Larry Muhlstein

March 1st 2015

Rough Wittgensteinian Outline of *Communication and Causality*

1. Shannon’s communication theory requires that the meanings of signals be known *a priori*
   1. It treats only the transmission of signals.
   2. Computers and telecommunications equipment do know the “meanings” of their signals a priori.
      1. This is built into their workings. Computers and telecommunications equipment do not need to learn the meanings of their signals.
         1. The signals being the information transmitted, not how this information relates to meaning in the person. This was irrelevant to Shannon, as he was only concerned with transmitting information without loss
2. But the meanings of signals are not known to people a priori.
3. They cannot be known a priori because they vary between languages and even between people.
   1. They are therefore *synthetic* and *a posteriori* and must be learned from experience.
      1. Of course we have natural propensities to learn language and to learn languages and meanings of certain kinds more readily than those of others. However, the exact nature of the trade-off between nativism and empiricism does not matter here. What matters is only that we need to learn the meanings of words.
         1. The few points related to the degree to which we require native properties to learn language that matter to our argument are detailed later. Roughly, it is the case that we need to have some sort of bias towards learning some possible languages rather than others and we need to approximately share these biases between our communication partners.
4. And we observe that they are indeed learned from exposure to the world.
5. It is relatively clear how these meanings might be learned in the cases where the meanings and the signals are observed to co-occur.
   1. Though the specifics are not necessarily clear, we know that people can, at the most basic level, observe correlations between when a word is used and the state of the world at that moment
      1. The indeterminacy problem remains, but it is lightened a bit because many situations do not require a precise meaning for a word, but only an approximate meaning, which is often available.
         1. Furthermore, cross-situational experience with these words increases the specificity of their meanings
   2. We have many models to account for how this process works, and though it is much more complicated than we have shown, we will leave the issue to rest because it does not bear on the argument at hand.
6. It is not so clear, however, how these meanings are learned when the meaning is not observable, but the signal is.
   1. There needs to be something to pair to the signal, otherwise there is no way to differentiate between possible meanings.
      1. So if this component of the meaning is not directly observable, it must be indirectly inferable.
7. I contend that these meanings are learned through a process of causal inference
   1. This process at its simplest level is similar to the process given by Alva Noë’s sensorimotor contingency theory
   2. It is learned by observing the relations between what is said and what is done, of course given the state of the environment and other things inferable about the mental state of the partner.
8. In general, we can use observed relations or contingencies between what we do and what we observe to learn about how the world works in a causal manner.
   1. We know this from epistemological theories, learning theories, sensorimotor contingency theory, and others. Moreover, it is obvious.
9. We can use observed relations or contingencies between what we say and how our partner reacts to learn about how they understand our utterances in a causal manner.
   1. This is because learning in general is such an epistemic process and learning signals and language are but special cases of this general process.
   2. The causal learning processes and models that we use, however, may be very broad, and this theory does not have bearing on the specific ones.
      1. I only claim that that these processes should be consistent with a general theory of causal learning as proposed by various people including, most notably, Judea Pearl
         1. They need to be consistent because the world works in a causal way, and because the only way to predict something in a causally-governed world is by mirroring the causal structure of that world in your learning and generalization process.
            1. It is not lost on the author that the nature of this causality may be contingent and that causality may be defined only by the mechanisms we can use to predict it. In this way, though the world has a certain kind causal structure, this structure is captured only by models that capture the structure of this causality, whatever it might be.

This mirroring and capturing of causality to enable prediction is how science as a whole functions. This is most apparent in physics, but exists for all sciences that attempt to predict.

* + - 1. Since we need our understandings of signal and language meaning to generalize beyond the current situation, at least to some degree, we need to capture the causal structure of the meaning space in our model. If we need to learn such a model because the meaning space is not predetermined, then we must have faculties that enable us to learn in a causal way.

1. We can choose these utterances to probe the detailed ways in which others understand language
   1. Though this relates to the world only when such utterances are grounded—when they are related to utterances about observable things
   2. We need to probe the detailed ways in which others learn language because everyone speaks at least a slightly different language. This is because everyone learns from different experiences and everyone’s generalization processes are at least a little different. Therefore, when communicating with people, we must learn to speak their language as they learn to speak ours. That is, we must coordinate on a language that is to be used between us.
      1. We observe this type of coordination to be highly prevalent if not omnipresent in discourse. See the work of Simon Garrod and Martin Pickering for more details.
   3. Another reason that it must be the case that we probe the languages of others is because we have the ability to recognize conflicts between our understandings of signals, words, and utterances and those of others. Without some way to probe the meanings that others associate with their signals and utterances, we would have no way of detecting this conflict. Each of our utterances would constitute a semantic ship, all of which would pass undetected in the darkness of dead reckoning.
2. We can also choose these utterances to communicate things
   1. Utterances are understood in terms of the effects they have on the state of the discourse. This is of course within the current discourse. This is the “working meaning” of the utterance.
      1. Just like Wittgenstein’s “Slab!” meant to hand over a slab, the meanings of other utterances are determined by what they do, internally or externally to the state. Of course this is modulated by the context and the manner of utterance in a particular way.
      2. This mirrors the enactive view that actions are understood in terms of their perceivable effects, but extends this to be that actions are understood in terms of the effects that can be inferred by the perceivable effects. This is equivalent to adding unobservable variables between the action and the perception in a causal graph.
      3. This predicts that utterances will have different meanings in different discourses, as they have different effects in different discourse states. We observe this to be the case.
         1. We can further describe this in terms of the goals of the utterance and of the discourse. Different utterances are produced and different meanings are interpreted depending on how they interface with goals.
            1. If we are out in the desert and we are burning and you utter “let’s find some shade,” “shade” might be satisfied by a building because the goal is simply to get out of the sun. If we are in the park enjoying the weather and want to read a computer screen and the same utterance is produced, then “shade” might not include an indoor environment, but only a covering from the sun.

