Chapter 1

Modeling human intelligence with probabilistic program induction

1.1 Objective

Here i would try to provide a broader perspective on the authors idea.Tenenbaum has been working on the core idea of this paper for quite some time - compositionality, causality, and learning to learn.Though i tried to stick with the authors present work,i have to draw considerably from his other works too.(Building Machines That Learn and Think Like People - 2016).Wherever i couldn't completely digress to my heart's content(space constraints) i have cited the original resource from which i learned as i think it's important to deeply appreciate certain ideas to see where a new idea fits into the fabric.If the line of ideas presented here seems out of place they may either be so or a brief pondering would fit them onto the edifice.

1.1.1 The Learning problem

If we look at the theoretical basis of machine learning we see that there is this trade off between model complexity and generalizability.We need more data for a more complex model to generalize well - (captured by vc dimension or much simpler bias-variance trade off).But humans seems to do one shot learning.Authors objective was to achieve human level performance in a task not through traditional statistical approach(refer *extras*(1)) but from the understanding of how humans do it ?

1.1.2 What they have shown ?

Authors have shown that through BPL(bayesian program learning) the engine can learn to perform one shot learning on par with humans.Following picture outlines their achievement..

By looking at the segway we can easily identify it's compositional parts - we can iden-

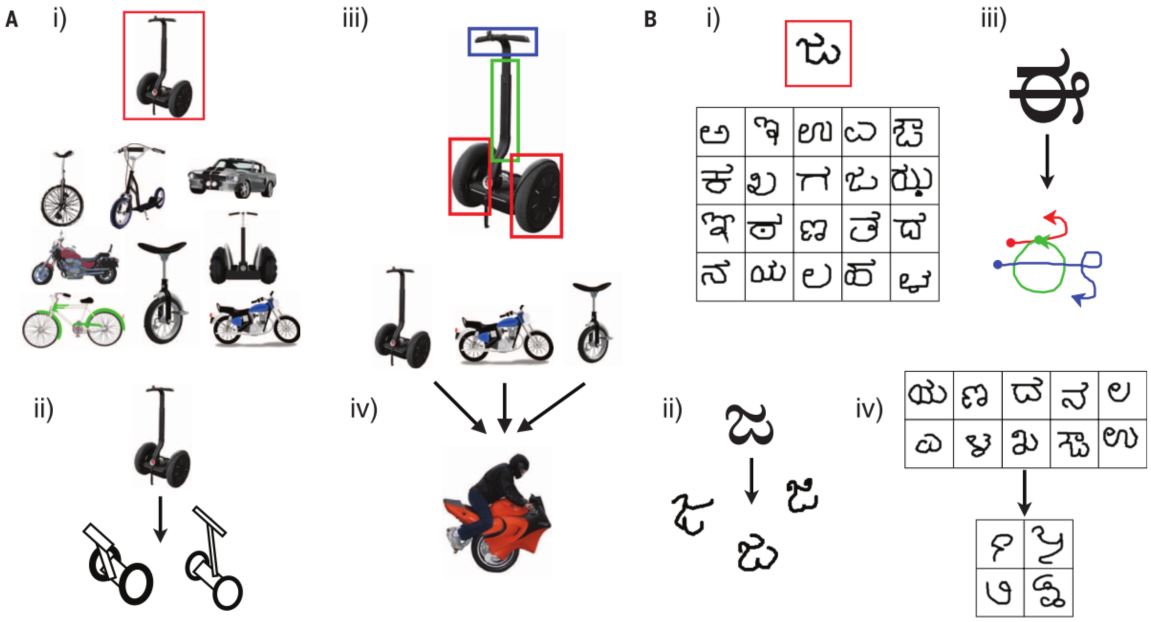


Fig. 1. People can learn rich concepts from limited data. (**A** and **B**) A single example of a new concept (red boxes) can be enough information to support the (i) classification of new examples, (ii) generation of new examples, (iii) parsing an object into parts and relations (parts segmented by color), and (iv) generation of new concepts from related concepts. [Image credit for (A), iv, bottom: With permission from Glenn Roberts and Motorcycle Mojo Magazine]

