Artificial Intelligence: A Child's Play

Ravi Kashyap

SolBridge International School of Business / City University of Hong Kong

February 18, 2019

Keywords: Artificial Intelligence; Turing Test; Curiosity; Confidence; Uncertainty; Trial and Error

ACM Codes: I.2.0: General; I.2.6: Learning; I.2.8: Problem Solving, Control Methods, and Search

AMS: 68Q32 Artificial intelligence; 68T05 Learning and adaptive systems; 91E40 Memory and learning

Contents

1 Abstract **2**

2 The Benchmark for Brainpower **2**

2.1 Questionably Simple Yet Complex Benchmark 3

2.2 Intelligence for What Sake? 5

2.3 Minds versus Machines 7

2.4 Becoming Smarter than Albert Einstein! 8

3 A Journey to the Land of Unintended Consequences **13**

3.1 I Don't Know, A Great Answer 13

3.2 Acing the Turing Test 15

3.3 Imitation in the Imitation Game 16

3.4 Mexican Chihuahua solving Korean Puzzles under a Mush-Room 17

3.5 Merry-Go-Round of Trials, Errors and Revisions 18

3.6 Gifts from the Realm of the Unintentional 19

3.7 Evolutionary Tricks for the Empirics 21

4 A Road-map for Intelligence **21**

5 Conclusions and Possibilities for Future Research **25**

6 Acknowledgments and End-notes **28**

7 References **33**

8 Appendix: From Words to Symbols, A Curious and Confident Model of Intellect 46

8.1 Notation and Terminology for Key Results 46

8.2 Bass Model of Diffusion for Information Accumulation 47

8.3 Knowledge Store 47

8.4 Bhattacharyya Distance for Information Comparison 48

8.5 Dimension Reduction before Information Comparison 49

1 Abstract

We discuss the objectives of any endeavor in creating artificial intelligence, AI, and provide a possible alternative. Intelligence might be an unintended consequence of curiosity left to roam free, best exemplified by a frolicking infant. This suggests that our attempts at AI could have been misguided, what we actually need to strive for can be termed artificial curiosity, AC, and intelligence happens as a consequence of those efforts. For this unintentional yet welcome aftereffect to set in a foundational list of guiding principles needs to be present. We discuss what these essential doctrines might be and why their establishment is required to form connections, possibly growing, between a knowledge store that has been built up and new pieces of information that curiosity will bring back. As more findings are acquired and more bonds are fermented, we need a way to, periodically, reduce the amount of data; in the sense, it is important to capture the critical characteristics of what has been accumulated or produce a summary of what has been gathered. We start with the intuition for this line of reasoning and formalize it with a series of models (and iterative improvements) that will be necessary to make the incubation of intelligence a reality. Our discussion provides conceptual modifications to the Turing Test and to Searle's Chinese room argument. We discuss the future implications for society as AI becomes an integral part of life.

2 The Benchmark for Brainpower

As a first step, we recognize that one possible categorization of different fields can be done by the set of questions a particular field attempts to answer. Since we are the creators of different disciplines, but not the creators of the world in which these fields need to operate, the answers to the questions posed by any domain can come from anywhere or from phenomenon studied under a combination of many other disciplines. Hence, the answers to the questions posed under the realm of artificial intelligence (AI) can come from seemingly diverse subjects, such as, physics, biology, mathematics, chemistry, marketing, economics, theater, music and so on. This suggests that we might be better off, identifying ourselves with problems and solutions, which tacitly confers upon us the title, Problem Solvers, instead of calling ourselves physicists, biologists,

psychologists, marketing experts, economists and so on.

It would not be entirely incorrect to state that the majority of the attempts at solving problems start with posing well defined questions and finding corresponding answers. As we linger on the topic of Questions & Answers, Q&A, we need to be cognizant that any answer we wish to seek would depend on some Definitions and Assumptions, D&A. But if we change those D&A we might get different Q&A, even telling us that Q&A and D&A might be in our very DNA, the biological one, which are always changing (Alberts, Johnson, Lewis, Raff, Roberts & Walter 2002; End-notes 4; 5).

The problem of designing intelligence artificially can be a rather trivial task, depending on which organism's brainpower acts as our gold standard. For simplicity and to be specific, we could confine the creation of intelligence outside the confines of biological organisms; but it will become clear later on that many parts of our discussion apply to the goals of increasing intelligence within biological organisms as well; not to mention a suitable definition, as discussed above, could render biological and non-biological organisms under the same category of sentient beings. We discuss the objectives of any endeavor in creating artificial intelligence (sections 2.1, 2.2, 2.3), AI, and provide a possible alternative. Intelligence might be an unintended consequence of curiosity left to roam free, best exemplified by a frolicking infant (section 2.4). This suggests that our attempts at AI could have been misguided, what we actually need to strive for can be termed artificial curiosity, AC, and intelligence happens as a consequence of those efforts.

For this unintentional yet welcome aftereffect to set in a foundational list of guiding principles needs to be present (section 3). We discuss what these essential doctrines might be and why their establishment is required to form connections, possibly growing, between a knowledge store that has been built up and new pieces of information that curiosity will bring back. As more findings are acquired and more bonds are fermented, we need a way to periodically reduce the amount of data; in the sense, it is important to capture the critical characteristics of what has been accumulated or produce a summary of what has been gathered. We start with the intuition for this line of reasoning and formalize it with a series of models (and iterative improvements) that will be necessary to make the incubation of intelligence a reality (section 4 provides a roadmap for creating intelligence with the technical parts relegated to the Appendix so that it can be comprehended by almost anyone interested in AI). Our discussion provides conceptual modifications to the Turing Test and to Searle's Chinese room argument (sections 3.2, 3.3, 3.4). We discuss the future implications for society as AI becomes an integral part of life.

2.1 Questionably Simple Yet Complex Benchmark

As we embark on the journey to apply the knowledge from other fields to AI, we need to be aware that artificial intelligence is "Simply Too Complex", since through time, AI has just been about beating a benchmark. The complications are mainly to select the right standards to compete with. This problem is compounded due to the fact that nobody really knows what is intelligence, especially when considering arti-

ficial systems: (Legg & Hutter 2007) take a number of well known informal definitions of human intelligence and extract their essential features, which are then mathematically formalized to produce a general measure of intelligence.

To facilitate a reference point for the rest of the article, we define intelligence as below (and perhaps adding to the confusion; section 3.1 clarifies why this might still be a positive outcome).

Definition 1. *Intelligence is the ability to connect elements of previously attained information to effect a decision. Nothing lasts forever and hence, no decision is good forever; but the longer a decision serves its purpose, the greater the intelligence of the agent making that decision.*

With this definition, it is implied that even a very intelligent decision (such as in a game of chess) that falls short of meeting its objective, since it has to counter a decision with greater intelligence, fails the benchmark. Intelligence is with respect to the situation and its demands, highlighting that we can only make relative comparisons of intelligence. It should also become clear that intelligence requires the ability to collect pieces of information and to connect them towards a decision making goal. Since decisions are not going to be valid indefinitely (this could be stated explicitly as an assumption), we need to continue to use these abilities. We further restrict our discussion to the sub class of living organisms, termed the homo sapiens and the agents being created by them using computers and related software to possess intelligence. We specifically check how this definition can be applied in the context of the Turing Test in section 3.2.

With no disrespect to any adults, it would not be entirely wrong to label children as better and faster learners than adults. (Holt 2017) shows that in most situations our minds work best when we use them in a certain way, and that young children tend to learn better than grownups (and better than they themselves will when they are older) because they use their minds in a special way, which is a style of learning that fits their present condition.

(Russell & Norvig 1995) is a comprehensive discussion of the concept of an intelligent agent. (Wooldridge & Jennings 1995) discuss the most important theoretical and practical issues associated with the design and construction of intelligent agents. They divide these issues into three areas (clearly any division cannot completely rule out overlap between the components).

1. Agent theory is concerned with the question of what an agent is, and the use of mathematical formalisms for representing and reasoning about the properties of agents.

2. Agent architectures is about the software engineering models of agents; this area is primarily concerned with the problem of designing software or hardware systems that will satisfy the properties specified

by agent theorists.

3. Agent languages are software systems for programming and experimenting with agents; these languages may embody principles proposed by theorists.

Neural networks are one approach to artificial intelligence (AI) that are modeled on the brain (Haykin 1998; Haykin 2009; Castelvecchi 2016). These systems, loosely inspired by the densely interconnected neurons of the brain, mimic human learning by changing the strength of simulated neural connections on the basis of experience. Unfortunately, such networks are also as opaque as the brain, though they promised to be better than standard algorithms at dealing with complex real-world situations. Instead of storing what they have learned in a neat block of digital memory, they diffuse the information in a way that is exceedingly difficult to decipher.

Deep learning is a three-decade-old technique in which massive amounts of data and processing power help computers to crack messy problems that humans solve almost intuitively, from recognizing faces to understanding language (Jones 2014; LeCun, Bengio & Hinton 2015); Such methods fall under a broader category termed machine learning, which aims to program computers to use example data or past experience to solve a given problem (Alpaydin 2014); using the data to decipher pattern is also known as training the system. Deep learning models are built as artificial neural networks and use a cascade of multiple layers of nonlinear processing units, allowing computational models to learn representations of data with multiple levels of abstraction. Each successive layer uses the output from the previous layer as input. (Deng & Yu 2014) is overview of deep learning methodologies and their applications to a variety of information processing tasks. (Schmidhuber 2015) is a comprehensive survey about deep learning in neural networks.

2.2 Intelligence for What Sake?

To be precise, this is not about creating intelligence to barter for the Japanese drink (though that seems like a wise exchange and might have been done many times before).

A central aspect of our lives is uncertainty and our struggle to overcome it. Over the years, it seems that we have found ways to understand the uncertainty in the natural world by postulating numerous physical laws. The majority of the predictions in the physical world hold under a fairly robust set of circumstances and cannot be influenced by the person making the observation, and they stay unaffected if more people become aware of such a possibility. In the social sciences, the situation is exactly the contrary. (Popper 2002) gave a critique and warned of the dangers of historical prediction in social systems.

We need intelligent decision making because of the uncertainty in the world we live in. Hence perhaps,

the one central theme in this entire article is Uncertainty. The dynamic nature of the social sciences, where changes can be observed and decisions can be taken by participants to influence the system, means that along with better models and predictive technologies, predictions need to be continuously revised; and yet unintended consequences set in (Kashyap 2016); and as long as participants are free to observe the results and modify their actions, this effect will persist.

A hall mark of the social sciences is the lack of objectivity. Here we assert that objectivity is with respect to comparisons done by different participants and that a comparison is a precursor to a decision.

Assumption 1. *Despite the several advances in the social sciences, we have yet to discover an objective measuring stick for comparison, a so called, True Comparison Theory, which can be an aid for arriving at objective decisions.*

For our present purposes, the lack of such an objective measure means that the difference in comparisons, as assessed by different participants, can effect different decisions under the same set of circumstances. Hence, despite all the uncertainty in the social sciences, the one thing we can be almost certain about is the subjectivity in all decision making. This lack of an objective measure for comparisons, makes people react at varying degrees and at varying speeds, as they make their subjective decisions. A decision gives rise to an action and subjectivity in the comparison means differing decisions and hence unpredictable actions. This inability to make consistent predictions in the social sciences explains the growing trend towards comprehending better and deciphering the decision process and the subsequent actions, by collecting more information across the entire cycle of comparisons, decisions and actions.

Another feature of the social sciences is that the actions of participants affects the state of the system, effecting a state transfer which perpetuates another merry-go-round of comparisons, decisions and actions from the participants involved. This means, more the participants, more the changes to the system, more the actions and more the information that is generated to be gathered. Hence perhaps, an unintended consequence of the recent developments in technology has been to increase the complexity in our lives in many ways.

(Simon 1962) points out that any attempt to seek properties common to many sorts of complex systems (physical, biological or social), would lead to a theory of hierarchy since a large proportion of complex systems observed in nature exhibit hierarchic structure; that a complex system is composed of subsystems that, in turn, have their own subsystems, and so on.

This might hold a clue to the marvel that our minds perform; abstracting away from the dots that make up a picture, to fully visualizing the image, that seems far removed from the pieces that give form and meaning to it. To help us gain a better understanding of the relationships between different elements of information,

as we will see in section 8, we use a metric built from smaller parts, but gives optimal benefits when seen from a higher level. Contrary to what conventional big picture conversations suggest, as the spectator steps back and the distance from the picture increases, the image becomes smaller yet clearer.

(McManus and Hastings 2005) clarify the wide range of uncertainties that affect complex engineering systems and present a framework to understand the risks (and opportunities) they create and the strategies system designers can use to mitigate or take advantage of them. (Lawson 1985) argues that the Keynesian view on uncertainty (that it is generally impossible, even in probabilistic terms, to evaluate the future outcomes of all possible current actions; Keynes 1937; 1971; 1973), far from being innocuous or destructive of economic analysis in general, can give rise to research programs incorporating, amongst other things, a view of rational behavior under uncertainty, which could be potentially fruitful.

These viewpoints hold many lessons for AI designers and could be instructive for researchers looking to create methods to compare and build complex systems, keeping in mind the caveats of dynamic social systems.

2.3 Minds versus Machines

We currently lack a proper understanding of how, in some instances, our brains (or minds; and right now, it seems, we don't know the difference!) make the leap of learning from information to knowledge to wisdom (See Mill 1829; Mazur 2015 for more about learning and behavior). A simple criterion for the problem of creating artificial intelligence would make it a child's play. Perhaps, the real challenge is to replicate the curiosity and learning an infant displays. Intellect might be a byproduct of Inquisitiveness, demonstrating another instance of an unintended yet welcome consequence (Kashyap 2016). If ignorance is bliss, intrusion might just be the opposite and bring misery. As the saying goes, Curiosity Terminated the Cat and ... (The movie Terminator should tell us about other unintended consequences that might pop up in the AI adventure: Cameron & Wisher 1991).

