# Computational Mechanisms of Human Communication (and philosophical issues)

## Introduction

Current theory of communication is a highly fragmented affair. We have models and theory for many aspects of communication. [Go through the phenomena modeled each with a sentence and a citation. Ex: Information theoretic models capture the information transfer inherent in a communication process (Shannon, Ramscar, etc.)] But no single theoretical framework can account for all of the phenomena, and there are some which are not satisfactorily captured by even a single framework. [Go through the theories and very briefly discuss the phenomena that they do not capture and where they do not generalize.] Sperber and Wilson have pointed out that we should not expect such a unifying theory of communication as communication involves a vast number of processes of many kinds that are not alike [get an actual quote]. Though this is certainly true, it claims only that we cannot expect a unifying theory at a certain level of abstraction. [Go into more detail about what we cannot expect.] On the other hand, if we take a step backwards, it is possible to notice some common features across all communication. Though constructing a theoretical framework will certainly not provide the same level of constraint and mechanistic prediction as a theory of language and of grammar, it enables us to gain a perspective on communication that captures the fundamental features at work. By taking this view, we can then understand more specific mechanisms of communication in a new and rather helpful way. The commonalities across the diverse communication mechanisms will become apparent, and we will be able to explain other previously unnoticed or unformalized aspects of communication with respect to this framework. This is the task that this paper sets out to do.

[Need to motivate the need for a new theory. Explain some problems with current work.]

* Current theories are incomplete and fail to generalize
* They do not take into account conceptual alignment and the lowest level inferences about word meanings, especially in the abstract.
* They don’t really account for phenomena such as discourse alignment
* They are too rigid to account for the full extent of cognitive processes that go on when using language. It is not just a means of message passing and information transfer.
* No one theory accounts for all the phenomena we observe
* Many of the more complex phenomena are not accounted for at all within a formal framework
* Current NLP systems understand utterances as utterances and relatively independently of what they call context. Context can be much better represented as a contingency and then the prediction can run off of utterance contingencies as a basis instead of on the utterance itself
* Language acquisition in current theories is COMPLETELY INDEPENDENT of language use
* Noisy channel models are very prevalent in natural language recognition, but they are not really correct
* Information theoretic models
  + Do not account for the origin of common ground, nor do they account for conceptual alignment
  + Were developed for machine communication, which involves different constraints and processes.
* Coordination and convention based models
  + Lewis
  + Herb Clark
  + Frank and Goodman
  + These generally involve assumptions about common ground and about conceptual alignment
* Artificial grammars
  + Insufficiently flexible. Even so, they only fully account for communication when combined with a LAD and a coordination theory such as Lewis’
* Basic predictive models of communication
  + I.e. Andy Clark
  + These tend to be insufficiently detailed and do not account for the functions of words and language. They are a promising start though.
* Simulation semantics
  + This captures a lot of the nature of language understanding, given that the person knows and understands the language. However, they do not delve into the nature of communication
* Goal-directed planning and theory of mind
  + Action understanding as inverse planning (Baker, Saxe, Tenenbaum)
  + Adena Schachner’s work
  + This assumes that there are always goals that are being inferred and does not delve into the workings of language too much. Another good start though.
* Communication as invoking similarity
  + Meltzoff “Like Me”
  + Uri Hasson’s work
  + This introduces the important principle of similarity, but does not touch a theory of either language or communication. It serves more as neurological and experimental evidence that similarity is modulated by language and successful communication
* Flexible concepts and categories and inference
  + Ad Hoc Cognition
  + This brings up an essential realization that concepts are not fixed and modular and that communication involves heavily context-driven inference. However, as far as I can tell, there is no explicit proposal for a theory of communication or coordination
* Turing test like models
  + Not so much a theory of communication, but a good set of example cases that tend to demonstrate inferences about the mental states of others
* Cues and correction
  + Work like that of Ashley Micklos
  + Not sufficient because misunderstanding cues can only be generated when there is recognition of misunderstanding, which, in this theoretical framework, is inferred through misunderstanding cues, yielding a circular framework unless other processes are appealed to
* Words as handles on concepts
  + This is a powerful associations paradigm, but does not propose specific mechanisms of association and does not account for conceptual alignment

[First explain how these work and the extent of the phenomena they capture. Later demonstrate the shortcomings of each when applied to the general case of communication.]