Figure 1.1

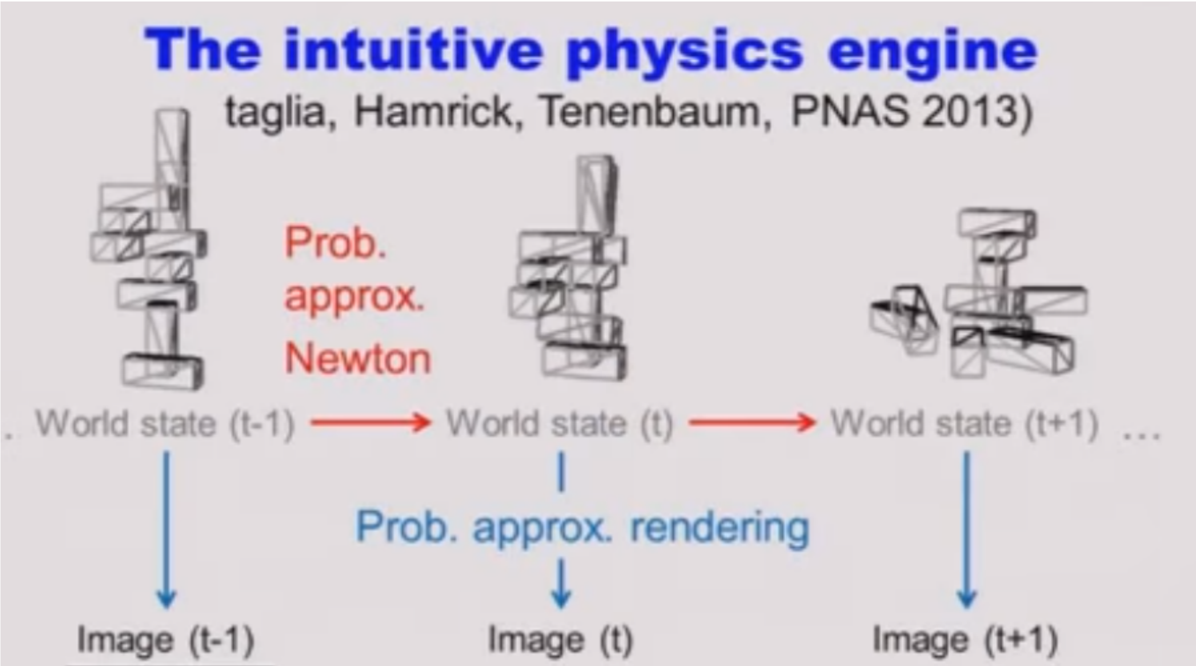
tify the wheels,suspension etc,draw it,generate new models by combinig with previous similar models.They have taken various hand written letters from various languages and tested their engine on 4 different tasks that involves learning a concept from single example,generating new examples,parsing whole into parts and relations and generating new concept from old ones.The detail of the engine are omitted in the paper though i would provide some of the details in the next section.But it has to (a) build causal models of the world that support explanation and understanding, rather than merely solving pattern recognition problems; (b) ground learning in intuitive theories of physics and psychology, to support and enrich the knowledge that is learned; and (c) harness compositionality and learning-to-learn to rapidly acquire and generalize knowledge to new tasks and situations.

They have incorporated inductive learning quite successfully through BPL- show a new letter - some body has drawn it - It must be sequence of strokes - strokes match most closely to 'A' - so it must be A.