This brings up the question of Art and Science in the creation of AI (and everything else in life?); which are more related than we probably realize, "Art is Science that we don't know about; Science is Art restricted to a set of symbols governed by a growing number of rules". This frame of mind and approach to seeking knowledge could be termed, Science without Borders, but combined with the Arts. While the similarities between art and science, should give us hope; we need to face the realities of the situation. Right now, arguably, in most cases, we (including computers and intelligent machines?) can barely make the jump from the information to the knowledge stage; even with the use of cutting / (bleeding?) edge technology and tools. This exemplifies three things:

1. We are still in the information age. As a route to establishing this, consider this: Information is Hidden; Knowledge is Exchanged or Bartered; Wisdom is Dispersed. Surely we are still in the Information Age since a disproportionate amount of our actions are geared towards accumulating unique data-sets for the sole benefits of the accumulators.

2. Automating the movement to a higher level of learning, which is necessary for dealing with certain doses of uncertainty, is still far away.

3. Some of us missed the memo that the best of humanity are actually robots in disguise, living amongst us.

Hence, it is not Mind versus Machine (Not even, Man versus Machine or MAN vs MAC, in short; for the underlying concepts on which modern computers are built and what the future holds see: Davis 2011; Perrier, Sipper & Zahnd 1996; Denning 2005; Amir, Ben-Ishay,Levner, Ittah,Abu-Horowitz & Bachelet 2014; Thompson, Gokler, Lloyd & Shor 2016; End-notes 6, 8). Not even MAN and MAC against the MPC (Microsoft Personal Computer: Freiberger & Swaine 1999; Garland 1977; Campbell-Kelly 2001; Manes & Andrews 1993; Carlton & Annotations-Kawasaki 1997; Wonglimpiyarat 2012; Corcoran, Coughlin & Wozniak 2016; End-notes 9, 10, 11)? It is MAN, MAC and the MPC against increasing complexity! (Also in scope are other computing platforms from the past, present and the future: Williams 1997; Ifrah, Harding, Bellos and Wood 2000; Leuenberger & Loss 2001; Ceruzzi 2003; Armbrust, Fox, Griffith, Joseph, Katz, Konwinski & Zaharia 2010; Zhang, Cheng & Boutaba 2010; End-notes 12, 13, 14, 15). This increasing complexity and information explosion is perhaps due to the increasing number of complex actions perpetrated by the actors that comprise the social system. The human mind will be obsolete if machines can fully manage society and we might have bigger problems on our hands than who is taking care of things. We need, and will continue to need, massive computing power and all the intelligence we can create, to mostly separate the signal from the noise. In this age of (Too Much) Information, it is imperative for Man and Machine to work together to uncover nuggets of knowledge from buckets of nonsense.

2.4 Becoming Smarter than Albert Einstein!

If our goal is to create artificial intelligence, (or anything else), we should aim for the sky, in the hope that we might at-least end up reaching the treetops. This takes us to the central assumption of this paper, which then becomes the ultimate benchmark to beat for any intelligent system.

Assumption 2. *Albert Einstein is the most intelligent human being that has ever lived. It has been remarked, albeit anecdotally, that his Intelligence Quotient (IQ) was between 160 to 190, give or take a few points. For simplicity, and perhaps also because Albert Einstein is more well know than other super smart individuals, we overlook the fact that other people have recorded higher levels of IQ.*

We completely stay away from the debate about the limitations of using IQ as an indicator of intelligence (Weinberg 1989; Bartholomew 2004 describe the status of controversies regarding the definition of intelligence, whether intelligence exists and, if it does, whether it can be measured, and the relative roles of genes versus environments in the development of individual differences in intelligence) since it will not make a conceptual difference for our discussion. (Ceci & Liker 1986) suggest that IQ is unrelated to skilled performance at the racetrack and to real-world forms of cognitive complexity that would appear to conform to some of those that scientists regard as the hallmarks of intelligent behavior; (DeDonno 2016) find that IQ fails to predict certain aspects of learning of Hold'em poker, a game of skill with significant complexity attributes resembling real-life activities such as stock market investing and shopping for a home; for the importance not only of conventional analytical intelligence but also skills such as common sense, creativity, knowledge that is usually not expressed or taught, and wisdom (not captured or hard to measure using presently known standardized tests), needed for real world problem solving, see: (Okuda, Runco & Berger 1991; Wagner & Sternberg 1985; Sternberg 2018).

A few other interesting viewpoints (including intelligence in man-made systems, which includes the possibility that our world was created by some of us from the future or even the past, after we have evolved to transcend time) are: (Hernández-Orallo & Dowe 2010) discuss the idea of a universal anytime intelligence test, that is a test that should be able to measure the intelligence of any biological or artificial system that exists at this time or in the future; (Martínez-Plumed, Ferri, Hernández-Orallo & Ramírez-Quintana 2017) warn about the need to be careful when applying human test problems to assess the abilities and cognitive development of robots and other artificial cognitive systems; (Hernández-Orallo, Martínez-Plumed, Schmid, Siebers & Dowe 2016) contend that there is poor understanding about what intelligence tests measure in machines and whether they are useful to evaluate AI systems, concluding that AI is still lacking general techniques to deal with a variety of problems at the same time, though a more careful understanding of what intelligence tests offer for AI may help build new bridges between psycho-metrics, cognitive science, and AI; though we believe that casting a wider net across all artificial disciplines is necessary, as discussed in section 2.

If we then make the simplifying assumption that if we are somehow able to capture all other attributes and higher dimensions of intelligence, (or desirable features for solving problems), that help to deal with the uncertainty in our lives into a single numeric score (and call it IQ, which could be Intelligence Quotient

or Infinite-Intellect Quest or Imagination Quotient, for lack of a better term), we could state that: we live in a world that requires around 2000 IQ points to consistently make correct decisions (End-note 19). But the problem is that the best of us, by assumption 2 above, has less than 200 IQ points. Hence, perhaps, we need someone like IQ-Man who is friends with Super-Man (for society's fascination with superheroes or super-humans see: Eco and Chilton 1972; Reynolds 1992; Fingeroth 2004; Haslem, Ndalianis and Mackie 2007; Coogan 2009). But since these supreme beings are nowhere to be found (Super-Man at-least can be seen in movies; IQ-Man is truly as of now, nowhere to be found), the rest of us could use the clues mentioned below (both for dealing with our problems and to create intelligence in machines) towards a possible solution to the question: Can we become smarter than Albert Einstein?

1. The Miraculous Circle of Trial and Error

   (a) With each try and subsequent failure, we learn a way to improve and move closer to success. Each improvement brings a better way to accomplish something, or in a way enhanced IQ. But success lasts only till it will fail, and we need to try something else and start all over again (Ismail 2014; End-note 19).

2. Lessons from other Relevant Episodes in History

   (a) The errors need not all be due to our efforts. We can learn from instances, where similar things have been tried and see what we can glean from the mistakes of others (For excellent introductions on the lessons history holds, see: Durant 1968; Malomo, Idowu & Osuagwu 2006).

3. Team Work

   (a) If a team of agents has the common purpose of accomplishing something, the effect is increased intelligence, as long as, no one is looking to sabotage the efforts of others. This is also known as the wisdom of the crowd (Giles 2005). What one person might overlook, another might notice with the overall effect, that the betterment of everyone involved, is accomplished.

4. Insatiable Curiosity and the Desire to Learn

(a) Any agent, that continues to be overwhelmingly curious (Reio Jr, Petrosko, Wiswell & Thongsukmag 2006 discuss the conceptualization and measurement of curiosity; also see: Loewy 1998; Loewenstein 1994; Berlyne 1954; 1966; Litman & Spielberger 2003), which will lead to collecting new pieces of information, might continue to have an uptick in the overall intelligence. This suggests that our attempts at AI could have been misguided, what we actually need to strive for can be termed artificial curiosity and intelligence happens as a consequence of those efforts. But this requires certain basic things to be established, which we will discuss informally in section 3 and in our formal model in section 8.

(Gopnik, Meltzoff & Kuhl 1999) argue that evolution designed us to both teach and learn, indicating that nurture is our nature, and the drive to learn is our most important instinct (perhaps as important as, or, even more important, than our instinct to survive). They reason that even very young children, as well as adults, use some of the same methods that scientists use to conduct research, and to learn about the world. (Campbell 1956) notes a formal parallel between some of the characteristics of organic evolution and trial and error learning.

Any discussion of children and grownups is incomplete without making explicit when does childhood end? Here, we are not asking what is childhood, since that is perhaps harder to define, but it would be a safe assumption that most humans have somewhat of a childhood, however brief that might be.

Definition 2. *The end of childhood is when curiosity and confidence are overtaken by the other concerns that life brings.*

New data show that infants use computational strategies to detect the statistical and prosodic patterns in language input, and that this leads to the discovery of phonemes and words (Kuhl 2004). (Oja 1982) derives a new class of unconstrained learning rules using a simple linear neuron model and shows that the model neuron tends to extract the principal component from a stationary input vector sequence (see: McCulloch & Pitts 1943; Nass & Cooper 1975; Takeuchi & Amari 1979; for more about models on neuron activity and the many roles that have been assigned to individual neurons from computational machines to analog signal processors).

Language and learning is most likely to be a two way street, in the sense that the rules by which infants perceive information, the ways in which they learn words, the social contexts in which language is communicated and the need to remember the learned entities for a long time, probably influenced the evolution of language (Kuhl 2004).

(Bush & Mosteller 1955; 2006) present a mathematical model for simple learning. Changes in the probability of occurrence of a response in a small time are described with the aid of mathematical operators. The parameters which appear in the operator equations are related to experimental variables such as the amount of reward and work. (LeBlanc & Weber-Russell 1996) present a computer simulation designed to capture the working memory demands required in the comprehension of arithmetic word problems, based on the belief that understanding arithmetic word problems involves a complex interaction of text comprehension and mathematical processes.

Criterion 1. *To learn anything, any agent first needs to learn a medium through which the learning can occur. Simply put, to start to learn, we first need to learn a language.*

(Lenneberg 1967) hypothesized that language could be acquired only within a critical period, extending from early infancy until puberty: "the coming of language occurs at about the same age in every healthy child throughout the world, strongly supporting the concept that genetically determined processes of maturation, rather than environmental influences, underlie capacity for speech and verbal understanding".

(Johnson & Newport 1989) try to check whether it should be the case that young children are better second language learners than adults and should consequently reach higher levels of final proficiency in the second language. By testing the English proficiency attained by 46 native Korean or Chinese speakers, who had arrived in the United States between the ages of 3 and 39, and who had lived in the United States between 3 and 26 years by the time of testing, the conclusion that a critical period for language acquisition extends its effects, to second language acquisition, was supported. (Newport 1990) considers evidence from several studies of both first and second language acquisition suggesting that normal language learning occurs only when exposure to the language begins early in life.

(Sutton & Barto 1998) provide an excellent introduction to understand intuitively the ideas of reinforcement learning and the general connection between its parts. They define reinforcement learning as learning what to do, how to map situations to actions, so as to maximize a numerical reward signal.

It is interesting to note, that there is contrasting evidence. (Snow & Hoefnagel-Höhle 1978) test the hypothesis, that second language acquisition will be relatively fast, successful, and qualitatively similar to first language only if it occurs before the age of puberty, by studying the naturalistic acquisition of Dutch by English speakers of different ages. It was found that the subjects in the age groups 12-15 and adults made the fastest progress during the first few months of learning Dutch and that at the end of the first year, (the subjects were tested 3 times during their first year in Holland to assess several aspects of their second language ability), the 8-10 and 12-15-year-olds had achieved the best control of Dutch. The 3-5 year-olds scored lowest on all the tests employed. These data do not support the critical period hypothesis for language

acquisition.

A point we need to keep in mind is that perhaps, English language (and many languages), especially its pronunciations and grammar are not the easiest to learn, due to the many nuances it has, that do not generalize easily. It would then make sense to develop a language that is more structured and free of ambiguity (Stageberg 1968 has a discussion with some examples).

3 A Journey to the Land of Unintended Consequences

A glimpse of what a journey towards the land of unintended consequences holds can be seen by reminding ourselves that all knowledge creation is but an unintended consequence. We start with an attempt to understand the papers written by others, (literature review of knowledge already created or experiments performed under what conditions) and end up with papers of our own (results that add what is missing or suggest improvements). Although to be precise, as researchers, we do want to intentionally create new knowledge, but the exact new knowledge we end up creating is unintentional; we stumble upon it as we wander around the knowledge that is already in place. This is simply because, our intentions can only cater to what we already know, or, to existing knowledge; new knowledge, which is unknown, has to come from the realm of the unintentional.

3.1 I Don't Know, A Great Answer

(Taleb 2007) in his landmark book, the Black Swan, talks about the unread books in the personal library of legendary Italian writer, Umberto Eco, and how over time, this unread collection gets larger. Hence, it would not be incorrect, to say that, there is more that, we don't know, than, what we know; the more we know, the more, there will be to know. But that should not stop us and the agent, from trying to seek the answers, or, even from making a guess, as a starting point.

Hence, an answer admitting, "I Don't Know" is a great answer is most situations. When we design any system or model, especially in AI, questions and answers are important, since that is the primary way to assess the presence of intelligence. But what becomes more important are our definitions and assumptions. To supplement our definition of intelligence, we provide the following cardinal assumption,

Assumption 3. *The knowledge that has been accumulated over time is lesser than the knowledge that is yet to gathered. With this assumption, an answer of "I Don't Know" becomes, not just a correct answer, but it is an invitation to the person asking the question to teach the agent how to answer the question.*

So the agent is always learning, and the reason is, simply, due to what we discussed before: we don't

know most things and hence, the learning, usually never stops. If the person asking the question is not satisfied with the answer, he or she now has a responsibility to teach the agent, to improve upon the answer produced. A failure to create intelligence in any agent is a failure on the part of the teacher in finding a teaching methodology appropriate for the agent. This also implies that:

Criterion 2. *Creating intelligence is not only about writing software code, it is about having the best teachers that humanity has produced, being available to teach the later generations, be it human or machines.*

We now consider the fundamental question of whether we need complicated models or merely stronger beliefs. We state this as our essential doctrine.

