Learning, Feedback and Information in Self-Organizing Communication Systems

Matthew Spike (matthew.spike@ed.ac.uk)
Kevin Stadler (kevin.stadler@ed.ac.uk)
Simon Kirby (simon@ling.ed.ac.uk)
Kenny Smith (kenny@ling.ed.ac.uk)

Language Evolution and Computation Research Unit, School of Philosophy, Psychology & Language Sciences
University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh, EH8 9AD, UK

Abstract

Communication systems reliably self-organize in populations of interacting agents under certain conditions. The various fields which model this – game theory, cognitive science and evolutionary linguistics – make different assumptions about the learning and behavioral processes which are responsible. We created an exemplar-based framework to directly compare these approaches by reproducing previously published models. Results show that a number of mechanisms are shared by the systems which can construct optimal communication. Three general factors are then proposed to underlie any self-organized learned system.

Keywords: cultural evolution; communication; self-organization; reinforcement learning; feedback learning; observational learning

Introduction

Human communication is a mostly learned behavior, while signaling behavior in the natural world appears to have a major genetic component. While Darwinian natural selection is argued to be the driving force behind the development of such innate capacities (e.g. Scott-Phillips et al., 2012 and Oliphant, 1996), the origin of learned communication is less clear. Effective communication requires consensus within a population; how is this reached given the arbitrary mapping between signal and meaning? In the absence of external or internal guidance, the emergent agreement must be the effect of not just global factors, such as how populations are connected and change over time, but crucially local ones also, for example how individuals learn and interact. Population-level behavior can therefore provide insights into aspects of human cognition.

The problem of self-organization of learned communication systems has been investigated by researchers working in game theory, artificial intelligence and evolutionary linguistics. The approaches taken by the different fields have much in common: all investigations focus on how two or more agents can effectively arrive at a mutually agreed set of conventions for a number of objects when agents are able to provide feedback to each other. Instead of observing a world state, speakers are said to randomly pick a topic from a communicative context. Key differences from the signaling game are that agents can indicate their intended referent in a communicative context. Lewis (1969) devised his classic signaling game in line with game-theoretic principles. A speaker’s signal triggers an action in the hearer: the resulting payoff, and thus reinforcement, depends on the state of the world, which is known only to the speaker. If the number of signals, acts and equiprobable states are all held at two, with equal non-conflicting payoffs, the game is proven to always converge upon an optimal signalling system (Beggs, 2005). Adjusting any of these parameters, however, quickly leads to pooling equilibria, where non-optimal communication strategies become attractors in the system. Barrett (2006) shows that while such sub-optimal situations will unavoidably occur when there are more than two possible states, systems can generally escape the pooling equilibria by enforcing memory limitations or including negative reinforcement (punishment of unsuccessful signals).

Steels’ 1998 seminal Talking Heads experiment gave rise to a plethora of naming games which investigate how static populations can converge on functional and efficient naming conventions for a number of objects when agents are able to provide feedback to each other. Instead of observing a world state, speakers are said to randomly pick a topic from a communicative context. Key differences from the signaling game are that agents can indicate their intended referent in the case of communicative failure in some ‘extra-linguistic’ manner (so-called corrective feedback), and that agents can introduce new signals (or names).

Such systems inevitably develop functional communication, but each object ends up with large number of synonyms, a result of the ability to innovate novel signals. By introducing competition between synonyms for the same object, the systems are driven into an efficient state where each object is known by only one label. De Vylder & Tuyls (2006) provide a mathematical proof that amplification of the input distribution of names is indeed sufficient to guarantee con-
vergence of the naming game. Agents that implement such amplification are said to employ lateral inhibition to dampen name competitors, the most well-known being Baronchelli et al. (2006)’s minimal strategy. Baronchelli (2010) shows that only the hearer need be modified for effective convergence.

Taking yet another approach, iterated learning is the collective term for a large number of computational and experimental studies which combine varieties of observational learning with intergenerational population turnovers (Kirby et al., 2008). Oliphant & Batali (1997) is one such example: their obverter strategy is derived from the mathematical result that if agents have perfect information about the internal state of the population, choosing signals by maximizing the chance of correct interpretation always results in the population converging on optimal communication. In simulations where agents use only incomplete information about the population gained through intergenerational learning, the obverter strategy still results in population convergence. In another study, Smith (2002) investigated the role of learning bias using populations of agents represented by Hebbian networks. Results showed that biases against homonymy and synonymy are necessary to produce optimal signaling.

The engine which drives the evolution of optimal signaling is variously stated: for reinforcement learning, it is communicative success; for the feedback models, it is the information gained through mutual alignment. Learning in the above models is horizontal; it takes place in static, closed groups. Intergenerational or vertical learning is employed by observational learners in iterated learning models which focus on individual learning biases, and obverters which stress the importance of explicitly maximizing the chance of being understood. A comparison of the above approaches leads to few clear conclusions regarding which learning and interaction features are responsible for convergence. Table 1 shows how the models contrast over many dimensions. The following section describes how the models were reproduced in a unified framework.