1.1.3 What is BPL ?

Bayesian Program Learning (BPL) represents concepts as simple stochastic programs that is, structured procedures that generate new examples of a concept when executed. These programs allow the model to express causal knowledge about how the raw data are formed, and the probabilistic semantics allow the model to handle noise and perform creative tasks. Structure sharing across concepts is accomplished by the compositional reuse of stochastic primitives that can combine in new ways to create new concepts.
Rather than cluttering the writeup with the details of the experiment let's try to intuitively understand what's happening under the hood.We have probabilistic graphical models(Daphne kollers book and coursera course are amazing resource) that helps us to build general purpose causal models.Probabilistic programs are just more expressive versions of pgm's.HMM'S are great at this task but they miss much of the structure just like how flow charts miss much of the structure in a traditional computer program.A word isn't just a finite dimensional vector nor physics can be a transition matrix.Instead in BPL authors have used seperate programs to induce intuitive physics and psychology.They are built on two core insights:
1.Humans come with intuitive prior's - like an understanding of physical objects,intentional agents,their properties etc..
2.Learning is theory building(child as a scientist not data analyst.) BPL has been built by many scholars under tanenbaum for instance the physicse engine(which closely parallels

with game engines).I would also suggest to read Deep Convolutional Inverse Graphics Network - tanenbaum to understand further details.Similarly an intuitive psychology engine is also built in .We know through evolution enduring properties of the world are bulit in as priors,we don't come with blank slate.Then it is sensible to build corresponding priors into the algorithm.Some technical details and the results are covered in next section.



1.1.4 The details and the results

The following details are not directly mentioned but are very crucial to understand:
Authors have assumed the primitives(below figure) before performing the experiment.Notable point is that their accuracy is not so crucial to the performance of the algorithm as the sequence of strokes is learned by the algorithm which is more crucial for the accuracy.In one setting of the experiment,the algorithm is provided with a new example that is not in the training data,the compositional parts are inferred and then based on a similarity metric it is classifed appropriately.So,as assumed we are give basic elements like a turn,hook,wave etc called primitives.When these primitives are connected sequentially parts are formed - those that can be drawn with a single stroke.Here are the most crucial parts of the algorithm:

Search space of the algorithm is restricted in the following two ways: 1.The fact that we are drawing single stroke line from primitives constrains their location. 2.Now consider two stroke letters- like the combination of epsilon and the hook.We have three line ends there.Now by varying relative position and adding motor noise we generate different samples.It's important to notice how we are limiting search space here.