Criterion 3. *The intelligent agent has to believe that it has the ability to learn, and the confidence to request lessons regarding answers, that it is unable to generate satisfactorily.*

Combining confidence with the great answer, which follows from our assumption 3, we get a better answer, which is "let me try".

When an agent is not learning, it should ideally be teaching (other agents or anyone else). Perhaps because, teaching and learning are highly interconnected and the best way to learn is to teach. A realization that the roles of students and teachers are constantly getting interchanged, originates from a belief that everyone has something to teach to everyone else. When we are teaching we are also learning from someone else, when we are learning we are really teaching ourselves. To be clear, although, most of us probably know this, learning does not just represent reading textbooks, or, doing assignments, though these are important components of learning. Learning can happen, when we are doing anything, that we enjoy doing. This can be built into the reward system of the agent, so that it accumulates points for aspects that it likes. Different agents could be made to like different things, so that we build a random enjoyment component that learns from different activities.

Hence, if any agent has to learn a lot (or everything really?), instead of trying to find the right teachers, we should make everyone its teacher. Since we have to respect our teachers, the agent now has to respect everyone. A consequence of everyone becoming a teacher, and since the roles of teacher and student can interchange, is that everyone also becomes everyone else's student. And the result might be that everyone will respect one another. Isn't that one of the objectives, and perhaps an unintended consequence, of making everyone intelligent?

Efforts at learning and teaching, usually end up confronting two monsters: Confusion and Frustration,

both of which, though, scary and ugly to begin with, can be powerful motivators, as long as, we don't let them bother us. Confusion is the beginning of Understanding. Necessity, is the mother of all creation / innovation / invention, but the often forgotten father, is Frustration. What we learn from the story of, Beauty and the Beast, is that, we need to love the beasts to find beauty. Hence, if we start to love these monsters (Confusion and Frustration), we can unlock their awesomeness and find truly stunning solutions.

Hence, our agent has to remain confident and ask questions, when it does not have an answer. This can also be stated as,

Solution 1. *Life for an intelligent agent is all about having confidence and the right teachers and /or students.*

3.2 Acing the Turing Test

(Moor 1976) puts forth the argument that the real value of the imitation game (also known as the Turing Test, TT, Turing 1950; End-note 16) lies not in treating it as the basis for an operational definition but in considering it as a potential source of good inductive evidence for the hypothesis that machines think. (French 1990) argues that the very capacity of the TT to probe the deepest, most essential areas of human cognition makes it virtually useless as a real test for intelligence. (French 2000) chronicles the comments and controversy surrounding the first fifty years of the TT, concluding that it will remain importance and relevant to future generations of people living in a world in which the cognitive capacities of machines will be vastly greater than they are now.

(Copeland 2000) suggests, based on unpublished material by Turing, that the Turing test withstands objections that are popularly believed to be fatal. (Harnad 1992) shows that it is important to understand that the TT is not, nor was it intended to be, a trick; how well one can fool someone is not a measure of scientific progress. The TT is an empirical criterion: It sets AI's empirical goal to be to generate human scale performance capacity. This goal will be met when the candidate's performance is totally indistinguishable from a human's. Until then, the TT simply represents what it is that AI must endeavor eventually to accomplish scientifically.

(Saygin, Cicekli & Akman 2000) conclude that the Turing Test has been, and will continue to be, an influential and controversial topic. (Von Ahn, Blum & Langford 2004) discuss the Completely Automated Public Turing Test to Tell Computers and Humans Apart, CAPTCHA, which is a automatically generated test, which most humans can pass, but that current computer programs cannot pass. This is somewhat of a paradox since a CAPTCHA is a program that can generate and grade tests that it itself cannot pass; this finds application in many places on the internet to ensure that computer programs are not substituting for

humans.

An often omitted criteria that needs to be considered when administering the Turing test is the ability, or, the level of skill of the person conducting the test. Surely, different individuals are satisfied with different levels of impersonation. When we see any drama, play or movie that depicts the life of any real person, (while reminding ourselves that movies might not be real, but real life can become movies); different people are satisfied with different levels of acting ability. We all know, that the person playing the role in the theatrical version is not the same as the person that is being enacted. But in many cases, (perhaps, in most cases, when it is well produced), we leave feeling satisfied with the result of the replication. The lesson for us here is this: how far does the test administrator need to go to believe that the computer program perfectly duplicates the human test subject.

Our definition 1 of intelligence implies that the benchmark for intelligence has been surpassed if the question is answered to the satisfaction of the person administering the Turing test. In this context answering a question is the decision making at display.

3.3 Imitation in the Imitation Game

Let us now consider another example of imitation in the imitation game, which was a recently released movie about the role of Alan Turing in the second world war (Proudfoot 2015; You 2015; Guo 2015). The actor in the movie, Benedict Cumberbatch (Porter 2014), does a sensational job portraying the real Alan Turing (though this is a subjective evaluation, if someone disagrees, termed a disbeliever, then it would be fair to state that they now have the responsibility of doing a better role play). To go into length on how Benedict Cumberbatch (or any disbeliever, forced to turn into a better actor) accomplished this would require another paper or a few books of their own (Hagen 1991; 2009 are masterpieces on how to be convincing actors). ***The short answer would be that, an actor believes that he can play the part he is chosen to play, which is what an agent chosen to display intelligence must first be made to believe. This is about not about dishonesty or deception, it is about belief and confidence. As discussed in section 3.1, true confidence comes when we admit we don't know something and we are willing to try.***

The manner in which Benedict Cumberbatch (End-note 18, 3) plays the main character in the movie, Imitation Game, leads us to state the Real Enigma of the Imitation Game as: Which Alan Turing is the More Convincingly Brilliant Mathematician? This question merely inquires as to whether, Alan Turing or Benedict Cumberbatch, would pass a stage test for actors who had to convince the audience they were mathematicians. Anyone that would make the argument that acting like a mathematician does not make a real mathematician, needs to be reminded that acting like a mathematician is the first step to being a mathematician. Once this belief is instilled, time, familiarity with the steps and the notation related to mathematics, supplemented with our road-map for intelligence (section 4) will take care of creating real mathematicians. (Kashyap 2017b) is

an application of our curious and confident approach to creating intelligence in the financial markets.

3.4 Mexican Chihuahua solving Korean Puzzles under a Mush-Room

(Searle 1980) argued, that the fact that machines, can be devised to respond to input, with the same output that a mind would give, does not mean that mind and machine are doing the same thing: for the latter lacks understanding (See End-note 17 for Searle's Chinese-Room thought experiment; Also, Searle 1982; 1990; 2001; 2004 are later discussions; Preston & Bishop 2002 has a collection of essays on this crucial challenge). Searle was in fact against the notion of strong AI, which is that human minds are, in essence, computer programs. That is an appropriately programmed computer with the right inputs and outputs, would thereby have a mind in exactly the same sense human beings have minds. All mental activity, is simply the carrying out of some well-defined sequence of operations, frequently referred to, as an algorithm.

(Penrose 1989) claims that there are aspects of consciousness that cannot be replicated within any computer model, no matter how sophisticated, as long as the model is based on an algorithm. He presents an overview of the present state of physical understanding and tries to show that an important gap exists at the point where, quantum and classical physics meet, as well as to speculate on how the conscious brain might be taking advantage of whatever new physics is needed to fill this gap to achieve its non-algorithmic effects.

Searle's example has had a profound impact on the discussions related to AI for the last many years. However, as a counter argument, let us consider, instead of an American (John Searle) juggling with Chinese characters, he has no clue about, in a closed room, using instructions in English, a language he understands; what if it was a Mexican Chihuahua solving puzzles posed using Korean characters, sitting under a giant Mush-Room (perhaps having devoured the mushroom and hence being influenced by it in ways that we do not quite yet comprehend, but for the purposes of this test, the effects are only beneficial. For the hallucinogenic effects of mushrooms see: Schwartz & Smith 1988; Samorini 1992; Musshoff, Madea & Beike 2000; Halpern 2004), and giving out the right answers back in the form of Korean characters, but only barks in response to everything else.

Does it matter, whether the Chihuahua is only using certain training it has been given, to use rules to arrange Korean characters, or, whether it is the Mushroom causing the miracle or something else? For all practical purposes, the Chihuahua is an intelligent creature, (since it is able to present the right set of Korean characters as a solution to the puzzles or questions we pose), that simply does not speak the same language as we do. We do not understand its barking nor does it understand the voice tones we produce, or maybe it pretends that it does not understand (it can be argued, though we won't continue this line of reasoning, that we understand less of what dogs say, than what dogs understand of what we say; who is more intelligent then?). For simplicity and for rhetorical reasons, let us just say that the effects of the mushroom, last for as long as the Chihuahua is alive, or, until we are still interested in asking it questions using Korean Characters?

To substantiate this counter viewpoint, does it really matter if we are simply using rules

to solve a puzzle or if we are actually understanding how the solution was arrived at, if we completely believe that we understand the solution and merely use rules to arrive at the solution? This is not about being dishonest, or, passing lie detector tests; since,if we believe we know the answer and if we are able to consistently generate the answer, it does not matter how we got the answer; since we should now be deemed intelligent enough as we have come up with the answers.

3.5 Merry-Go-Round of Trials, Errors and Revisions

Usually, on our first attempt to answer any question, we may not get the correct, or, the best answer. This is where, the trial and error part, kicks in. But once, we start somewhere, we learn from our mistakes and improve upon our explanations. In this Question & Answer context, we define any question as a good question and a good answer as something that we only think of later or something we find after a few iterations of trial and error.

(McCarthy & Hayes 1969) is a discussion of the main issues in philosophy that also arise in AI. John MaCarthy, who is credited with coining the term "Artificial Intelligence" defined it as "the science and engineering of making intelligent machines" (McCarthy 2004). (Beck & Arnold 1977) discuss this iterative approach to estimate parameters used in Engineering and the Sciences. Many improvements in the sciences and engineering happen through a series of refinements: (Wolfe 2005) is a discussion of how successive designs of fighter planes, where a failure potentially meant the loss of life of the pilots, brought our incremental improvements, which have now make possible the reality of space exploration; End-note 19 is a mention by Taleb of why it is important to create an environment where the errors are less costly, or, trial with small errors are preferable. Though sometimes, errors are unavoidable as in plane crashes, which subsequently lead to safer air travel for later passengers.

(Swanson 1977) recognizes the essential role of trial and error in accessing to scientific literature, pointing the way toward improved information services illuminating potential inconsistencies that have beset many retrieval exercises (This has strong implications for our knowledge store discussed in section 8.3).

(Doidge 2007) presents classic cases from the frontiers of neuroscience that chronicle the biological changes happening in the brain driven by external impetuses, revealing that adapting to new circumstances and learning to deal with adversity are almost hard wired into us. In essence what they reveal is that the brain constantly changes as situations change, which tells us that what we need to contend with or mimic is a moving target.

(Young 2009) is about trial and learning in a social or economic game theory setting (Gibbons 1992). A person learns by trial and error if he occasionally tries out new strategies, rejecting choices that are, erroneous, in the sense that they do not lead to higher payoffs. In an economic game, however, strategies can become erroneous due to a change of behavior by someone else, triggering a search for new and better strategies. In

economics, it is insightful to establish conditions under which the Nash equilibrium property (Nash 1950) can be established. But in real life, equilibrium is a dynamic, constantly changing state (like a see-saw) due to the subjectivity in all decision making and the differing perceptions of the individuals involved and hence the trial and error never ceases.

Intelligence and learning also involve the ability to guess or the ability to make decisions when the best choice is not exactly clear. Observed data can be consistent with many models, and therefore which model is appropriate, given the data, is uncertain (Ghahramani 2015). Similarly, predictions about future data and the future consequences of actions are uncertain. Probability theory provides a framework for modeling uncertainty. A machine can use such models to make predictions about future data, and take decisions that are rational given these predictions.

In all efforts at creating intelligence, we make an unstated assumption that human beings are capable of intelligence. But, we are not born intelligent. It takes years of nurturing and tutoring for us to become intelligent and we display different abilities and aptitude for different things, or the intelligence of different individual could be in different skills. How could we then have expectations that something, that we deem not to have the capacity for intelligence, has to become intelligent in a relatively short span of time. This holds a strong message for us that, to create intelligence artificially, might require years of training for an agent.

In a typical classroom, some kids end up with more conventional forms of intelligence in comparison to others and as assessed by our benchmark or measure, due to creating more connections and retaining the relevant bits of information they receive. Using our assumption 2, we can reword this as: in a world full of intelligent human beings, only a handful of us become Albert Einsteins. Hence, we would expect a similar sort of situation when trying to create AI, we need to start with a group of agents, with different parameters and let them wander around and see what innate abilities they pick up. Accordingly, we need to further those skills that were naturally acquired. The circle of trial, error and corrections needs to be happening constantly.

3.6 Gifts from the Realm of the Unintentional

(Fogel 2004) chronicles that infantile amnesia, the apparent loss of memory about one's own infancy, has been accepted as fact for at least a few thousand years. (Waldfogel 1948) reveals a serious gap in our knowledge regarding childhood memories, though there seems to be an abundance of clinical evidence regarding the fact that repressed childhood experiences may be significant for adult behavior. The evidence, though plenty, cannot be constructed as proof of the universality or the predominance of infantile amnesia, due to the authenticity of the data used in these studies.

