsky (1981) has suggested a link between music and meanings. Recent work by Lu, J. et al. (2012) has translated brain waves to music; this avenue to sensibility must be explored. If fruitful, we might view the brain as a "music box" continually seeking harmony while resolving even-present dissonance. This play of resonance may account for sensibility and irrationality.

In any case, Turing's examiner would surely ask both human and machine the question, asked with annoying frequencies by most three year-olds, "why" and expect a "sensible" response in the dialogue.

2.3 Ethical Reasoning

In a previous paper (Ennis, 2013) I noted:

> *Mikhail (2007) frames the following poignant question relevant in the pursuit of an ethically-based artificial general intelligence (AGI): "Is there a universal moral grammar and, if so, what are its properties?" Stated otherwise, is there a set of rules that govern the formation of all ethically acceptable behaviors across cultures?*
>
> *Evidence can be found on any kindergarten playground across the global community that ethical reasoning is at play. In what part of the human experience is some construct of "fairness and harmony" non-existent? This construct may seem suspended or violated at various times, but an innate awareness of fairness and harmony resides within us all—even in our early childhood interactions (Smith, et al., 2013).*
>
> *Fairness may be defined differently across individuals, families and cultures, but yet it resonates within all social structures even if pathways to it are blocked. Fairness to some implies non-bias equality of quantity and quality. However, this definition rarely works out well without the consideration of context.*
>
> *For instance, is it fair to an eight year-old sister to be treated equally with her four year-old brother, or vice-a-versa? Most parents would conclude unequal treatment is far more "fair" that an unwavering pursuit of equality. Much to the consternation of young siblings, most parents conclude that it does not have to be equal to be fair. Fairness is contextual to age, abilities, available resources, etc.*

If fairness is not somehow achieved or at least approximated, we humans recognize that harmony (dynamic balance) within a system may be threatened or disrupted. Back to the family system—sibling disputes over fairness can disrupt the sense of harmony for all in the family.

What remains in the pursuit of ethical reasoning is not the question of a set of ethical rules that are proven to be universal, but rather can a grammar—a functional ethical DNA be established? By using that DNA of ethical reasoning, can a diversity of contextual rules be fashioned and situations evaluated for ethical acceptability? Is that DNA applicable in the formation of ethical rules and parsing of existing rules across cultures—even when the rules seem in conflict?

A solution to that ethical DNA (e-DNA) and subsequent management of it is paramount in the quest for artificial general intelligence (AGI) (Gubrud, 1997). This e-DNA should account for the human sense of fairness and harmony across a multitude of contexts. Asimov (1950) proposed such a moral code with his three laws of robotics, but we need a more fundamental code from which these laws and others might be derived. As Pana (2006) states, "We do not have to implement a moral code, but to create a moral intelligence, we can aspire to a condition of potentiality, not the generation of some fixed reality."

The examiner of AGI will quickly perceive the ability of the human to seek fairness and harmony. But will the machine pass this test? The answer is or should be of upmost concern for all in the enterprise of building AGI systems. Without ethical reasoning, AGI may be very intelligent but it will not resemble child or adult human-level intelligence regarding ethical reasoning. Such intelligence may find no difficulty in prescribing and enacting decisions that humanity may find utterly unethical and disastrous.

3 Assumptive Pathways to Artificial General Intelligence

With the above assumptions of language, sensibility and ethics in mind, a pathway to AGI is suggested below.

3.1 Process Sensory Input

All forms of sensory input (visual, auditory, taste, touch, smell) must eventually fit within a model for AGI. However, visual images and written words seem sufficient to begin. These inputs must be received from those inputting data or auto-gathered across data fields. The inputs must then be ignored, discarded or filed in retrievable though adjustable filters.

3.2 Learning through Pattern Recognition

Learning occurs through recognition of patterns. These may be new patterns or recognition of previously established patterns that can reinforce learning.

Learning can occur through analogy. This comparative process allows accelerated pattern recognition—previously recognized patterns are leveraged to identify new patterns.

Learning is dependent upon innate rules of thumb. These rules a baby is born with. Without them all humanity would need to rediscover higher-level rules of thumb for reasoning and communicating. Innate rules of thumb are apparent in children at an early age.

Learning occurs through processing new input. It also occurs through processing feedback from previous decision consequences. These inputs allow for reinterpretation of prior inputs and formation of higher-level rules of thumb—held as solid but adjustable.