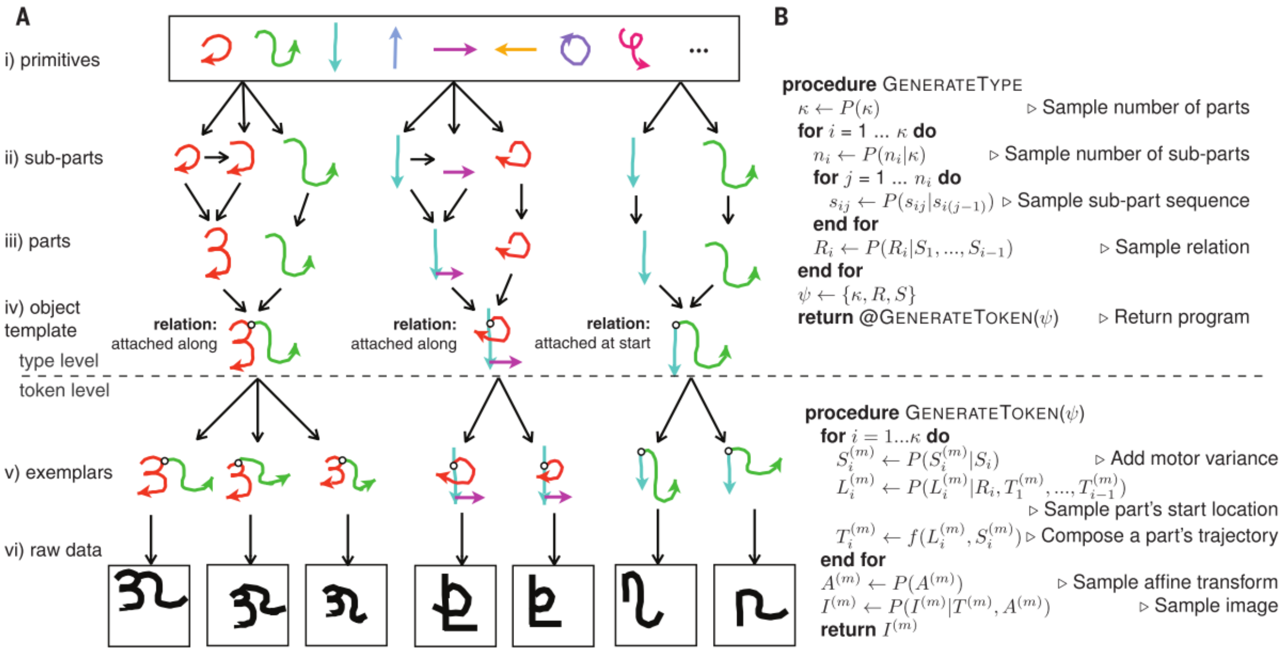


Fig. 3. A generative model of handwritten characters. (**A**) New types are generated by choosing primitive actions (color coded) from a library (i), combining these subparts (ii) to make parts (iii), and combining parts with relations to define simple programs (iv). New tokens are generated by running these programs (v), which are then rendered as raw data (vi). (**B**) Pseudocode for generating new types ψ and new token images $I^{(m)}$ for $m = 1, ..., M$. The function $f(\cdot, \cdot)$ transforms a subpart sequence and start location into a trajectory.

Note that BPL is a generative model so there can be multiple programs with different parses but still that isn't a problem as illustrated..

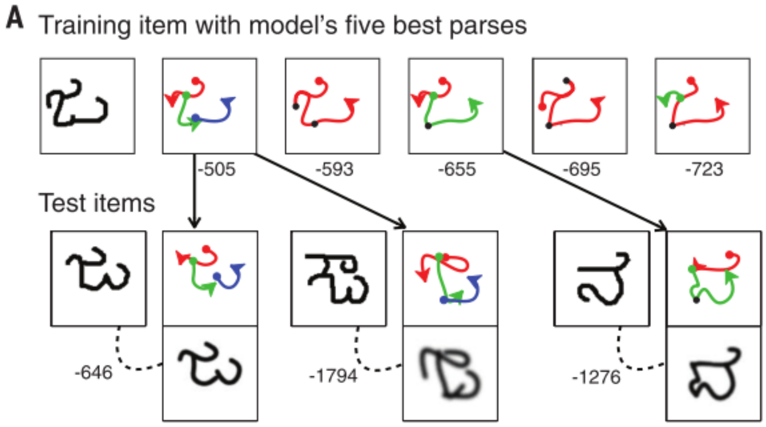


Fig. 4. Inferring motor programs from images. Parts are distinguished by color, with a colored dot indicating the beginning of a stroke and an arrowhead indicating the end. (**A**) The top row shows the five best programs discovered for an image along with their log-probability scores (Eq. 1). Subpart breaks are shown as black dots. For classification, each program was refit to three new test images (left in image triplets), and the best-fitting parse (top right) is shown with its image reconstruction (bottom right) and classification score (log posterior predictive probability). The correctly matching test item receives a much higher classification score and is also more cleanly reconstructed by the best programs induced from the training item. (**B**) Nine human drawings of three characters (left) are shown with their ground truth parses (middle) and best model parses (right).

RESULTS:Most notable part of the results is that BPL with leisons have dramatically increased error rate.It shows that by adding additional constraints by providing the knowledge of how the characters are made search space is reduced thus decreased error rate.