Most of these have a lot of explanatory power within their spheres of influence, but none is sufficiently general to capture the nature of communication. Each is a special case and there is much debate and fogginess at the boundaries and overlaps of the phenomena each tool has been used to analyze. In order to repair this, it will be useful to specify a more general framework for understanding communication. Not only will this framework help to resolve conflicts between existing systems, but it will also help shed light on other phenomena that are essential for communication, but of which there has been little explicit study. We will begin with coordination problems such as those that were analyzed by David Lewis, but we will approach them in a slightly different manner informed by the new tools and perspectives of our time including Bayesian inference and generalization theory. From this beginning we will expand the domain to include broader domains of language and eventually we will relate the present model to the other explanatory frameworks covered above.

The initial theory will explore coordination in terms of contingencies between what is said and what is done in a similar manner to Lewis’ *Convention*. This idea of contingencies will then be expanded to include relations between actions and states more generally and then to cover, not only how contingencies are used for coordination, but also how observations of similar contingencies can help establish conventional contingency relations. This joint framework for learning and inference as well as for communication will become the backbone of the theory, which will later be expanded to cover a broader and more complex set of linguistic and non-linguistic communication practices through the application of Bayesian probabilistic learning theory and generalization theory. In contrast to the eventual generality and power of the theory as fully articulated and formed, we begin, as Lewis did, with a simple exploration of coordination problems.

## Coordination scenarios

In this section I will go through a variety of scenarios that require coordination and use these scenarios to introduce some features of the coordination and communication theory. These features and principles will be covered only roughly in this section, but hopefully their immersion within a concrete scenario will help serve as conceptual scaffolding to help us to construct the more formal theory that will begin in the next section.

Coordination problems necessarily involve a sort-of inference about the goals and actions of others. But they do not require explicit reasoning about their accompanying beliefs and mental states. This is because all that is required is coordination of behavior. As this is a simple case of a more general communication process that may involve these explicit inferences, we begin here. It should be noted, however, that while the theory is being developed in these coordination scenarios, its real power comes out only in the more general case of communication. Though many features of coordination problems have been accounted for in a formal framework (Lewis 1969), there does not yet exist a framework that generalizes beyond this to account for all the phenomena and processes of communication that we find essential.

[Discuss contingencies and coordination. Define contingencies.]

Contingencies involve observing the structure of the changes in observations that are introduced by performing an action. These are inferential contingencies. What they help to infer are the contingencies at work in the other agent. I.e. the relations between receiving a certain signal in a certain environmental state and their resultant behavior. Of course this is complicated dramatically by the unobservable internal state of the other agent. Though the contingency is also unobservable, it can be inferred through probing, but if it depends on other unobservable properties, then this inference cannot be complete. How we fill in the gaps in this incompletely determined inductive process will be covered in the section on generalization. For now we focus on the much simpler case where the action of the other person depends solely on the state of the world that is jointly observable. We will demonstrate the mechanism for inference of these simplified contingencies and how this enables coordination and communication.

1. Chess game
   1. Introductory for the notation and for the three types of contingency interpretations
      1. Simple scenario where moves made by the opponent enable inferences about the opponent’s understanding of the rules
      2. The three types are:
         1. Expected responses
            1. Those that are consistent with the current predictive model of the opponent
         2. Unexpected, but post-hoc explainable responses
            1. This means that they did not fit into the prior model of the opponent’s behavior or mental states, but that it was possible to find an alternative model to account for them. It is a kind of hypothesis switching

These types of inferences show a nice similarity to predictive models of communication. However, we can probably show that the prediction is not necessary, only that an intuitive theory about the other person’s mental state or intended message is necessary

* + - 1. Unreconcillable responses
         1. Those for which a consistent predictive model of the opponent’s behavior cannot be found
         2. These often result in the player giving a signal of misunderstanding or of interacting with the opponent in order to reconcile their misalignment
  1. Not particularly good evidence for why contingencies are the quantity of interest
     1. This example mostly deals with the types of inferences about the opponent, not the dependency of these inferences on the state of the world