(Nadel & Zola-Morgan 1984) indicate that some memory systems in our body become functional at birth or shortly thereafter, whereas others become active following a period of postnatal neurogenesis. Also, studies

have shown that localized brain damage typically leads to selective, rather than general, memory defects, suggesting that, the postnatal maturation of a specific neural system, lies at the root of infantile amnesia. (Howe & Courage 1993) conclude that infantile amnesia is a chimera of a previously unexplored relationship between the development of a cognitive sense of self and the personalization of event memory. They examine this hypothesis in the context of related developments in language and social cognition. (De Brigard 2014) is a philosophical discussion of the phenomenon of remembering along with a historical perspective and reviews of critical findings in the psychology and the neuroscience of remembering.

Despite the many unknown aspects of infantile amnesia, it is clear that the formative years of any human being are not remembered. Perhaps, an unintended consequence of not knowing who we really were, before we got a better idea of who we were becoming, is to reduce the confusion, as we learn to explore and form a conception of what we are. Or, maybe, evolution only deemed worthy of remembering, only what we remember, which is after we had a better idea of what was happening around us.

Also perhaps, the most important element of AI is to ignite curiosity within the agents, because, once an agent gets inquisitive, learning happens almost by itself, after that. An unintended consequence can be overconfidence and needs to be monitored for closely. At periodic intervals, the agent has to be corrected so that positive learning is rewarded and mistakes are removed. To prevent the abuse of excessive intelligence, perhaps, the teachers who train the agent also need to impart moral behavior and empathy towards the, so called, less intelligent.

Success is, a very relative term. In the extreme case, which, we study a bit about, finance, one person's success (profit) could be someone else's failure (loss). That being said, to triumph in creating AI and almost everything else, it is important to know where we are, and start the journey towards, where we want to be. An unintended consequence of taking the first step on a journey, means that the percentage of the distance left to be traveled reduces from infinity to a finite number. So once we start the trip, it becomes manageable immediately. The subjectivity in how we compare things means that the benchmark for AI will be constantly changing, which means we need our agents to keep on learning, just as we need to do the same as well.

A further glance in the direction of unintended consequences might show that in the process of creating knowledge or intelligence, and trying to understand the world better or make it a better place, we might just end up understanding one another better, perhaps, becoming more tolerant in the process, an unintended yet very welcome consequence; making us wonder whether the the true purpose of all knowledge creation might be to make us more tolerant.

As AI becomes tightly interwoven with many aspects of our daily lives, another unintended consequence would be the many jobs that would no longer any human intervention. While on the surface, this might seem like a grave threat, this trend would force human beings to look inward into what truly makes them human and realize the greater potential of their minds.

3.7 Evolutionary Tricks for the Empirics

Anyone reading this paper, might have certain well founded reservations, regarding the theoretical and conceptual connections outlined here, because of the apparent lack of empirical tests. For that, we would like to point out that the present paper, is based on millions of years of experimentation. This test, which is still going on, and which is still creating intelligence, is nothing but evolution. To replicate this test, might take another few millions years (but perhaps, can be done using machines in a shorter time, though it might still take a few years). This paper also provides the theoretical basis, for many new empirical tests and is based on one (or many?) continuous and complete, empirical experiment(s), happening everywhere in our cosmos.

As epitomized by the magical movie, Forrest Gump, (Groom 1986), Shit Happens, which can also be interpreted as Life Happens (See Palmer 2002; Mandik 2007; Schudson 2007; Campos-Arceiz 2009; for shit happening, including perhaps, both intended and unintended consequences, positive or negative, depending on the viewpoint, in sports, news, philosophy and everything else), which might imply to us, that the most intelligent organism that has been created, thus far, was not really created, but it happened, because of mother nature's trials and errors, also known as evolution (Darwin 1859; Dawkins 1976; Eldredge 2005; Scott 2009 are a tiny collection of works on this phenomenal topic, starting from the pathfinder, to more recent views including debates on other possibilities).

As discussed in section 3.5, this cycle of revisions is also at play in the numerous species that inhabit our planet. A large avenue for future research would be to explore the level of intelligence in different kinds of creatures and the extent to which they display the characteristics described here and perhaps be open to other fundamental principles omitted here (intelligence is also about being open to possibilities). It might also reveal that evolution and its twin sister, reproduction, are passing on genetic improvements geared towards increased intelligence to later generations. Such studies rooted in the basics of biology might even reveal bottlenecks to increasing intelligence based on the physical properties of any system. We just need to keep reminding ourselves, that what happens is usually unintentional, but it just might bring about wonderful consequences, once it happens, even if, some of us might only see the benefits, or, the impediments, at any point in time.

4 A Road-map for Intelligence

While a detailed axiomatic approach to uncertainty, unintended consequences and sapience is postponed for another time (or perhaps another lifetime), the present assortment can be summarized as the below "how to guide for intellect", or, a road-map of the essential elements required to create artificial intelligence (which are elaborated in sections 3.1, 3.2, 3.4, 3.5, 3.6). ***Each step in the following algorithm or pseudo-code can be tested as a separate scientific hypothesis (relevant evidence including pointers to mathematical ingredients from section 8 and further references are given in the corresponding***

point below); but surely, greater the coherence between the components that encapsulate the below concepts, better the intended outcome.

While(Agent is Alive or The World has not Ended)

Begin

1. A language certification is necessary.

   (a) From assumption 1, we need to ensure that the agent can pick up advanced concepts by having been certified that a certain minimum level of language abilities have been acquired. If a certain threshold is not met in terms of language skills, it is back to the language classroom for this agent (Condition 1). (Lightbown, Spada, Ranta & Rand 1999) is a comprehensive discussion on how languages are learned, especially from the point of interest of classroom teachers.

   (b) To be clear, the requirement from the agent can be something simple, like giving advice on a financial strategy. In this case, the inputs can simply be the time series of numbers and the output can be just a Buy, Sell or Hold indication, since all of finance through time has involved only these three simple outcomes. The complications are mostly to get to these three results, which the agent can conjure up in its own way. But its interface with the external world, need not be anything too involved.

   (c) On the other extreme, if we wish to create agents that are mathematicians, we would need a precise representation of mathematical rules and results using explicitly clear notation and terminology. Both the input and output have to adhere to exactly defined and rigorous statements. The language in this case becomes a sequence of steps that flow from the preceding one to the subsequent one using the rules of mathematics or already established results or based on logical arguments. To begin with, we only need to capture basic operations in mathematics and more advanced results will be connected to the basic results using fundamental rules or results already established. As discussed in section 2, breeding true intelligence requires us to transcend artificial domains created by us, such as finance or mathematics; but to make the problem of implementation more manageable, focusing on simplified language requirements specific to any field is prudent.

2. A formal model that collects new pieces of information from the various possible choices

   (a) We model collection of information, using the Bass Model of Diffusion (Bass 1969; Mahajan

1985; Mahajan, Muller & Bass 1991; Bass, Krishnan & Jain 1994; Michalakelis, Varoutas & Sphicopoulos 2010; Jiang & Jain 2012; End-note 20; section 8.2) which is used extensively in marketing to study the adoption of new products. We view a new product being adopted by someone as being equivalent to the agent collecting the adopter. The person that just got newly adopted by the product is the new piece of information from the perspective of our model of information collection. Collecting new pieces of information is how we mimic curiosity in our agents. New models of curiosity would benefit immensely from suggestions on how to develop children to be lifelong learners and to nurture the scientist within all kids (Calkins & Bellino 1997; Ramey-Gassert 1997; Gamble & Cota-Robles 2015; End-note 24). Here, we use illustrate with the two opposite ends of the spectrum mentioned in step 1. The new piece of information could be a new time series of financial data if our requirement is related to finance, based on the expectation that the agent has to give advice on a financial strategy. The new piece of information could be a new mathematical result if our requirement is mathematical, based on the expectation that we wish to create mathematicians.

3. A collection of agents that pick different pieces of information using the collector model

   (a) This is simply a group of agents with different parameters of innovation and adoption for the Bass Model of diffusion (section 8.2). We could also use many developments in the use of models of curiosity (Sato, Takeuchi & Okude 2011 present an experience-based curiosity model which indicates individual's real time curiosity within a city regarding how well the individual knows the city. It aims to understand individual's real time interests by not relying on information the people input intentionally, but by understanding behavior data, done through environmental sensing from mobile devices that are capable of sensing the environment around the individual and not necessarily by interacting with the people directly).

4. A belief instilled in the agent that it/he/she is intelligent

   (a) This has been captured in criterion 3 and condition 1. To elaborate, confidence in one's abilities generally enhances motivation, making it a valuable asset for individuals with imperfect willpower (Bénabou & Tirole 2002). For this the agent has to be an actor, believing that it can play the part of someone who is intelligent. There is a vast literature on the importance of self-confidence

and its relation to performance in different fields such as sports and language acquisition: (Feltz 1988; Clément, Dörnyei, & Noels 1994; Noels, Pon, & Clément 1996;). An overemphasis on global self-esteem is perhaps not ideal and (Owens 1993) discusses the implications for understanding the differential impact of negative and positive self-evaluations on emotional and social well-being.

5. A measure to judge how closely the new information collected matches the information already stored

   (a) To aid in this effort to extract meaning from chaos, we summarize the application of the theoretical results from (Kashyap 2017a) to AI studies. The central concept rests on a novel methodology based on the marriage between the Bhattacharyya distance, a measure of similarity across distributions, and the Johnson Lindenstrauss Lemma, a technique for dimension reduction, providing us with a simple yet powerful tool that allows comparisons between data-sets representing any two distributions, perhaps also becoming, to our limited knowledge, an example of perfect matrimony (sections 8.4, 8.5). This methodology is also necessary to assess how similar newly gathered information is to the knowledge store that we already have (section 8.3).

   (b) A subtle point here is that if we keep receiving, or collecting, the same information multiple times, the degree of similarity of the newly received information with the information we have already collected, condensed (dimension reduced) and stored will increase over subsequent iterations. Hence after a few rounds of certain information being repeatedly received, it will be stored almost in its entirety. This captures the fundamental principle of how learning happens by repetition. The specific implementation models will need to fine tune the number of times something is received before most of it is retained, but at a high level this crucial concept has to happen. This also tells us that the more often some information is being received, the more important it is and the more completely it needs to be stored. The case where we need a lot of precision is for any new mathematical result, which has to be an exact combination of fundamental rules or already established results.

6. A method to keep reducing the information store so that less (and the most essential) data, only almost a summary, needs to be maintained

   (a) Any existing variance reduction (Shlens 2014 is about Principal Component Analysis) or compression technique (Lynch 1985; Storer 1988) can be used or the method from point 5, can be used as well to reduce the data store to a smaller dimension.

7. A regular period of deep sleep.

   (a) The importance of sleep in human beings in not fully understood, but it is beyond debate that sleep is essential and has numerous benefits (Robertson, Pascual-Leone & Press 2004; Ellenbogen 2005; Blischke, Erlacher, Kresin, Brueckner & Malangré 2008; Aly & Moscovitch 2010; Nere, Hashmi, Cirelli & Tononi 2013). Any agent requires a period of sleep, where the steps 5, 6 are carried out without any other external disturbances, so its confidence is not shaken and this belief in itself, or, himself, or, herself, is not destroyed (step 4). This can suggest a hypothesis that, what happens during our sleep might be that our creators (Mother Nature or Evolution or Whatever), might be giving suggestions to us in our sleep and these pointers for a better life, (or greater intelligence, or, whatever their purpose, might be), might manifest themselves as dreams. The steps 5, 6 can be done in the background while the agent is not necessarily asleep, but a period of complete focus on the above two steps might be helpful.

End // The While Loop Ends Here, but it must go on Forever

The formal mathematical elements are discussed in section 8. These quantitative measures can be applied across aggregations of smaller elements that can aid the AI agent by providing simple yet powerful metrics to compare groups of entities and add to its knowledge store. The results draw upon sources from statistics, probability, marketing, economics / finance, communication systems, pattern recognition and information theory; becoming an example of how elements of different fields can be combined to provide answers to the questions raised by a particular field.

5 Conclusions and Possibilities for Future Research

We have discussed the intuition for why we also need the best teachers and not just the best computing science designers to create artificial intelligence. An unintended consequence of establishing curiosity and confidence in an agent, expected to become intelligent, might well be intelligence. We have considered why, even though we wish to create intelligence and make the agent pass tests of intelligence, the gift of intelligence might be something from the realm of the unintentional; We have provided the mathematical tools and formal elements of what such an endeavor might require, which includes models of diffusion, distance measures and dimension reduction, among other things.

1. The possibilities for what improvements are necessary are endless, since we are just beginning. But once we instill confidence in the agent that has to become intelligent, it can ask the questions to learn better answers; that is we try something, observe the mistakes and make corrections depending on the level of progress we deem satisfactory.

2. Another important aspect is to try to establish intelligence in simpler real life applications and then take the lessons to the more complex design of a completely autonomous intelligent creature. There is a lot of activity, in this space, on many individual fronts. We could start with the financial markets, doing rudimentary household tasks, driving etc. (all of which are happening), but combining the trial and errors from all of these experiences are essential toward our greater goal of AI. It would be helpful to start with the existing level of understanding and the latest developments in text parsing, speech recognition, and so on. As we put these parts together, the loop of trying and learning from mistakes has to continue forever (or for a very long time).

3. As the likelihood of having to co-exist with so called artificially created intelligent beings increases, we will need to learn to be tolerant and focus on what truly makes us human and realize the greater potential of our existence.