### Replications

An exemplar-style model was used to replicate the four models described above so that the effect of their different design features could be compared directly. Exemplar models have been employed to solve linguistic problems such as categorization (see e.g. Pierrehumbert, 2001). Learning involves storing packets of perceptual information with discrete category labels. Our framework represents each exemplar as a simple pairing between a signal and a meaning, where ‘signal’ can also be read as ‘name’, and ‘meaning’ is equivalent to both objects in naming games as well as world-states and actions from signalling games. When an agent maps a signal to a meaning, a single exemplar is stored. As such, the framework does not represent a fundamental departure from network and association weight models, but does suggest the simplification of aspects of these models in ways which are detailed below.

A stored exemplar is atomic, and can not be modified in any way apart from wholesale deletion. Production and interpretation of signals can be deterministic or stochastic. With stochastic methods (excepting obverters) the probabilities of producing or interpreting a signal s from a total S signals in association with meaning m from a total M meanings are given in Formula 1 below, where n_{i,j} represents an agent’s count of exemplars associating meaning i and signal j. Deterministic methods (also known as winner-take-all or WTA) always select the signal or meaning which yields the highest probability.

\[
P(s|m) = \frac{n_{ms}}{\sum_{i=1}^{S} n_{mi}} \quad \text{and} \quad P(m|s) = \frac{n_{ms}}{\sum_{j=1}^{M} n_{js}} \quad (1)
\]

Our framework is able to capture deterministic and stochastic behavior, as well as both static and changing populations, and the various manipulations of agents’ internal representations employed by each of the models discussed above. For the sake of comparison, some parameters are held constant throughout all simulations presented here: populations consist of 10 agents and there are 5 available signals and meanings, where each meaning is equally likely to be selected. Populations are unstructured, with any two agents equally likely to interact. For models using vertical learning, a single new agent is trained on the data of the existing population at each iteration. The new agent then replaces the oldest member of the population.

In closed groups without population turnover, two agents are picked at random from the group at each time step, with one designated the speaker and the other the hearer. After each interaction, the hearer is updated according to the particular rules of that model, specified below. When lateral inhibition of synonyms and/or homonyms is employed,

---

**Table 1: Model Comparison**

<table>
<thead>
<tr>
<th>Transmission model type</th>
<th>Barrett</th>
<th>Steels</th>
<th>Oliphant &amp; Batali</th>
<th>Smith</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modify hearer/speaker?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning features</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Production &amp; reception</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| horizontal mathematical                  |           |           |                    |          |
| H & S                                   |           |           |                    |          |
| mutual payoff                           |           |           |                    |          |
| forgetting/negative reinforcement       |           |           |                    |          |
| stochastic                              |           |           |                    |          |

| horizontal associative                  |           |           |                    |          |
| H & S                                   |           |           |                    |          |
| feedback                               |           |           |                    |          |
| inhibition                              |           |           |                    |          |
| deterministic                           |           |           |                    |          |
| stochastic                              |           |           |                    |          |
a newly stored exemplar results in the deletion of one randomly selected exemplar with competing signal/meaning associations.\textsuperscript{1} When a memory limit is included in a model, this is instantiated by enforcing a maximum number of $n$ stored exemplars per agent. When this is exceeded, one exemplar is selected at random for deletion.

Communicative success was measured analytically by looking at the outcome of all possible communicative interactions over the entire population after each time step. 10,000 individual simulations were run for each configuration of each replication, and the number of iterations taken for each to converge on optimal signaling over the population was recorded. The cumulative distribution of converged populations over time was then plotted, as seen in Figures 1–4.

1. The reinforcement models used by Skyrms and Barrett employ Roth-Erev learning (Roth & Erev, 1995), which maps exactly onto the exemplar model where behaviour is directly proportional to the relative frequency of memory tokens. When agents produce a signal for a given meaning, they do so by selecting stochastically from all stored exemplars associated with that meaning; interpretation is done similarly. Crucially, however, a new exemplar memory is only stored in the case of communicative success.\textsuperscript{2} Repli-
always chosen.  

\[ P(s|m) = \frac{P(m|s)}{\sum_{i=1}^{S} P(m|i)} \] (2)

The simulations showed that, for both WTA and stochastic production, populations would only converge on optimal signaling either in combination with continuous replacement of old agents (iterated learning), or when agents had a fixed memory capacity in static populations.