The rules of grammar for word order in sentence context are learned rather than assumed or pre-programmed. And in like fashion, formal rules of logic are learned.

Auto-learning can occur through auto-gathering and processing of data. This data is temporarily mapped and adjusted into a more appropriate location by rules of thumb related to the words or images in context.

3.3 Evaluating Decisions for Ethical Acceptability

As much as ethical reasoning has fallen out of favor in our current post-modern rationality, even the construct of tolerance is heavily laden with ethical acceptability. Some means of evaluating process and output decisions for ethical acceptability must be achieved. "Ethical DNA Model for Artificial General Intelligence" posits such a means (Ennis, 2013).

3.4 Imagination of Possibilities

As inputs increase and are linked within varying emotional intensities, imagination (i.e. dreaming) becomes possible. Even as a three-year old child lives in an imaginative play world, so AGI must have an ability to imagine what is not actual. Without imaginative powers, AGI will eventually fail in the eyes of the Turing examiner.

3.5 Optimization of Decisions

Optimization of decisions is a truly human intelligent pursuit. Achieving a goal involves uncertainty. People seek the best result. That best involves both sound reasoning that is congruent with prior rules of thumb and mindsets as well as, in time, a positive evaluation of decision outcomes.

Congruence decisions can be conceptualized as acceptable. Though not always optimal, congruence can serve as a benchmark in pursuit of optimization.

Dissonant decisions are conceptualized with a range from warning to dangerous. Dissonance might possibly lead to a redistribution of rules of thumb.

Comic decisions are conceptualized to facilitate dissonance and redistribution of rules of thumb. [The use of human comedy to create cognitive dissonance is an art form used across cultures to make room for alternate solutions to common problems.]

Paradoxical solutions are conceptualized as optimal. The adult human mind runs headlong into paradoxical reasoning. That is, thoughts A and B are held to be congruent when viewed separately but when viewed together they seem contradictory (dissonant). Often the optimal solution for a system may appear paradoxical. A paradoxical conclusion can be seen as a means of declaring the limits of the human mind to solve a problem that is based within our spatial and temporal limitations. Optimizing decisions in fields ranging from global economics to physics can be viewed through this paradoxical pathway.

Turing's examiner may well pass AGI that resembles child-like thinking because paradox is seldom on a three year-olds mind. However, true adult-level AGI must account for paradoxical optimization—the best solution is sometimes a paradoxical conclusion.

3.6 Resolve Conflict

Every human being experiences the quandary of arguing with others or not agreeing with self. For instance, few people hold all the same rules of thumb for social interaction at age 20 that they held so tightly at age ten. Added to this conflict are conflicts with other mindsets (individuals, nations, etc.). Human intelligence gives considerable energy to resolving conflicts within and between mindsets. Lack of resolution can have mild to disastrous results from individual mental confusion to wars between nations.

Conflict resolution within and between mindsets may possibly be negotiated through the weighted influence of mindsets and a paradoxical central construct of ethical acceptability in order to diminish dissonant conflict. AGI must pass this test as well.

3.7 Solidifying Rules of Thumb

In order to make fast decisions, human being establish rules of thumb rather than sorting through all data inputs to re-logic every decision variation. We all have our rules of thumb for what behavior to employ when it is raining. And we usually defer to those rules of thumb rather that process all available data regarding water composition, rate and velocity of rainfall, etc. before making our clothing choices on a rainy day.

Types of rules of thumb can be conceptualized to include: innate (ethical reasoning DNA), metaphoric (simple comparative rules), situational (simple consequential rules) and abstract (complex comparative and consequential rules). Abstractions, refined through sensible analogies, facilitate formation and adjustment of rules of thumb. And rules of thumb are prioritized and adjustable with additional input.

AGI may best be built by forming rules of thumb from a baby-mind to adult-level intelligence versus dumbing down from adult to child. This fragile process must be overseen and adjusted as AGI accounts for pain-pleasure in the human experience.

Any examiner will perceive the use of rules of thumb. Omission of rules might be detected by the onslaught of data a machine might use to justify their thought and output decisions.

3.8 Selective Memory Recall

Memory storage allows for inputs, patterns, rules of thumb, etc. to be accessible without constantly in focused consciousness. Memory, with its large storage capacity, can be accessed through prompts for recall.

Recall in humans is always limited—commonly referred to as selective memory. Though some human brains have been shown to have total memory across a progression of time, few humans actual possess such total recall ability.