1. Serving at Chipotle
   1. Demonstrates contingencies based on the station. i.e. given “black or pinto beans?” the reasonable responses of the customer are different from given “chicken, steak, or veggie?” and this cleanly defined set of responses yields a simple example of contingency-based reasoning about interlocutor understanding.
   2. Still retains a relatively simple space and representation
2. Doctor’s office visit
   1. Demonstrates how we can use contingencies alongside similarity assumptions to explicitly infer others’ mental states
   2. Different types of pains and sensations can be inferred through probing the other person and assuming a parallel between their mental state-response contingencies and your own
      1. This serves as a special case to the more general similarity/generalization topic and a simple introduction
      2. This process is done through the neurodynamic processes that are currently lumped under the mirror neuron moniker
3. Turing test
   1. Give linguistic examples of all the phenomena explored so far
   2. This should help show how all of these processes have interesting analogs in discourse
      1. The three types of contingency interpretations can be demonstrated in a variety of discourse examples
         1. Expected responses are simple to show
         2. Unexpected but reconcilable responses are the types of situations where a response was unexpected, but where it leads to a new interpretation of the communicator’s intended message or mental state
         3. Unreconcillable responses are those that simply do not make sense and for which the listener cannot find an interpretation that would enable them to make sense
      2. Contingency processes can be demonstrated simply by discourse situations where certain phrases would make sense only given an earlier state or piece of common ground
         1. This can also be used to back-infer a state that was not clearly known if the utterance would only make sense in a given state or with a given piece of common ground. This inference might be understood in a Bayesian Viterbi type way
         2. This helps introduce communication and coordination as a certain type of inference procedure
   3. It will set the stage for the formalization and for the eventual addition of generalization theory

## A formal approach to signaling and coordination

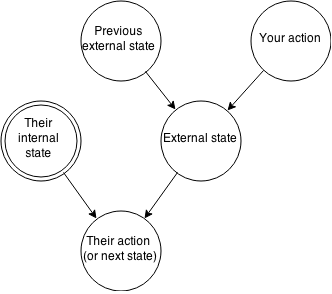
[For preliminary notes, see the document entitled “optimal communication”]

Before we develop a formal approach, let us explore the nature of coordination more precisely.

[Pylyshyn style black box except analyzed with the tools of computational learning theory.]

Contingencies in an active manner are similar to active sensing. We sense the result of the changes we introduce to the world or to another mind. In vision, these changes are introduced to the perceived world by changing our relationship to the world, in echolocation, they are produced by sending a probe of sorts into the world, and in language there is a similar pattern.

Introduce the causal graph here to illustrate the information flow in the contingency process.



This generalized causal model captures the effect of your action on the internal state and the effect of the external state and the partner’s internal state on the partner’s action. In this way it illustrates the causal pathways involved for active contingency inference, passive contingency inference, and inference of continued processes over time for a turn-taking discourse.

## Establishing conventions and common ground

## Generalization and inference

* For dealing with scenarios, states, and actions, not before seen
* For dealing with compositionality
* For learning the roles of components in a compositional system
* For inferring the role of the unobservable internal state in the contingencies
  + This requires both similarity and generalization
    - Similarity works through generalization. We notice similarities between ourselves and the other and we therefore can use these similarities to infer that there exist other unconformable similarities in the relationship between internal state and external state-action contingencies. Though these similarities are unconformable, we can gain information about them through interdependency-type contingency analyses. Of course, there is always the possibility of exchangeability with respect to the internal state. A paradigm case of this is color phenomena.

Show that inference is dependent on hypotheses and priors

Demonstrate that hypotheses priors are instantiations of generalization bias

Explore where this generalization bias comes from and what determines the correct generalization bias.

Just as sensorimotor contingencies serve as predictive mechanisms and are learned in a general way, so are signal-response contingencies in intersubjective communication.

We can take the black box thought experiment from the earlier section and add generalization to it and therefore show that there must be similar generalization biases.

It is important to EMPHASIZE in this work that we are not making any assumptions at all about the nature of the problem in the model. This enables us to use this as a theory about how humans can model these interactions without leaning on things that we already know about the structure of these processes. If we build our knowledge of these processes into our model of how we model these processes, then we are not actually capturing how we come to this knowledge in the first place. Though a highly abstract theoretical model such as this will not make specific predictions as much as provide a framework for the theory, it is highly principled in that it acts more as a philosophical derivation than an engineering-style model, the later of which is increasingly prevalent in the cognitive science field. In essence, we have made a decision to favor theoretical purity over modeling detail in the hope that we can provide a principled approach to the study of human communication.