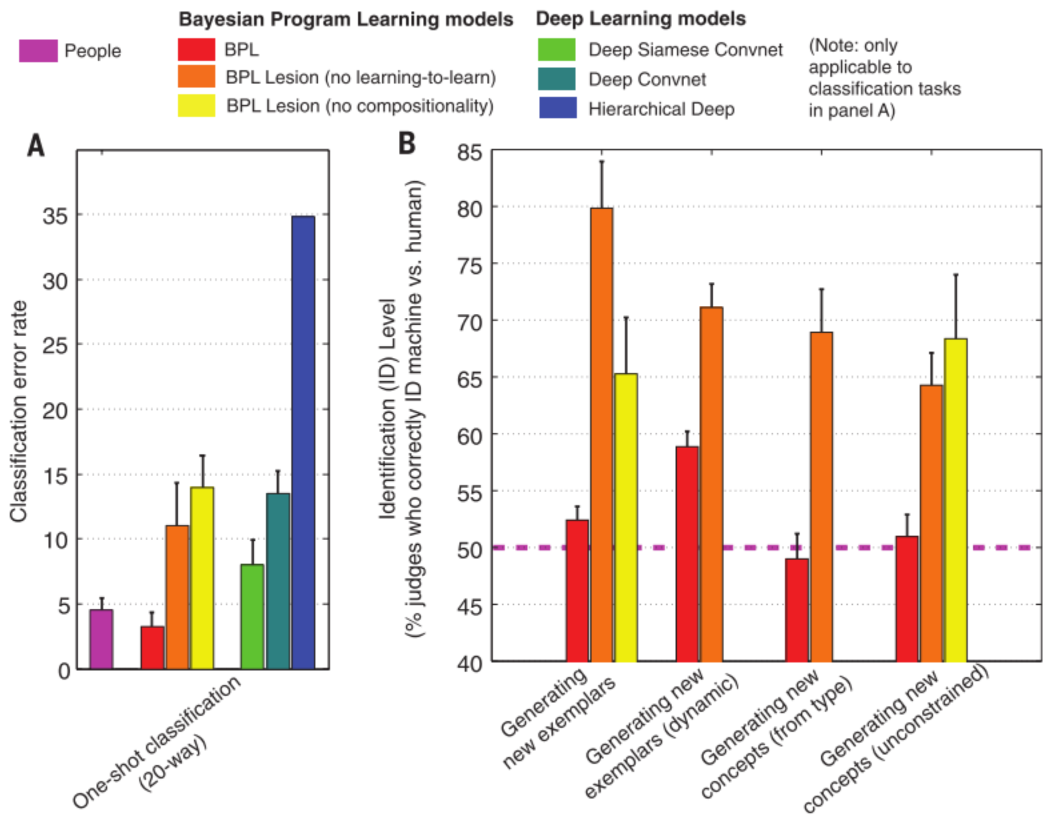


Fig. 6. Human and machine performance was compared on (A) one-shot classification and (B) four generative tasks. The creative outputs for humans and models were compared by the percent of human judges to correctly identify the machine. Ideal performance is 50%, where the machine is perfectly confusable with humans in these two-alternative forced choice tasks (pink dotted line). Bars show the mean ± SEM [*N* = 10 alphabets in (A)]. The no learning-to-learn lesion is applied at different levels (bars left to right): (A) token; (B) token, stroke order, type, and type.

1.1.5 The not so cool parts of the paper

1.The notion of providing inductive bias is sensible but authors have cleverly chosen the right problem setup to easily come up with right primitives(they have used deep learning).Consider language learning or visual scene understanding - what are the right primitives ? 2.For instance the authors haven't included line width into their basic primitive structure.Same sequence of strokes but different line widths could result in different estimation of the probabilities. 3.According to authors own note, BPL, sees less structure in visual concepts than people do. It lacks explicit knowledge of parallel lines, symmetry, optional elements such as cross bars in 7s, and connections between the ends of strokes and other strokes. 4.Now,here's the biggest drawback.It seems that the smartest people in the field are asking the wrong question.Firstly,We have to understand that humans are born with strong priors that are geared towards our survival in this universe.Under evolutionary time scale what has been constant - laws of physics and basic human nature has provided us with strong prior.We are not as fast learners as the A.I community led us to believe.If we are to put under slightly different laws of physics then i would bet we are far from one shot learners.A recent paper clearly drives this point-INVESTIGATING HUMAN PRIORS FOR PLAYING VIDEO GAMES(uc berkeley).