This is achieved due to equivalence classes of meaning derived from the goals. The goal of preventing skin from burning is achieved by any means of getting out of the sun, but the goal of enjoying the weather and reading a computer screen is achieved only by finding an outdoor cover. The interpretation of the utterance “let’s find some shade” is dependent on the equivalence class of states that achieve the goal at hand.

This is consistent with the idea of ad-hoc concepts. The concept of shade is dependent on the goals and the context.

1. The component of an utterance’s purpose that is designed to tell us about the way our interlocutor uses language can be seen as the *exploration component*
2. The component of an utterance’s purpose that is designed to communicate something and therefore to move directly towards a goal can be seen as the utterance’s *exploitation component*
3. Another way to look at this involves the distinction between pragmatic and epistemic actions
   1. The exploration component is the *epistemic action component*
   2. The exploitation component is the *pragmatic action component*
4. We propose that speakers choose utterances primarily based on these two components
   1. Of course this choice is informed by the structure and nature of their goals, the confidence of their beliefs about the ways in which their interlocutors use language, the sensitivity of the goals to the language, and the importance of getting everything right
5. When making these inferences, we also bring other abstract knowledge to bear on the problem
   1. This knowledge includes anything that would give us an expectation as to how the partner is likely to use language and is likely to hold meaning including knowledge about their language community, their cultural affiliations, and more.
      1. Such knowledge can also be jointly inferred with the meaning of their language, as certain usages are likely to point towards language communities and regularities, and such regularities are likely to influence the interpretation of their language.
   2. We also make use of knowledge about their similarity to us in order to bias what we think they are likely to mean by a phrase (as they are likely to mean what we mean if they are similar to us, and in order to bias how they are likely to generalize concepts
      1. A key problem in language learning and learning in general is that there exist an infinite number of ways to generalize beyond observed data with none of these ways privileged over any other in an absolute sense. In order to learn and generalize, we must have *inductive biases* that cause us to favor certain types of generalizations over others. If everyone’s inductive biases were *independent* of each other’s, then it would be impossible to predict how other generalize their concepts, and therefore to map meaning to a shared language. In order to prevent this problem, we rely on the fact that all people have *similar inductive biases* to enable us to infer that the concepts that others associate with words are likely to be similar to ours given the data we observe. This helps to weaken what would otherwise be the death grip of Quine’s indeterminacy of translation.
         1. A metric space provides the basis for a bias, but the bias consists of both a metric and a function. This function is the shape of the bias or the generalization structure of the bias.
            1. The body provides a metric space in terms of sensory perception and motor actions and the perceived distances and similarities between various forms of each. The mind provides computational biases and therefore a shape/generalization structure.

Since we have similar bodies and similar minds, we can guarantee that we have similar biases. This enables us to co-generalize and therefore to communicate.

1. There are some meanings that are more difficult to express than others using a given language
   1. This difficulty is modulated by the precision or sensitivity of the meaning (how close it is to other meanings), the connections between this meaning and other meanings, the distance from ground, and the degree of exchangeability of the meaning
2. There are other meanings that are impossible to get evidence for and must be trusted or guessed based on non-communicative means
   1. These are the meanings that are fully exchangeable
   2. These meanings cannot be communicated in a verifiable way, meaning that there is no way for another agent to verify whether the meanings are one way or another using communication
      1. The meanings that are fully exchangeable includes the set of subjective phenomena
         1. The “explanatory gap” in philosophy of mind emerges in this way from the epistemic features of communication. Subjective beings cannot verifiably communicate their subjective experiences because they exchangeable with respect to the set of communicable utterances about them. They are neither observable nor ground-able in the observable and therefore they are *communicatively exchangeable*
3. A full theory of communication and causality should be able to predict, given a language, the difficulty of verifying the meanings of different words to a given degree as well as—independent of language, which meanings are not communicable
   1. Such a theory would be built as a Bayesian probably approximately correct model of learning