(Nilsson 2006) argues for the development of general-purpose, educable systems that can learn and be taught to perform any of the thousands of jobs that humans can perform, rather than work toward the goal of automation by building special-purpose systems; but the message we put forth is that the lessons from seemingly trivial tasks need to be weaved towards higher ambitions; (Russell, Dewey & Tegmark 2015) has some examples for futher areas of research in building intelligent agents; (Bottou 2014) suggests that instead of trying to bridge the gap between machine learning systems and sophisticated "all-purpose" inference mechanisms, we can instead algebraically enrich the set of manipulations applicable to training systems, and build reasoning capabilities from the ground up.

(Lake, Ullman, Tenenbaum & Gershman 2017) review progress in cognitive science, suggesting that truly human-like learning and thinking machines will have to reach beyond current engineering trends in both what they learn and how they learn it. Specifically, they argue that these machines should build causal models of the world (including intuitive theories of physics and psychology) that support explanation and understanding, rather than merely solving pattern recognition problems. (Yannakakis & Togelius 2015) give a high-level overview of the field of artificial and computational intelligence in games. (Chesani, Mello & Milano 2017) propose to solve mathematical puzzles by means of computers, starting from text and diagrams describing them, without any human intervention.

That, as of today, the limited success in creating artificial intelligence (in machines, humans and elsewhere)

is due to fundamental limitations with the current thinking and the absence of certain basic principles in the majority of attempts in this space right now. There are numerous journals, articles and scientific efforts aimed at creating artificial intelligence. The limited success can be attributed to many building blocks of intelligence being absent in those efforts. This paper seeks to provide the foundational elements for intelligence and addresses the drawback with the present efforts. Hence, the basic ideas outlined in the paper will and should be of interest to anyone interested in creating intelligence.

Surely, one paper cannot completely accomplish the task where millions of other efforts have failed; but what it can hope to do is guide future efforts to areas that will lead to greater success, inspired by how intelligence perhaps happens in us. That being said, a road-map which can act as an algorithm and also a set of hypothesis that can be tested and implemented computationally are provided. Just because we (readers / reviewers / students) do not see a connection or do not understand something, does not mean there is no connection or nothing to be understood. This paper puts forth the suggestion that in such cases, when something is not clear, to increase intelligence we need to ask questions and relate it to what we already know. Otherwise, intelligence will not increase. This message is constantly espoused in classrooms worldwide and but when we deem ourselves above that, whatever our role in the intelligence creation Eco-system, we have stopped to learn. This paper is as much about intelligence in humans as in machines. Intelligence happens as an unintended consequence due to curiosity left free. If something seems irrelevant, then we have stopped being curious and the level of intelligence plateaus off.

We have a great example of intelligent beings that have been created, which is us. We could debate as to whether this creator is evolution or an intelligent designer. But until higher powers intervene and provide the ultimate solution to create intelligence, we have to make do with marginal methods exemplified by curiously confident, trials and errors, such as this composition puts forth.

As we wait for the perfect solution, it is worth meditating upon what superior beings would do when faced with a complex situation, such as the one we are in. It is said that the Universe is but the Brahma's (Creator's) dream (Barnett 1907; Ramamurthi 1995; Ghatage 2010). Research (Effort / Struggle) can help us understand this world and maybe decipher the key to intelligent agents; Sleep (Ease / Peace of Mind) can help us create our own world; surely, creating intelligent beings would be a much smaller part of this new world (and reminding ourselves that sleep has many direct benefits to increase cognitive abilities). We just need to be mindful that the most rosy and well intentioned dreams can have unintended consequences and turn to nightmares (Nolan 2010; Lehrer 2010; Kashyap 2016).

Native to Australia (Clark 1993; End-note 23), "Koalas spend about 4.7 hours eating, 4 minutes traveling, 4.8 hours resting while awake and 14.5 hours sleeping in a 24-hour period" - (Nagy & Martin1985). See also (Smith 1979; Moyal 2008). The benefits of yoga on sleep quality are well documented (Cohen, Warneke, Fouladi, Rodriguez and Chaoul-Reich 2004; Khalsa 2004; Manjunath & Telles 2005; Chen, Chen, Chao,

Hung, Lin & Li 2009; Vera, Manzaneque, Maldonado, Carranque, Rodriguez, Blanca & Morell 2009; Endnote 22).

A lesson from close by and down under: We need to "Do Some Yoga and Sleep Like A Koala".

6 Acknowledgments and End-notes

1. Numerous seminar participants suggested ways to improve the manuscript. The views and opinions expressed in this article, along with any mistakes, are mine alone and do not necessarily reflect the official policy, or, position of either of my affiliations, or, any other agency. Dr. Yong Wang, Dr. Isabel Yan, Dr. Vikas Kakkar, Dr. Fred Kwan, Dr. William Case, Dr. Srikant Marakani, Dr. Qiang Zhang, Dr. Costel Andonie, Dr. Jeff Hong, Dr. Guangwu Liu, Dr. Humphrey Tung and Dr. Xu Han at the City University of Hong Kong; Dr. Richard Sylla, Dr. Adam Brandenburger, Dr. Richard Freedman, Dr. Robert Engle, Prof. Larry Zicklin, Prof. Seth Freeman, Dr. Laura Veldkamp, Dr. Ignacio Esponda at New York University; Dr. Liam Lenten at La Trobe University; and Dr. Paul Joseph, Dr. M. N. Neelakantan, Dr. V. K. Govindan, Dr. Moiuddin Kutty, Dr. M.P. Sebastian, Mr. Murali Krishnan at the National Institute of Technology Calicut; provided advice and more importantly encouragement to explore and where possible apply cross disciplinary techniques.

2. The students of SolBridge International School of Business continue to make us believe that the concepts in this paper are correct by displaying unbridled curiosity, which then unleashes remarkable remarks relevant to the topics discussed in the classroom. The faculty members of SolBridge International School of Business provided patient guidance and valuable suggestions on how to further this solution, and most importantly about the practical implications of even considering such an approach. In particular a discussion with Dr. Nurmukhammad Yusupov about the length of formula sheets for examinations, provided the insight that, over time as a semester or course progresses, the formula sheet (or, a summary of the lessons learnt) should not keep getting longer, but we need to be able to condense the material to keep it to a manageable size, so that the ideas can be more easily remembered and recalled, with perhaps pointers to a larger reference.

3. Certain six year old's that we know when questioned, "How to make computers intelligent?", responded by saying - "Have two computers; Use Google on one computer to find the answer and make the other computer use this answer". This remark gives us assurance that the concepts put forth in this paper about the way children learn by being curious, confident and most importantly by imitating good role models while attempting to solve problems, without getting frazzled, hold the key to increased

intelligence.

4. Deoxyribonucleic acid (DNA) is a molecule composed of two chains (made of nucleotides) that coil around each other to form a double helix carrying the genetic instructions used in the growth, development, functioning and reproduction of all known living organisms and many viruses. DNA, Wikipedia Link

5. Maybe, DNA hold the lessons from the lives of every ancestor we have ever had. Evolution is constantly coding the information, compressing it and passing forward, what is needed to survive better and to thrive, building what is essential right into our genes. For information storage in DNA and related applications see: Church, Gao & Kosuri 2012; Kosuri & Church 2014.

6. Universal Computing Machine, Wikipedia Link In computer science, a universal Turing machine (UTM) is a Turing machine (Minsky 1967; End-note 7) that can simulate an arbitrary Turing machine on arbitrary input. The universal machine essentially achieves this by reading both the description of the machine to be simulated as well as the input thereof from its own tape.

7. Turing Machine, Wikipedia Link A Turing machine is a mathematical model of computation that defines an abstract machine, which manipulates symbols on a strip of tape according to a table of rules. Despite the model's simplicity, given any computer algorithm, a Turing machine capable of simulating that algorithm's logic can be constructed (Sipser 2006).

8. Computer, Wikipedia Link A computer is a device that can be instructed to carry out sequences of arithmetic or logical operations automatically via computer programming.

9. MAC or Macintosh, Wikipedia Link The Macintosh (pronounced as MAK-in-tosh; branded as Mac since 1998) is a family of personal computers designed, manufactured, and sold by Apple Inc. since January 1984.

10. Personal Computer, Wikipedia Link A personal computer (PC) is a multi-purpose computer whose size, capabilities, and price make it feasible for individual use.

11. MAC vs MPC, Wikipedia Link Apple Computer, Inc. v. Microsoft Corporation, was a copyright infringement lawsuit in 1994 in which Apple Computer, Inc. (now Apple Inc.) sought to prevent Microsoft and Hewlett-Packard from using visual graphical user interface (GUI) elements that were similar to those in Apple's Lisa and Macintosh operating systems. Mac vs PC also refers to the rivalry between the two companies to dominate the personal computer market.

12. History Computing, Wikipedia Link The history of computing is longer than the history of computing hardware and modern computing technology and includes the history of methods intended for pen and

paper or for chalk and slate, with or without the aid of tables.

13. Computing Platform, Wikipedia Link A computing platform or digital platform is the environment in which a piece of software is executed. It may be the hardware or the operating system (OS), even a web browser and associated application programming interfaces, or other underlying software, as long as the program code is executed with it.

14. Cloud Computing, Wikipedia Link Cloud computing is shared pools of configurable computer system resources and higher-level services that can be rapidly provisioned with minimal management effort, often over the Internet. Cloud computing relies on sharing of resources to achieve coherence and economies of scale, similar to a public utility.

15. Quantum Computing, Wikipedia Link Quantum computing is computing using quantum-mechanical phenomena, such as superposition and entanglement. A quantum computer is a device that performs quantum computing. Such a computer is different from binary digital electronic computers based on transistors. Whereas common digital computing requires that the data be encoded into binary digits (bits), each of which is always in one of two definite states (0 or 1), quantum computation uses quantum bits or qubits, which can be in superpositions of states.

16. Turing Test, Wikipedia Link The Turing test, developed by Alan Turing in 1950, is a test of a machine's ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human. Turing proposed that a human evaluator would judge natural language conversations between a human and a machine designed to generate human-like responses. The evaluator would be aware that one of the two partners in conversation is a machine, and all participants would be separated from one another. The conversation would be limited to a text-only channel such as a computer keyboard and screen so the result would not depend on the machine's ability to render words as speech (Turing originally suggested a teleprinter, one of the few text-only communication systems available in 1950). If the evaluator cannot reliably tell the machine from the human, the machine is said to have passed the test. The test does not check the ability to give correct answers to questions, only how closely answers resemble those a human would give.

17. Searle's Chinese room thought experiment, Wikipedia Link Searle's thought experiment begins with this hypothetical premise: suppose that artificial intelligence research has succeeded in constructing a computer that behaves as if it understands Chinese. It takes Chinese characters as input and, by following the instructions of a computer program, produces other Chinese characters, which it presents as output. Suppose, says Searle, that this computer performs its task so convincingly that it comfortably passes the Turing test: it convinces a human Chinese speaker that the program is itself a live Chinese

speaker. To all of the questions that the person asks, it makes appropriate responses, such that any Chinese speaker would be convinced that they are talking to another Chinese-speaking human being.

This was originally phrased as: Searle supposes that he is in a closed room and has a book with an English version of the computer program, along with sufficient paper, pencils, erasers, and filing cabinets. Searle could receive Chinese characters through a slot in the door, process them according to the program's instructions, and produce Chinese characters as output.

The question Searle wants to answer is this: does the machine literally "understand" Chinese? Or is it merely simulating the ability to understand Chinese? Searle calls the first position "strong AI" and the latter "weak AI". Searle writes that "according to Strong AI, the correct simulation really is a mind. According to Weak AI, the correct simulation is a model of the mind." He also writes: "On the Strong AI view, the appropriately programmed computer does not just simulate having a mind; it literally has a mind."

18. Most people when posed the question: "Is Benedict Cumberbatch a mathematician of extraordinary ability?", would answer in the negative. This answer comes about, without most of us having met him, or, knowing whether he has been studying mathematics secretly for years but not having obtained any formal degree in the field; again highlighting, how we jump to conclusions. Unlikely as it seems, it is still probabilistically possible that he might be an exemplary mathematician. It is worth pointing out that most of us have this belief about Mr. Cumberbatch, without even knowing what his educational background is; though in this case, there are no surprises since, a quick search on Wikipedia or Google Benedict Cumberbatch, WIkipedia Link will reveal that his many years of formal training have been in acting.

19. Nassim Taleb and Daniel Kahneman discuss Trial and Error / IQ Points, among other things, at the New York Public Library on Feb 5, 2013. As Taleb explains, "it is trial with small errors that leads to progress". That being said, if there are big errors that might incapacitate the person trying the trial from further trials; as long as someone else has observed the attempts with huge errors, the rest of society benefits from it; assuming, of course, that the big blow up has left a non-trivial portion of society intact, or at-least not too shaken up. This concept is also illustrated in the next point about learning from the lessons history holds for us (point 2). (Ismail 2014) mentions the following quote from Taleb, "Knowledge gives you a little bit of an edge, but tinkering (trial and error) is the equivalent of 1,000 IQ points. It is tinkering that allowed the industrial revolution". This means that to match trial and error we need 1000 IQ points. But trial and error could still give the wrong outcomes. We can try and fail many times and still be wrong. So in our paper we make the assumption that we need 2000

IQ points to consistently make the right decisions. The subtle point that arises from this discussion is that: we need 2000 IQ points to be right all the time, but the problem is that the best of us has less than 200 IQ points. So Albert Einstein is nothing but a retard. Let us let this sink in; if Einstein is retarded we have no hope. So how do we deal with such hopeless retards? Should we not be more tolerant and considerate? Is it right to hold grudges against retards? Would it not be the right thing to try to be helpful the next time we see any retards, which is any human being.

20. Bass Model of Diffusion, Wikipedia Link The Bass Model or Bass Diffusion Model was developed by Frank Bass. It consists of a simple differential equation that describes the process of how new products get adopted in a population. The model presents a rationale of how current adopters and potential adopters of a new product interact. The basic premise of the model is that adopters can be classified as innovators or as imitators and the speed and timing of adoption depends on their degree of innovativeness and the degree of imitation among adopters. The Bass model has been widely used in forecasting, especially new products' sales forecasting and technology forecasting. Mathematically, the basic Bass diffusion is a Riccati equation (End-note 21) with constant coefficients.