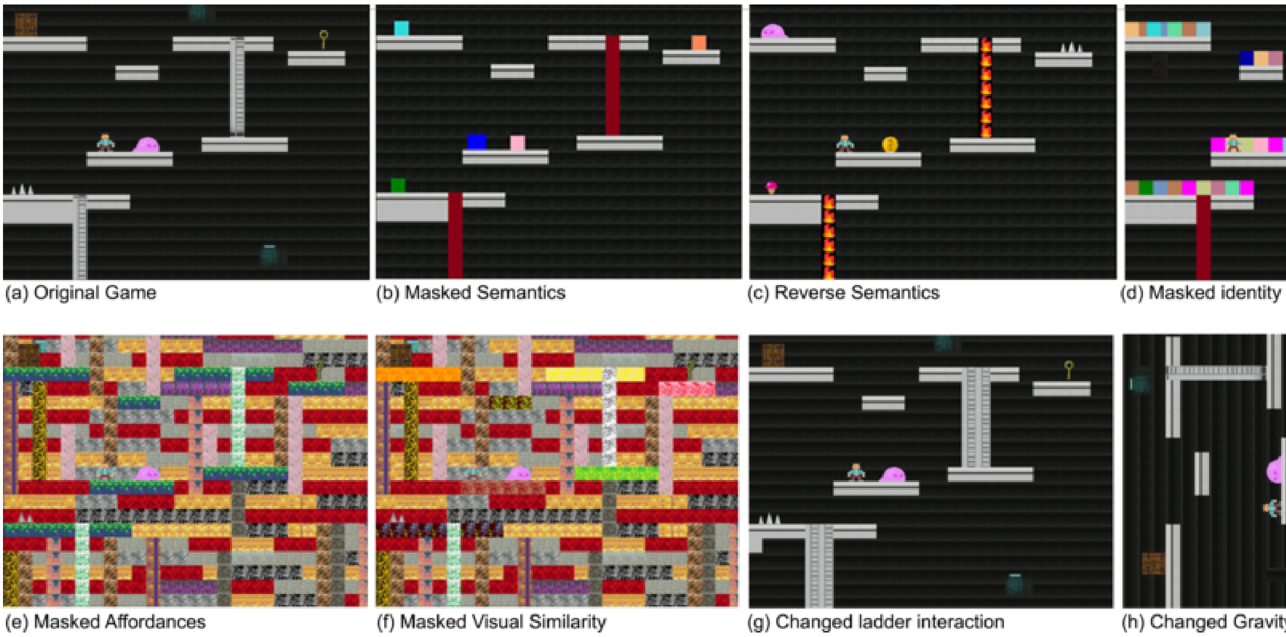


Figure 2: **Various game manipulations**. (a) Original version of the game. (b) Game w objects to ablate semantics prior. (c) Game with reversed associations as an alternate w semantics prior. (d) Game with masked objects and distractor objects to ablate the conce (e) Game with background textures to ablate affordance prior. (f) Game with backgrou and different colors for all platforms to ablate similarity prior. (g) Game with modifi hinder participant's prior about ladder interactions. (h) Rotated game to change partic about gravity.

A few minor changes like gravity direction has multiplied the learning time by few decades.Being equipped with strong prior knowledge can sometimes lead to constrained exploration that might not be optimal in all environments (Lucas et al., 2014; Bonawitz et al., 2011).Investigating human priors for every task and incorporating that into A.I is no dumber than rule based systems.

1.1.6 *Extra*

The following text is not part of the critique but this serves to provide additional context to the arguments in the text.

(1)I think we can break the history of A.I into three eras...

