In this document, I lay out the bones of a theory of how language develops in the human mind. This theory depends crucially on the theoretical work of Noam Chomsky. It also crucially references Richard Dawkin's ideas about memetic selection. It also relies crucially on Purdy's theory of askesis, which hopefully you have received a copy of the paper that lays out that theory. It also relies crucially on the theory of Pirahã exceptionalism, due to Daniel L. Everett. It also relies crucially on the concept of recursive neural networks, as pioneered by Richard Socher, and exemplified in the paper (https://nlp.stanford.edu/pubs/SocherBauerManningNg_ACL2013.pdf) by Socher, Bauer, Manning, and Ng. Finally, it relies on the argument made in the paper (https://www.vicarious.com/2018/02/07/learning-concepts-through-sensorimotor-interactions/) "Behavior is Everything Towards Representing Concepts with Sensorimotor Contingencies", by Nicholas Hay, Michael Stark, Alexander Schlegel, Carter Wendelken, Dennis Park, Eric Purdy, and Tom Silver.

Please help me fill in the missing details. You can ask anyone for help, as long as you add their name to the paper.

1 Concept Learning Challenges

We recapitulate here some unpublished work due to Purdy (2015). We define a number of concept learning challenges that we take to be a gold standard for the correct use of concepts and language.

We start from a relatively naive view of a concept: a concept has some denotation (a dictionary definition), and some extension. (I.e., some set of situations in which adept users of the concept will hold that the concept is present, and some other set of situation in which adept users will hold that the concept is absent.)

We posit the following challenges, in rough order of difficulty:

1. Identify when the concept is present.
2. Identify when the concept is absent.
3. Manipulate one's environment until the concept is present.
4. Manipulate one's environment to achieve some arbitrary goal, all the while maintaining the absence of the concept. E.g., if the concept is "agent touches red object", then

2 Behavior isn't Everything, but it is a Start

We note the paper Behavior is Everything. We consider it to be neo-Skinnerian bunk, on the one hand (it does not pay the requisite toll to the Cognitive Revolution) and a fundamental contribution in the history of AI on the other hand (it demonstrates that modern deep learning systems are fully capable of solving the first two of the Concept Learning Challenges in a way that is

deeply grounded in sensorimotor experience and that requires absolutely no prior knowledge).

We thus believe that this work affords the best place to start for a set of experiments conclusively demonstrating that computers are capable of using language at a human level. The only thing missing from the PixelWorld environment is the ability to perform speech acts. We envision these being produced letter by letter by the agent; the agent should be able to produce speech acts up to a certain maximum size on every turn.

In return, the agent should receive speech acts back from the Teacher, a software component that produces a small list of stock response speech acts to situations in a dumb, hard-coded manner. The idea is that the Teacher is a native speaker who has learned a concept by rote, while the Agent is a native speaker who has learned a concept in a manner that is deeply grounded in sensorimotor experience.

3 A theory of Pirahã Exceptionalism

I posit that the thing that makes Pirahã such a powerful and durable language despite its many, many shortcomings is that it crucially incorporates a large number of continuous axes into its phonemic repertoire. In this theory, Pirahã speakers are sort of grandmasters of a new way of talking that the rest of us could stand to learn a thing or two about. (Slash, we already know these things in our own language, and use them to select tone, intonation, volume, and other such continuous attributes of our speech acts.)

Basically, let us suppose that each signifier comes with some number of continuous-valued attributes that the speaker can select in any way he or she chooses and the listener can interpret in any way he or she chooses. For the safety and comfort of both listener and speaker, let us bound all of these continuous values between -1 and 1. We posit that the interpretation side of this is performed by a mechanism akin to that of Socher et al.

How is language produced? We believe that it is chosen by a reinforcement learning agent to further both instrumental and ultimate goals of the agent. The agent makes use of what we term a Huffman encoding table with a discriminative post-filter; this is a short description of what we believe the cerebellum does. We have enclosed some notes on the cerebellum and Purdy's theory of askesis.

4 Galifreyan signifier notation

In this section, I lay out some notation that will hopefully act as a spur to the intuition, much as Feynman diagrams in physics or Leibniz notation in calculus do.

We will write English words and other linguistic signifiers in a funny way depending on how they were produced:

- $\langle word|$ is how we will write a word produced by the left brain

- $|word\rangle$ is how we will write a word produced by the right brain
- $\langle word\rangle$ is how we will write a word produced by the two halves of the brain working together. This is the typical pattern of an adult speaker of a particular language.
- $\widehat{|word|}$ is how we will write a word produced according to consensus reality and the accepted rules of the language in question
- $\widehat{\langle word|}$ is how we will write a word produced by the left brain, in accordance with consensus reality and the accepted rules of language
- $\widehat{|word\rangle}$ is how we will write a word produced by the right brain, in accordance with consensus reality and the accepted rules of language
- $\widehat{\langle word\rangle}$ is how we will write a word produced by the two halves of the brain working together, in accordance with consensus reality and the accepted rules of language. Such a word is what is generally called "true".

5 The Chomsky-Dawkins Law of Memetic Selection

Now that we have this notation, we can talk about different notions of truth with much more ease. We also introduce some rules for manipulating this notation:

The fundamental rule is that two signfiers can be combined into a larger signifier whenever they have the same "type", where type refers to the bracket notation introduced above.

Thus:

$$\langle Colorless\rangle\langle green\rangle\langle ideas\rangle\langle sleep\rangle\langle furiosly\rangle \tag{1}$$
$$\langle Colorlessgreen\rangle\langle ideas\rangle\langle sleep\rangle\langle furiosly\rangle \tag{2}$$
$$\langle Colorlessgreenideas\rangle\langle sleep\rangle\langle furiosly\rangle \tag{3}$$
$$\langle Colorlessgreenideassleep\rangle\langle furiosly\rangle \tag{4}$$
$$\langle Colorlessgreenideassleepfuriosly\rangle \tag{5}$$
$$\tag{6}$$

The above sentence is generally taken to be purest nonsense, i.e., it will never mean anything. However, language evolves, and things can come to mean things that they did not mean initially. If you translate the above sentence from Chomsky's language of thought to my own, you get something like the following:

$$\langle[Tasteless/bland(tothespeaker)/offensive(tothehearer)]French/existentialist/Sartre/schizophrenic/(abl \tag{7}$$

I hold that this sentence is actually deeply true, and gives one a good sense of which ideas will spread and which will not. For instance, Holocaust denialism and 9/11 trutherism both satisfy the conditions of the above sentence, and both are incredibly persistent despite the fact that both are obviously nonsense. Birtherism is another such idea, I would guess. Perhaps theories about the Kennedy assassination are another such example. Anyway, my point is that the sentence is less nonsense than it is some sort of law of memetic selection. (cf the work of Richard Dawkins, esp. The Selfish Gene) I have thus taken the liberty of naming it the Chomsky-Dawkins law of memetic evolution.

6 But how does this explain the acquistion of language?

Language is a tool for affecting the world around us. We use this tool efficiently. Efficiency is achieved using byte-pair encoding. (Google it!) This byte-pair enocding is accomplished neurally through the cerebellum, according to Purdy's theory of askesis, which you should have received a copy of with this document.