21. Riccati Equation, Wikipedia Link In mathematics, a Riccati equation in the narrowest sense is any first-order ordinary differential equation that is quadratic in the unknown function. In other words, it is an equation of the form

$$y'(x) = q_0(x) + q_1(x)\, y(x) + q_2(x)\, y^2(x)$$

where $q_0(x) \neq 0$ and $q_2(x) \neq 0$. If $q_0(x) = 0$ the equation reduces to a Bernoulli equation, while if $q_2(x) = 0$ the equation becomes a first order linear ordinary differential equation. The equation is named after Jacopo Riccati (1676–1754) (see Riccati 1724).

22. Yoga, Wikipedia Link Yoga is a group of physical, mental, and spiritual practices which originated in ancient India.

23. Australia or Down Under, Wikipedia Link The term Down Under is a colloquialism which is variously construed to refer to Australia and New Zealand.

24. Tips on Nurturing Your Child's Curiosity The more curious a child is, the more he learns. Nurturing your child's curiosity is one of the most important ways you can help her become a lifelong learner. Babies are born learners, with a natural curiosity to figure out how the world works. Curiosity is the desire to learn. It is an eagerness to explore, discover and figure things out.

7 References

1. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). Molecular Biology of the Cell, Garland Science, New York.

2. Aly, M., & Moscovitch, M. (2010). The effects of sleep on episodic memory in older and younger adults. Memory, 18(3), 327-334.

3. Alpaydin, E. (2014). Introduction to machine learning. MIT press.

4. Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., ... & Zaharia, M. (2010). A view of cloud computing. Communications of the ACM, 53(4), 50-58.

5. Amir, Y., Ben-Ishay, E., Levner, D., Ittah, S., Abu-Horowitz, A., & Bachelet, I. (2014). Universal computing by DNA origami robots in a living animal. Nature nanotechnology, 9(5), 353.

6. Banerjee, A. V. (1992). A simple model of herd behavior. The Quarterly Journal of Economics, 107(3), 797-817.

7. Banerjee, A. V. (1993). The economics of rumours. The Review of Economic Studies, 60(2), 309-327.

8. Barnett, L. D. (1907). The Brahma Knowledge. An Outline of the Philosophy of the Vedanta as Set Forth by the Upanishads and by Sankara. Wisdom of the East series. E.P. Dutton Publishing, Boston, Massachusetts.

9. Bartholomew, D. J. (2004). Measuring intelligence: Facts and fallacies. Cambridge University Press.

10. Bass, F. M. (1969). A new product growth for model consumer durables. Management science, 15(5), 215-227.

11. Bass, F. M., Krishnan, T. V., & Jain, D. C. (1994). Why the Bass model fits without decision variables. Marketing science, 13(3), 203-223.

12. Beck, J. V., & Arnold, K. J. (1977). Parameter estimation in engineering and science. James Beck.

13. Bénabou, R., & Tirole, J. (2002). Self-confidence and personal motivation. The Quarterly Journal of Economics, 117(3), 871-915.

14. Berlyne, D. E. (1954). A theory of human curiosity. British Journal of Psychology, 45, 180-191.

15. Berlyne, D. E. (1966). Curiosity and exploration. Science, 153. 25-33.

16. Bhattacharyya, A. (1943). On a Measure of Divergence Between Two Statistical Populations Defined by their Probability Distributions, Bull. Calcutta Math. Soc., 35, pp. 99-110.

17. Bhattacharyya, A. (1946). On a measure of divergence between two multinomial populations. Sankhyā: The Indian Journal of Statistics, 401-406.

18. Blischke, K., Erlacher, D., Kresin, H., Brueckner, S., & Malangré, A. (2008). Benefits of sleep in motor learning–prospects and limitations. Journal of human kinetics, 20, 23-35.

19. Bottou, L. (2014). From machine learning to machine reasoning. Machine learning, 94(2), 133-149.

20. Burges, C. J. (2009). Dimension reduction: A guided tour. Machine Learning, 2(4), 275-365.

21. Burkardt, J. (2014). The Truncated Normal Distribution. Department of Scientific Computing Website, Florida State University.

22. Bush, R. R., & Mosteller, F. (2006). A mathematical model for simple learning. In Selected Papers of Frederick Mosteller (pp. 221-234). Springer New York.

23. Bush, R. R., & Mosteller, F. (1955). Stochastic models for learning.

24. Calkins, L., & Bellino, L. (1997). Raising Lifelong Learners: A Parent's Guide. Harper Collins, 1000 Keystone Industrial Park, Scranton, PA 18512; toll-free.

25. Cameron, J., & Wisher, W. (1991). Terminator 2: judgment day (Vol. 2). USA.

26. Campbell, D. T. (1956). Perception as substitute trial and error. Psychological review, 63(5), 330.

27. Campbell-Kelly, M. (2001). Not only Microsoft: The maturing of the personal computer software industry, 1982–1995. Business History Review, 75(1), 103-145.

28. Campos-Arceiz, A. (2009). Shit happens (to be useful)! Use of elephant dung as habitat by amphibians. Biotropica, 41(4), 406-407.

29. Carlton, J., & Annotations-Kawasaki, G. (1997). Apple: The inside story of intrigue, egomania, and business blunders. Random House Inc..

30. Castelvecchi, D. (2016). Can we open the black box of AI?. Nature, 538(7623), 20-23.

31. Catania, B., & Zarri, G. P. (2000). Intelligent database systems. Addison-Wesley.

32. Ceci, S. J., & Liker, J. K. (1986). A day at the races: A study of IQ, expertise, and cognitive complexity. Journal of Experimental Psychology: General, 115(3), 255.

33. Ceruzzi, P. E. (2003). A history of modern computing. MIT press.

34. Chen, K. M., Chen, M. H., Chao, H. C., Hung, H. M., Lin, H. S., & Li, C. H. (2009). Sleep quality, depression state, and health status of older adults after silver yoga exercises: cluster randomized trial. International journal of nursing studies, 46(2), 154-163.

35. Chesani, F., Mello, P., & Milano, M. (2017). Solving Mathematical Puzzles: A Challenging Competition for AI. AI Magazine, 38(3), 83-97.

36. Chiani, M., Dardari, D., & Simon, M. K. (2003). New exponential bounds and approximations for the computation of error probability in fading channels. Wireless Communications, IEEE Transactions on, 2(4), 840-845.

37. Church, G. M., Gao, Y., & Kosuri, S. (2012). Next-generation digital information storage in DNA. Science, 1226355.

38. Clément, R., Dörnyei, Z., & Noels, K. A. (1994). Motivation, self-confidence, and group cohesion in the foreign language classroom. Language learning, 44(3), 417-448.

39. Cody, W. J. (1969). Rational Chebyshev approximations for the error function. Mathematics of Computation, 23(107), 631-637.

40. Cohen, L., Warneke, C., Fouladi, R. T., Rodriguez, M., & Chaoul-Reich, A. (2004). Psychological adjustment and sleep quality in a randomized trial of the effects of a Tibetan yoga intervention in patients with lymphoma. Cancer, 100(10), 2253-2260.

41. Coogan, P. (2009). The Definition of the Superhero. A comics studies reader, 77.

42. Copeland, B. J. (2000). The turing test. Minds and Machines, 10(4), 519-539.

43. Corcoran, P., Coughlin, T., & Wozniak, S. (2016). Champions in our midst: the Apple doesn't fall far from the tree. IEEE Consumer Electronics Magazine, 5(1), 93-98.

44. Darwin, C. (1859). On the origin of species by means of natural selection. 1968. London: Murray Google Scholar.

45. Dasgupta, S., & Gupta, A. (1999). An elementary proof of the Johnson-Lindenstrauss lemma. International Computer Science Institute, Technical Report, 99-006.

46. Davis, M. (2011). The universal computer: The road from Leibniz to Turing. AK Peters/CRC Press.

47. Dawkins, R. (1976). The selfish gene. Oxford university press.

48. De Brigard, F. (2014). The nature of memory traces. Philosophy Compass, 9(6), 402-414.

49. DeDonno, M. A. (2016). The influence of IQ on pure discovery and guided discovery learning of a complex real-world task. Learning and Individual Differences, 49, 11-16.

50. Deng, L., & Yu, D. (2014). Deep learning: methods and applications. Foundations and Trends® in Signal Processing, 7(3–4), 197-387.

51. Denning, P. J. (2005). Is computer science science?. Communications of the ACM, 48(4), 27-31.

52. Derpanis, K. G. (2008). The Bhattacharyya Measure. Mendeley Computer, 1(4), 1990-1992.

53. Doidge, N. (2007). The brain that changes itself: stories of personal triumph from the frontiers of brain science/Norman.

54. Durant, W. (1968). The lessons of history.

55. Eco, U., & Chilton, N. (1972). The myth of Superman.

56. Eldredge, N. (2005). Darwin: discovering the tree of life. WW Norton & Company.

57. Ellenbogen, J. M. (2005). Cognitive benefits of sleep and their loss due to sleep deprivation. Neurology, 64(7), E25-E27.

58. Feltz, D. L. (1988). Self-confidence and sports performance. Exercise and sport sciences reviews, 16(1), 423-458.

59. Fingeroth, D. (2004). Superman on the Couch: What Superheroes Really Tell Us about Ourselves and Our Society. A&C Black.

60. Fodor, I. K. (2002). A survey of dimension reduction techniques. Technical Report UCRL-ID-148494, Lawrence Livermore National Laboratory.

61. Fogel, A. (2004). Infancy: Accessing Our Earliest Experiences. Theories of infant development, 204.

62. Frankl, P., & Maehara, H. (1988). The Johnson-Lindenstrauss lemma and the sphericity of some graphs. Journal of Combinatorial Theory, Series B, 44(3), 355-362.

63. Frankl, P., & Maehara, H. (1990). Some geometric applications of the beta distribution. Annals of the Institute of Statistical Mathematics, 42(3), 463-474.

64. Freiberger, P., & Swaine, M. (1999). Fire in the Valley: the making of the personal computer. McGraw-Hill Professional.

65. French, R. M. (1990). Subcognition and the limits of the Turing test. Mind, 99(393), 53-65.

66. French, R. M. (2000). The Turing Test: the first 50 years. Trends in cognitive sciences, 4(3), 115-122.

67. Gamble, W. C., & Cota-Robles, S. (2015). Guiding Curiosity: Nurturing Young Scientists. BookBaby.

68. Garland, H. (1977). Design innovations in personal computers. Computer, 10(3), 24-27.

69. Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. Nature, 521(7553), 452-460.

70. Ghatage, S. (2010). Brahma's Dream. Anchor Canada, Penguin Random House, Manhattan, New York.

71. Gibbons, R. (1992). A primer in game theory. Harvester Wheatsheaf.

72. Giles, J. (2005). Wisdom of the Crowd. Nature, 438(7066), 281.

73. Gopnik, A., Meltzoff, A. N., & Kuhl, P. K. (1999). The scientist in the crib: Minds, brains, and how children learn. William Morrow & Co.

74. Groom, W. (1986). Forrest Gump. Random House Digital, Inc..

75. Guo, T. (2015). Alan Turing: Artificial intelligence as human self-knowledge. Anthropology Today, 31(6), 3-7.

76. Hagen, U. (1991). Challenge for the Actor. Simon and Schuster.

77. Hagen, U. (1973). Respect for acting. John Wiley & Sons.

78. Halpern, J. H. (2004). Hallucinogens and dissociative agents naturally growing in the United States. Pharmacology & therapeutics, 102(2), 131-138.

79. Harnad, S. (1992). The Turing Test is not a trick: Turing indistinguishability is a scientific criterion. ACM SIGART Bulletin, 3(4), 9-10.

80. Haslem, W., Ndalianis, A., & Mackie, C. J. (Eds.). (2007). Super/Heroes: From Hercules to Superman. New Academia Publishing, LLC.

81. Haykin, S. S. (2004). Neural networks: A comprehensive foundation.

82. Haykin, S. S. (2009). Neural networks and learning machines (Vol. 3). Upper Saddle River, NJ, USA:: Pearson.

83. Hernández-Orallo, J., & Dowe, D. L. (2010). Measuring universal intelligence: Towards an anytime intelligence test. Artificial Intelligence, 174(18), 1508-1539.

84. Hernández-Orallo, J., Martínez-Plumed, F., Schmid, U., Siebers, M., & Dowe, D. L. (2016). Computer models solving intelligence test problems: Progress and implications. Artificial Intelligence, 230, 74-107.

85. Holt, J. (1995). How Children Learn. Classics in Child Development.

86. Horrace, W. C. (2005). Some results on the multivariate truncated normal distribution. Journal of Multivariate Analysis, 94(1), 209-221.

87. Howe, M. L., & Courage, M. L. (1993). On resolving the enigma of infantile amnesia. Psychological bulletin, 113(2), 305.

88. Ifrah, G., Harding, E. F., Bellos, D., & Wood, S. (2000). The universal history of computing: From the abacus to quantum computing. John Wiley & Sons, Inc.

89. Ismail, S. (2014). Exponential Organizations: Why new organizations are ten times better, faster, and cheaper than yours (and what to do about it). Diversion Books.

90. Jacobs, P. S. (2014). Text-based intelligent systems: Current research and practice in information extraction and retrieval. Psychology Press.

91. Jiang, Z., & Jain, D. C. (2012). A generalized Norton–Bass model for multigeneration diffusion. Management Science, 58(10), 1887-1897.

92. Johnson, W. B., & Lindenstrauss, J. (1984). Extensions of Lipschitz mappings into a Hilbert space. Contemporary mathematics, 26(189-206), 1.

93. Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. Cognitive psychology, 21(1), 60-99.

94. Jones, N. (2014). The learning machines. Nature, 505(7482), 146.

95. Kashyap, R. (2016). Notes on Uncertainty, Unintended Consequences and Everything Else. Working Paper.

96. Kashyap, R. (2017a). Combining Dimension Reduction, Distance Measures and Covariance. Working Paper.

97. Kashyap, R. (2017b). Imitation in the Imitation Game. Working Paper.

98. Kattumannil, S. K. (2009). On Stein's identity and its applications. Statistics & Probability Letters, 79(12), 1444-1449.

99. Keynes, J. M. (1937). The General Theory of Employment. The Quarterly Journal of Economics, 51(2), 209-223.

100. Keynes, J. M. (1971). The Collected Writings of John Maynard Keynes: In 2 Volumes. A Treatise on Money. The Applied Theory of Money. Macmillan for the Royal Economic Society.

101. Keynes, J. M. (1973). A treatise on probability, the collected writings of John Maynard Keynes, vol. VIII.

102. Khalsa, S. B. S. (2004). Treatment of chronic insomnia with yoga: A preliminary study with sleep–wake diaries. Applied psychophysiology and biofeedback, 29(4), 269-278.

103. Kiani, M., Panaretos, J., Psarakis, S., & Saleem, M. (2008). Approximations to the normal distribution function and an extended table for the mean range of the normal variables.

104. Kimeldorf, G., & Sampson, A. (1973). A class of covariance inequalities. Journal of the American Statistical Association, 68(341), 228-230.

105. Kosuri, S., & Church, G. M. (2014). Large-scale de novo DNA synthesis: technologies and applications. Nature methods, 11(5), 499.

106. Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. Nature Reviews. Neuroscience, 5(11), 831.

107. Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. Behavioral and Brain Sciences, 40.

108. Lawson, T. (1985). Uncertainty and economic analysis. The Economic Journal, 95(380), 909-927.

109. LeBlanc, M. D., & Weber-Russell, S. (1996). Text integration and mathematical connections: A computer model of arithmetic word problem solving. Cognitive Science, 20(3), 357-407.

110. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

111. Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. Minds and Machines, 17(4), 391-444.

112. Lee, K. Y., & Bretschneider, T. R. (2012). Separability Measures of Target Classes for Polarimetric Synthetic Aperture Radar Imagery. Asian Journal of Geoinformatics, 12(2).

113. Lehrer, J. (2010). The Neuroscience of Inception. Wired 26 Jul. 2010. Web. 13 Aug. 2013.

114. Lenneberg, E. H. (1967). The biological foundations of language. Hospital Practice, 2(12), 59-67.

115. Leuenberger, M. N., & Loss, D. (2001). Quantum computing in molecular magnets. Nature, 410(6830), 789.

116. Lightbown, P. M., Spada, N., Ranta, L., & Rand, J. (1999). How languages are learned (Vol. 2). Oxford: Oxford University Press.

117. Litman, J. A., & Spielberger, C. D. (2003). Measuring epistemic curiosity and its diversive and specific components. Journal of Personality Assessment, 80, 75-86.

118. Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. Psychological bulletin, 116(1), 75.

119. Loewy, E. H. (1998). Curiosity, imagination, compassion, science and ethics: Do curiosity and imagination serve a central function?. Health Care Analysis, 6(4), 286-294.

120. Lynch, T. J. (1985). Data compression: techniques and applications. Lifetime Learning Publications.

121. Mahajan, V. (1985). Innovation diffusion. John Wiley & Sons, Ltd.

122. Mahajan, V., Muller, E., & Bass, F. M. (1991). New product diffusion models in marketing: A review and directions for research. In Diffusion of technologies and social behavior (pp. 125-177). Springer, Berlin, Heidelberg.

123. Malomo, A. O., Idowu, O. E., & Osuagwu, F. C. (2006). Lessons from history: human anatomy, from the origin to the renaissance. Int. J. Morphol, 24(1), 99-104.

124. Mandik, P. (2007). Shit happens. Episteme, 4(2), 205-218.

125. Manes, S., & Andrews, P. (1993). Gates: How Microsoft's mogul reinvented an industry-and made himself the richest man in America. Simon & Schuster.

126. Manjunath, N. K., & Telles, S. (2005). Influence of Yoga & Ayurveda on self-rated sleep in a geriatric population. Indian Journal of Medical Research, 121(5), 683.

127. Martínez-Plumed, F., Ferri, C., Hernández-Orallo, J., & Ramírez-Quintana, M. J. (2017). A computational analysis of general intelligence tests for evaluating cognitive development. Cognitive Systems Research, 43, 100-118.

128. Mazur, J. E. (2015). Learning and behavior. Psychology Press.

129. McCarthy, J., & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. Readings in artificial intelligence, 431-450.

130. McCarthy, J. (2004). What is artificial intelligence. URL: http://www-formal. stanford. edu/jmc/whatisai. html.

131. McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5(4), 115-133.

132. McManus, H., & Hastings, D. (2005, July). 3.4. 1 A Framework for Understanding Uncertainty and its Mitigation and Exploitation in Complex Systems. In INCOSE International Symposium (Vol. 15, No. 1, pp. 484-503).

133. Michalakelis, C., Varoutas, D., & Sphicopoulos, T. (2010). Innovation diffusion with generation substitution effects. Technological Forecasting and Social Change, 77(4), 541-557.

134. Mill, J. (1829). Analysis of the Phenomena of the Human Mind (Vol. 1, 2). Longmans, Green, Reader, and Dyer.

135. Minsky, M. L. (1967). Computation: finite and infinite machines. Prentice-Hall, Inc.

136. Moor, J. H. (1976). An analysis of the Turing test. Philosophical Studies, 30(4), 249-257.

137. Moyal, A. (Ed.). (2008). Koala: a historical biography. CSIRO PUBLISHING.

138. Musshoff, F., Madea, B., & Beike, J. (2000). Hallucinogenic mushrooms on the German market—simple instructions for examination and identification. Forensic science international, 113(1), 389-395.

139. Nadel, L., & Zola-Morgan, S. (1984). Infantile amnesia. In Infant memory (pp. 145-172). Springer US.

140. Nagy, K. A., & Martin, R. W. (1985). Field Metabolic Rate, Water Flux, Food Consumption and Time Budget of Koalas, Phascolarctos Cinereus (Marsupialia: Phascolarctidae) in Victoria. Australian Journal of Zoology, 33(5), 655-665.

141. Nash, J. F. (1950). Equilibrium points in n-person games. Proceedings of the national academy of sciences, 36(1), 48-49.

142. Nass, M. M., & Cooper, L. N. (1975). A theory for the development of feature detecting cells in visual cortex. Biological cybernetics, 19(1), 1-18.

143. Nere, A., Hashmi, A., Cirelli, C., & Tononi, G. (2013). Sleep-dependent synaptic down-selection (I): modeling the benefits of sleep on memory consolidation and integration. Frontiers in neurology.

144. Newport, E. L. (1990). Maturational constraints on language learning. Cognitive science, 14(1), 11-28.

145. Nilsson, N. J. (2006). Human-Level Artificial Intelligence? Be Serious!. AI Magazine, 26(4), 68-75.

146. Niu, S. C. (2002). A stochastic formulation of the Bass model of new-product diffusion. Mathematical problems in Engineering, 8(3), 249-263.

147. Noels, K. A., Pon, G., & Clément, R. (1996). Language, identity, and adjustment: The role of linguistic self-confidence in the acculturation process. Journal of language and social psychology, 15(3), 246-264.

148. Nolan, C. (2010). Inception [film]. Warner Bros.: Los Angeles, CA, USA.

149. Norstad, J. (1999). The normal and lognormal distributions.

150. Oja, E. (1982). Simplified neuron model as a principal component analyzer. Journal of mathematical biology, 15(3), 267-273.

151. Okuda, S. M., Runco, M. A., & Berger, D. E. (1991). Creativity and the finding and solving of real-world problems. Journal of Psychoeducational assessment, 9(1), 45-53.

152. Owens, T. J. (1993). Accentuate the positive-and the negative: Rethinking the use of self-esteem, self-deprecation, and self-confidence. Social Psychology Quarterly, 288-299.

153. Palmer, C. (2002). 'Shit happens': The selling of risk in extreme sport. The Australian Journal of Anthropology, 13(3), 323-336.

154. Parsaye, K., & Chignell, M. (1993). Intelligent Database Tools and Applications: Hyperinformation access, data quality, visualization, automatic discovery. John Wiley & Sons, Inc..

155. Penrose, R. (1989). The Emperor's New Mind: concerning computers, brains and the laws of physics. Oxford Paperbacks.

156. Perrier, J. Y., Sipper, M., & Zahnd, J. (1996). Toward a viable, self-reproducing universal computer. Physica D: Nonlinear Phenomena, 97(4), 335-352.

157. Piskorski, J., & Neumann, G. (2000). An intelligent text extraction and navigation system. In Content-Based Multimedia Information Access-Volume 2 (pp. 1015-1032).

158. Popper, K. R. (2002). The poverty of historicism. Psychology Press.

159. Porter, L. (2014). Benedict Cumberbatch, Transition Completed: Films, Fame, Fans. Andrews UK Limited.

160. Preston, J., & Bishop, M. J. (2002). Views into the Chinese room: New essays on Searle and artificial intelligence. OUP.

161. Proudfoot, D. (2015). What turing himself said about the imitation game. IEEE Spectrum, 52(7), 42-47.

162. Ramamurthi, B. (1995). The fourth state of consciousness: The Thuriya Avastha. Psychiatry and clinical neurosciences, 49(2), 107-110.

163. Ramey-Gassert, L. (1997). Learning science beyond the classroom. The Elementary School Journal, 97(4), 433-450.

164. Reio Jr, T. G., Petrosko, J. M., Wiswell, A. K., & Thongsukmag, J. (2006). The measurement and conceptualization of curiosity. The Journal of Genetic Psychology, 167(2), 117-135.

165. Reynolds, R. (1992). Super heroes: A modern mythology. Univ. Press of Mississippi.

166. Riccati, J. (1724). Animadversiones in aequationes differentiales secundi gradus. Actorum Eruditorum Supplementa, 8(1724), 66-73.

167. Robertson, E. M., Pascual-Leone, A., & Press, D. Z. (2004). Awareness modifies the skill-learning benefits of sleep. Current Biology, 14(3), 208-212.

168. Rubinstein, M. E. (1973). A comparative statics analysis of risk premiums. The Journal of Business, 46(4), 605-615.

169. Rubinstein, M. (1976). The valuation of uncertain income streams and the pricing of options. The Bell Journal of Economics, 407-425.

170. Russell, S. J., & Norvig, P. (1995). Artificial Intelligence: A Modern Approach. Prentice-Hall, Englewood Cliffs, 25, 27.

171. Russell, S., Dewey, D., & Tegmark, M. (2015). Research priorities for robust and beneficial artificial intelligence. AI Magazine, 36(4), 105-114.

172. Samorini, G. (1992). The oldest representations of hallucinogenic mushrooms in the world. Integration, 2, 3.

173. Sato, C., Takeuchi, S., & Okude, N. (2011). Experience-based curiosity model: Curiosity extracting model regarding individual experiences of urban spaces. Design, User Experience, and Usability. Theory, Methods, Tools and Practice, 635-644.

174. Saygin, A. P., Cicekli, I., & Akman, V. (2000). Turing test: 50 years later. Minds and machines, 10(4), 463-518.

175. Schudson, M. (2007). The anarchy of events and the anxiety of story telling.

176. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural networks, 61, 85-117.

177. Schwartz, R. H., & Smith, D. E. (1988). Hallucinogenic mushrooms. Clinical pediatrics, 27(2), 70-73.

178. Scott, E. C. (2009). Evolution vs. creationism: An introduction (Vol. 62). Univ of California Press.

179. Searle, J. R. (1980). Minds, brains, and programs. Behavioral and brain sciences, 3(3), 417-424.

180. Searle, J. R. (1982). The Chinese room revisited. Behavioral and brain sciences, 5(2), 345-348.

181. Searle, J. R. (1990). Is the brain's mind a computer program. Scientific American, 262(1), 26-31.

182. Searle, J. (2001). Chinese Room Argument, The. Encyclopedia of cognitive science.

183. Searle, J. R. (2004). Mind: a brief introduction. Oxford University Press.

184. Shlens, J. (2014). A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100.

185. Simon, H. A. (1962). The Architecture of Complexity. Proceedings of the American Philosophical Society, 106(6), 467-482.

186. Sipser, M. (2006). Introduction to the Theory of Computation (Vol. 2). Boston: Thomson Course Technology.

187. Smith, M. (1979). Behaviour of the Koala, Phascolarctos Cinereus Goldfuss, in Captivity. 1. Non-Social Behaviour. Wildlife Research, 6(2), 117-129.

188. Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. Child development, 1114-1128.

189. Soranzo, A., & Epure, E. (2014). Very simply explicitly invertible approximations of normal cumulative and normal quantile function. Applied Mathematical Sciences, 8(87), 4323-4341.

190. Sorzano, C. O. S., Vargas, J., & Montano, A. P. (2014). A survey of dimensionality reduction techniques. arXiv preprint arXiv:1403.2877.

191. Stageberg, N. C. (1968). Structural Ambiguity for English Teachers. In Selected Addresses Delivered at the Conference on English Education (No. 6, pp. 29-34). National Council of Teachers of English.

192. Stein, C. M. (1973). Estimation of the mean of a multivariate normal distribution. Proceedings of the Prague Symposium of Asymptotic Statistics.

193. Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. The annals of Statistics, 1135-1151.

194. Sternberg, R. J. (2018). Speculations on the role of Successful Intelligence in solving contemporary world problems. Journal of Intelligence, 6(1), 4.