1.Prehistory - 1980 Symbolic A.I(Minsky being the major proponent).Largely over influenced by the success of the computers and drawing analogs to the brain people believed that brains are mere symbol manipulating machines.

2.Intelligence as statistics on large scale (1980's to 2000):Given enough data any function

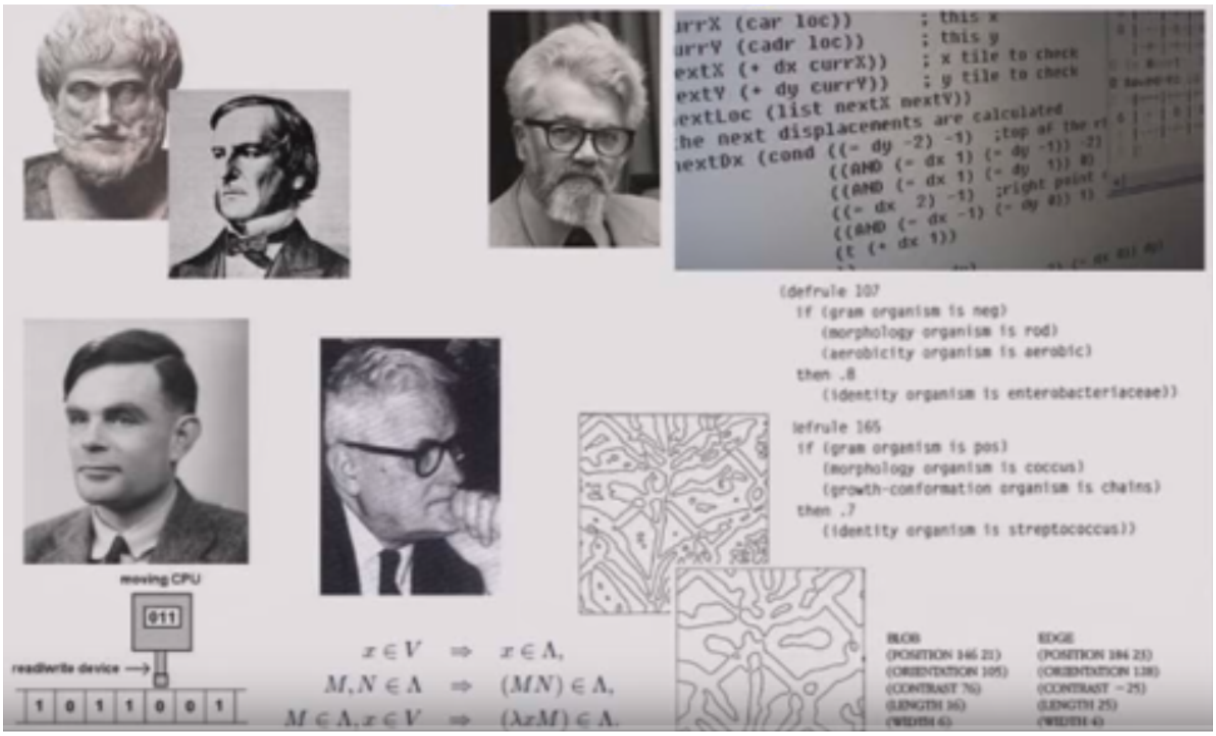


Figure 1.2: Symbolic era

can be estimated,so our brains must be doing data analysis.Much of the math of finding structure in the data(deep learning to spectral clustering methods) is worked out sort of underground in 60's but burst into the scene from 80's.

3.Present:We have made special purpose algorithms to beat human performance in several tasks.Consider state of computer vision - we have outperformed humans in object localization tasks but vision isn't just about that.When we see an image we also infer many intuitive things like the mental state of the person,possible future evolution of the system,social relationships of the people from their expressions etc.In fact this inability is one of the biggest impediment to the development of self-driving cars - visual scene understanding.People have realized that just by combining ideas from the previous two eras is not the way to go.Researchers like tanenbaum work on the intersection of cognitive science and A.I to develop more human like machines.