AGI should then be able to recall context and time-appropriate information. A dialogue with an AGI machine will reveal an ability or lack thereof to recall this type of information and then associate it with appropriate rules of thumbs.

3.9 Sequential Time Markers

It is not enough to recognize patterns of objects and logic. Time must be accounted for by AGI. Meanings in words are often time-sensitive. Thus, input must be marked as well as the formation of patterns and rules of thumb.

An inability to discern timed sequences is necessary for adult HLI.

3.10 Prediction of Process and Outcome Decisions

The foundation of predicting decisions is probabilistic cause and effect of imaged decision outcomes adjusted through the feedback of prior decision consequences. Within this rubric, prior decisions are factored in but determinative of future thoughts and events. Thus, parents might predict (with some degree of probability) that their three year-old son will decide to eat all of his vegetable today because he responded so well to negative consequences from last night's traumatic dinner experience at Grandma's house.

AGI will demonstrate some ability to predict. Whether successful or not, the propensity to predict is inherent in human-level intelligence.

3.11 Explanation of Decisions

AGI must have sensible reasons to some discernable degree. Decisions without an articulated rationale is less that human-level intelligence. These explanations can be conceptualized as congruent or dissonant with prior data associations. The play of congruence-dissonance might be negotiated through an approximation of music from brain waves.

4 Conclusion

The implied goal of this paper is that an AGI software program incorporating the above core assumptions and assumptive pathway will achieve human adult-level artificial intelligence after much input has been processed, the ethical grid has been established and many situational and abstract rules of thumb have been formed and refined.

Future research can employ these core assumptions as a mean of describing process (thought) decisions and output (behavior) decisions. And the assumptive pathway can guide formation of decision mapping structures and logical employments within those structures. [A precursor to a mapping model is put forth by Ennis, 2004.] These structures are best formulated to interface with future mapping of the neurons of brain and possibly employing layered memristors as a hardware means for better storage and manipulation of data that is often described on continuums rather than discretely.

Other assumptions and pathways may indeed be needed to fill in a road map to AGI. The assumptions put forth in this paper, I maintain, are essential to passing the Turing examination.

In addition to Turing's test, true HLI must account for irrationality and unethical behavior while hopefully presenting a means of moderating such human tendencies. Within the above assumptions for HLI and the assumptive pathways to AGI both tests are view to be achievable over time.

To that amazing goal, the field of AI continues with an uncertain end regarding success and the desirability of that achievement. May AGI achieve not only adult-level human intelligence but also the ability to perpetually seek paradoxical ethical optimization in ways that support fairness and harmony between human and machine desire for survival and thriving.

References

Derrida, J.: *Speech and phenomena and other essays on Husserl's theory of signs*. Trans. David B. Allison. Evanston, IL: Northwestern University Press, (1973) 14.

Doi, T.: *The anatomy of dependency: The key analysis of Japanese behavior*. Tokyo, Japan: Kodansha International (1981).

Ennis, R.: A theoretical model for research in intercultural decision making. *Intercultural Communication Studies*. 8 (2004) 113-124.

Ennis, R.: Ethical DNA model for artificial general intelligence. *The 10th International Conference on Modeling Decisions for Artificial Intelligence* (2013). Pages 56 -67 in USB Proceedings. ISBN: 978-84-695-9120-8

Gentner, D. and Namy, L.: Analogical process in language learning. *Association for Psychological Science*, 16(6) (2006) 297-301

Grady, C., et al.: Neural correlates of the episodic encoding of pictures and words. In *Proceedings of the National Academy of Sciences of the USA*. 95:2703-2708. (March 1998)

Lu, J. et al.: Scale-free brain-wave music from simultaneously EEG and fMRI recordings. *Plos One*. (November 14, 2012)

Mikhail, J.: Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences*. 11(4) (2007)143-152.

Minsky, M.: Interview in *Discovery Magazine*. (January 2007)

Minsky, M.: Music, Mind and Meaning. *Computer Music Journal*, Vol. 5, No. 3, Fall 1981

Turing, A.: Computing machinery and intelligence. *Mind*. 59:236. (1950)

Tversky, A. and Kahneman P.: Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131. (1974).

Wittgenstein, L.: *Philosophical Investigations*. Trans. G. E. M. Anscombe. Oxford: Blackwell Publishers. (1999)

January 2014