195. Storer, J. (1988). Data compression. Computer Science Press, Rockville, Maryland.

196. Swanson, D. R. (1977). Information retrieval as a trial-and-error process. The Library Quarterly, 47(2), 128-148.

197. Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: An introduction (Vol. 1, No. 1). Cambridge: MIT press.

198. Tahani, V. (1977). A conceptual framework for fuzzy query processing—a step toward very intelligent database systems. Information Processing & Management, 13(5), 289-303.

199. Takeuchi, A., & Amari, S. I. (1979). Formation of topographic maps and columnar microstructures in nerve fields. Biological Cybernetics, 35(2), 63-72.

200. Taleb, N. N. (2007). The black swan: the impact of the highly improbable. NY: Random House.

201. Teerapabolarn, K. (2013). Stein's identity for discrete distributions. International Journal of Pure and Applied Mathematics, 83(4), 565.

202. Thompson, K. F., Gokler, C., Lloyd, S., & Shor, P. W. (2016). Time independent universal computing with spin chains: quantum plinko machine. New Journal of Physics, 18(7), 073044.

203. Tou, F. N., Williams, M. D., Fikes, R., Henderson Jr, D. A., & Malone, T. W. (1982). RABBIT: An Intelligent Database Assistant. In AAAI (pp. 314-318).

204. Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433-460.

205. Vera, F. M., Manzaneque, J. M., Maldonado, E. F., Carranque, G. A., Rodriguez, F. M., Blanca, M. J., & Morell, M. (2009). Subjective sleep quality and hormonal modulation in long-term yoga practitioners. Biological psychology, 81(3), 164-168.

206. Von Ahn, L., Blum, M., & Langford, J. (2004). Telling humans and computers apart automatically. Communications of the ACM, 47(2), 56-60.

207. Wagner, R. K., & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: The role of tacit knowledge. Journal of personality and social psychology, 49(2), 436.

208. Waldfogel, S. (1948). The frequency and affective character of childhood memories. Psychological Monographs: General and Applied, 62(4), i.

209. Weinberg, R. A. (1989). Intelligence and IQ: Landmark issues and great debates. American Psychologist, 44(2), 98.

210. Williams, M. R. (1997). A history of computing technology. IEEE Computer Society Press.

211. Wolfe, T. (2005). The right stuff. Random House.

212. Wonglimpiyarat, J. (2012). Technology strategies and standard competition—Comparative innovation cases of Apple and Microsoft. The Journal of High Technology Management Research, 23(2), 90-102.

213. Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. The knowledge engineering review, 10(2), 115-152.

214. Yannakakis, G. N., & Togelius, J. (2015). A panorama of artificial and computational intelligence in games. IEEE Transactions on Computational Intelligence and AI in Games, 7(4), 317-335.

215. Yang, M. (2008). Normal log-normal mixture, leptokurtosis and skewness. Applied Economics Letters, 15(9), 737-742.

216. You, J. (2015). Beyond the turing test. Science, 347(6218), 116-116.

217. Young, H. P. (2009). Learning by trial and error. Games and economic behavior, 65(2), 626-643.

218. Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud computing: state-of-the-art and research challenges. Journal of internet services and applications, 1(1), 7-18.

219. Zogheib, B., & Hlynka, M. (2009). Approximations of the Standard Normal Distribution. University of Windsor, Department of Mathematics and Statistics.

8 Appendix: From Words to Symbols, A Curious and Confident Model of Intellect

8.1 Notation and Terminology for Key Results

- $D_{BC}\left(p_i, p_i'\right)$, the Bhattacharyya Distance between two multinomial populations each consisting of k categories classes with associated probabilities $p_1, p_2, ..., p_k$ and $p_1', p_2', ..., p_k'$ respectively.
- $\rho\left(p_i, p_i'\right)$, the Bhattacharyya Coefficient.
- $D_{BC-N}(p, q)$ is the Bhattacharyya distance between p and q normal distributions or classes.
- $D_{BC-MN}\left(p_1, p_2\right)$ is the Bhattacharyya distance between two multivariate normal distributions, $\boldsymbol{p_1}, \boldsymbol{p_2}$ where $\boldsymbol{p_i} \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$.
- $D_{BC-TN}(p, q)$ is the Bhattacharyya distance between p and q truncated normal distributions or classes.

- $D_{BC-TMN}\left(p_1, p_2\right)$ is the Bhattacharyya distance between two truncated multivariate normal distributions, $\boldsymbol{p_1}, \boldsymbol{p_2}$ where $\boldsymbol{p_i} \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \boldsymbol{a}_i, \boldsymbol{b}_i)$.

8.2 Bass Model of Diffusion for Information Accumulation

Collecting new pieces of information is the behavior parallel we draw to creating curiosity in our agents. One of the simplest forms of the Bass model and also the original one from the pioneer (Bass 1969) can be written as,

$$\frac{f\left(t\right)}{1-F\left(t\right)} = p + qF\left(t\right)$$

$$F\left(t\right) = \int_0^t f\left(u\right) du$$

Here,

$f\left(t\right)$, is the change of the installed base fraction or the likelihood of purchase at time t.

$F\left(t\right)$, is the installed base fraction.

p, is the coefficient of innovation.

q, is the coefficient of imitation.

Sales $S\left(t\right)$ is the rate of change of installed base (i.e. adoption), that is, $f\left(t\right)$ multiplied by the ultimate market potential m. This is given by,

$$S\left(t\right) = mf\left(t\right)$$

$$S\left(t\right) = m\frac{\left(p+q\right)^2}{p}\frac{e^{-(p+q)t}}{\left(1+\frac{q}{p}e^{-(p+q)t}\right)^2}$$

(Niu 2002) is a stochastic formulation of the Bass model of new product diffusion. As alternative models, we could use models used in economics for the spread of rumors (Banerjee 1993 has a discussion of information transmission processes, which for our purposes are similar to information collection processes) and herd behavior. (Banerjee 1992) a sequential decision model in which each decision maker looks at the decisions made by previous decision makers in taking her own decision, showing that the decision rules that are chosen by optimizing individuals will be characterized by herd behavior; i.e., people will be doing what others are doing rather than using their information.

8.3 Knowledge Store

We could use the developments in the field of text parsing and storing (Piskorski & Neumann 2000; Jacobs 2014), to create keyword based database(s), to hold bits of learning that the agent has gathered. The knowledge store has to be processed periodically to establish and reestablish the connections between the different stored elements. This feature would be of assistance in being able to recollect what has been learnt.

The connections are established based on the Bhattacharyya distance discussed next, and when appropriate, we use dimension reduction techniques so that this distance measure could be applied (For intelligent database systems and related query developments, see: Tahani 1977; Tou, Williams, Fikes, Henderson Jr, & Malone 1982; Parsaye & Chignell 1993).

8.4 Bhattacharyya Distance for Information Comparison

We use the Bhattacharyya distance (Bhattacharyya 1943, 1946) as a measure of similarity or dissimilarity between the probability distributions of the two entities we are looking to compare. These entities could be two information sources, two securities, groups of securities, markets or any statistical populations that we are interested in studying. The Bhattacharyya distance is defined as the negative logarithm of the Bhattacharyya coefficient.

$$D_{BC}\left(p_i, p_i'\right) = -\ln\left[\rho\left(p_i, p_i'\right)\right]$$

The Bhattacharyya coefficient is calculated as shown below for discrete and continuous probability distributions.

$$\rho\left(p_i, p_i'\right) = \sum_{i}^{k} \sqrt{p_i p_i'}$$

$$\rho\left(p_i, p_i'\right) = \int \sqrt{p_i\left(x\right) p_i'\left(x\right)} dx$$

Bhattacharyya's original interpretation of the measure was geometric (Derpanis 2008). He considered two multinomial populations each consisting of k categories classes with associated probabilities $p_1, p_2, ..., p_k$ and $p_1', p_2', ..., p_k'$ respectively. Then, as $\sum_i^k p_i = 1$ and $\sum_i^k p_i' = 1$, he noted that $(\sqrt{p_1}, ..., \sqrt{p_k})$ and $(\sqrt{p_1'}, ..., \sqrt{p_k'})$ could be considered as the direction cosines of two vectors in $k-$dimensional space referred to a system of orthogonal co-ordinate axes. As a measure of divergence between the two populations Bhattacharyya used the square of the angle between the two position vectors. If θ is the angle between the vectors then:

$$\rho\left(p_i, p_i'\right) = cos\theta = \sum_{i}^{k} \sqrt{p_i p_i'}$$

Thus if the two populations are identical: $cos\theta = 1$ corresponding to $\theta = 0$, hence we see the intuitive motivation behind the definition as the vectors are co-linear. Bhattacharyya further showed that by passing to the limiting case a measure of divergence could be obtained between two populations defined in any way given that the two populations have the same number of variates. The value of coefficient then lies between 0 and 1.

$$0 \leq \rho\left(p_i, p_i'\right) \leq = 1$$

$$0 \leq D_{BC}\left(p_i, p_i'\right) \leq \infty$$

We get the following formulae (Lee and Bretschneider 2012) for the Bhattacharyya distance when applied to the case of two uni-variate normal distributions.

$$D_{BC-N}(p,q) = \frac{1}{4}\ln\left(\frac{1}{4}\left(\frac{\sigma_p^2}{\sigma_q^2}+\frac{\sigma_q^2}{\sigma_p^2}+2\right)\right)+\frac{1}{4}\left(\frac{(\mu_p-\mu_q)^2}{\sigma_p^2+\sigma_q^2}\right)$$

σ_p is the variance of the $p-$th distribution,

μ_p is the mean of the $p-$th distribution, and

p, q are two different distributions.

The original paper on the Bhattacharyya distance (Bhattacharyya 1943) mentions a natural extension to the case of more than two populations. For an M population system, each with k random variates, the definition of the coefficient becomes,

$$\rho\left(p_1,p_2,...,p_M\right) = \int\cdots\int\left[p_1\left(x\right)p_2\left(x\right)...p_M\left(x\right)\right]^{\frac{1}{M}}dx_1\cdots dx_k$$

For two multivariate normal distributions, $\boldsymbol{p_1}, \boldsymbol{p_2}$ where $\boldsymbol{p_i} \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$,

$$D_{BC-MN}\left(p_1,p_2\right) = \frac{1}{8}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)+\frac{1}{2}\ln\left(\frac{\det\boldsymbol{\Sigma}}{\sqrt{\det\boldsymbol{\Sigma}_1\,\det\boldsymbol{\Sigma}_2}}\right),$$

$\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are the means and covariances of the distributions, and $\boldsymbol{\Sigma} = \frac{\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2}{2}$. We need to keep in mind that a discrete sample could be stored in matrices of the form A and B, where, n is the number of observations and m denotes the number of variables captured by the two matrices.

$$\boldsymbol{A_{m\times n}} \sim \mathcal{N}\left(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_1}\right)$$

$$\boldsymbol{B_{m\times n}} \sim \mathcal{N}\left(\boldsymbol{\mu_2}, \boldsymbol{\Sigma_2}\right)$$

8.5 Dimension Reduction before Information Comparison

A key requirement to apply the Bhattacharyya distance in practice is to have data-sets with the same number of dimensions. (Fodor 2002; Burges 2009; Sorzano, Vargas and Montano 2014) are comprehensive collections of methodologies aimed at reducing the dimensions of a data-set using Principal Component Analysis or Singular Value Decomposition and related techniques. (Johnson and Lindenstrauss 1984) proved a fundamental result (JL Lemma) that says that any n point subset of Euclidean space can be embedded in $k = O(log\frac{n}{\epsilon^2})$ dimensions without distorting the distances between any pair of points by more than a factor of $(1 \pm \epsilon)$, for any $0 < \epsilon < 1$. Whereas principal component analysis is only useful when the original data points are inherently low dimensional, the JL Lemma requires absolutely no assumption on the original data. Also,

note that the final data points have no dependence on d, the dimensions of the original data which could live in an arbitrarily high dimension. We use the version of the bounds for the dimensions of the transformed subspace given in (Frankl and Maehara 1988; 1990; Dasgupta and Gupta 1999).

Lemma 1. *For any $0 < \epsilon < 1$ and any integer n, let k be a positive integer such that*

$$k \geq 4 \left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3} \right)^{-1} \ln n$$

Then for any set V of n points in $\boldsymbol{R}^d$, there is a map $f : \boldsymbol{R}^d \to \boldsymbol{R}^k$ such that for all $u, v \in V$,

$$(1 - \epsilon) \|u - v\|^2 \leq \|f(u) - f(v)\|^2 \leq (1 + \epsilon) \|u - v\|^2$$

Furthermore, this map can be found in randomized polynomial time and one such map is $f = \frac{1}{\sqrt{k}} Ax$ where, $x \in \boldsymbol{R}^d$ and A is a $k \times d$ matrix in which each entry is sampled i.i.d from a Gaussian $N(0, 1)$ distribution.

(Kashyap 2017a) provides expressions for the density functions after dimension transformation when considering log normal distributions, truncated normal and truncated multivariate normal distributions (Norstad 1999; Horrace 2005; Kiani, Panaretos, Psarakis & Saleem 2008; Yang 2008; Burkardt 2014). These results are applicable in the context of many variables observed in real life such as stock prices, heart rates, inventory levels, and volatilities, which do not take on negative values. We also require the expression for the dimension transformed normal distribution and possibly, techniques for numerical approximations. (Cody 1969; Chiani, Dardari & Simon 2003; Zogheib & Hlynka 2009; Soranzo & Epure 2014), since it is a better candidate to model a reward variable, which could take on negative values. A relationship between covariance (Stein 1973; 1981; Kimeldorf & Sampson 1973; Rubinstein 1973; 1976; Kattumannil 2009; Teerapabolarn 2013) and distance measures is also derived. We point out that these mathematical concepts have many uses outside the domain of artificial intelligence.