4. Smith’s (2002) network model contained a total of 81 possible ‘update rules’ determining how learning affects internal representations. The exemplar framework rendered most of these counter-intuitive, leaving only two parameters: whether adding a new exemplar would result in lateral inhibition of competing synonyms and/or homonyms (or neither). The replication confirmed Smith’s analysis: inhibition of homonyms alone results in the extermination of both homonymy and synonymy. The reverse is not true, however: inhibiting synonyms does not affect homonymy. Moreover, the time taken to converge when homonymy inhibition is employed is apparently unaffected by the presence of an anti-synonymy bias, or whether WTA or stochastic strategies were used, as shown in Figure 3. With the correct bias in place, however, observational learners proved able to construct optimal signaling in both static and iterated learning populations.

When the four main models are compared using only horizontal transmission in a static population as in Figure 4, the convergence time for the hearer-only feedback and observational models appear to have identical distributions, and memory-limited obverters perform similarly as well. Negative reinforcement models take a significantly longer time to converge. As such, the requirements for each model to converge appear to be:

1. **Reinforcement learning**: negative reinforcement or memory limitations

2. **Corrective feedback models**: either no possibility of homonymy, or inhibition of homonyms.

3. **Obverter learning**: either vertical learning or limited memory

4. **Observational learning**: inhibition of homonyms is required

**Comments**

Based on our comparative simulations, the following conclusions can be drawn:

1. Simple reinforcement on the basis of successful communication is an ineffective way of establishing conventional signaling systems, leading to either non-convergence or very long convergence times in comparison to the other models. However, a much faster convergence is ensured if any form of deletion from memory is implemented, the most effective one being targeted negative reinforcement.

2. Corrective feedback as instantiated in the Steels models includes very large name or signal spaces. As a result,
homonymy is either impossible or unlikely. Communicative success in this case is unsurprising: even if every agent innovates their own signal for each meaning, eventually all agents throughout the population will have heard this token and will be able to correctly interpret it. This results in highly redundant labeling systems. Inhibiting synonyms leads to the eventual adoption of one-to-one mappings throughout a population. When the available signal space is limited, however, homonymy becomes a problem. Without the lateral inhibition of homonyms, convergence is not a certainty.

3. Smith’s (2002) models and the simplified Steels & Loetzsch (2012) models have extremely similar behavior because on one level of analysis they are the same: while Smith’s observational learning ignores referential uncertainty, that uncertainty actually plays no role in the feedback model. With corrective feedback, the intended referent is either correctly understood or else communicated after failure. The speaker’s intended communication is known independently of communicative success in both models.

4. ‘Feedback’ has several interpretations. Corrective feedback is described in Steels & Loetzsch (2012): the speaker indicates its intended interpretation. Reinforcement learning involves another form of feedback, where the speaker (or the environment) simply confirms whether or not the hearer has correctly understood. In Baronchelli (2011) and Vogt & Coumans (2003), feedback is defined as when the hearer informs the speaker how it has interpreted the signal.

We propose that the different kinds of “feedback” might be better characterized by looking at how information flows between speaker and listener. Corrective feedback in naming games ensures that the speaker always provides complete information about how it associates a particular meaning with a signal by unambiguously providing both the signal and the intended referent in every interaction. This guaranteed transmission of information is a feature shared by the observational models presented above. In reinforcement models, that information is only transmitted to a hearer after correct interpretation. Information flow from the hearer back to the speaker, on the other hand, is not present in the observational models which exhibit purely vertical transmission. Baronchelli (2011) shows that this flow is in fact unnecessary for the naming game without homonymy; the replications of the previous section show that this is also the case with homonymy (see Figure 2).

Feedback from hearer to speaker is critical for reinforcement learning, as confirmation of communicative success requires this information. The lack of ambiguity in other models ensures success, and thus removes the need for knowledge about communicative success. The flow of information from speaker to hearer is common to all the above models. The role of any relevant feedback, then, is to allow this information to pass at least some of the time.

5. Basic reinforcement models utilize only the general positive feedback provided after successful communication. Negative reinforcement goes one step further by using information available after failed communication to determine what the likely internal state of the speaker is not, and this difference in information is sufficient to lead to ideal signaling. However, the reliably transmitted information in other models is not by itself enough to guarantee optimality. Some force must lead to competition between homonyms. For observational models and in the naming game, this is lateral inhibition through deletion. For obverters, it is implicit in the way production is biased towards the most successful homonym.

6. Functional communication arises when signals unequivocally map to single meanings. Models which do not actually delete competing homonyms, such as basic reinforcement and obverters, must employ some form of non-targeted deletion. These effects arise through either vertical learning (by wiping out parts of the ‘collective memory’ through the ongoing replacement of agents) or memory constraints on individual agents. Vertical learning leads to a process analogous to genetic drift: there is a chance that with every new generation some tokens will not be learned and thus lost, reducing the diversity of signals for any given meaning. Equally, limiting individual agents’ memory capacity by deleting surplus exemplars causes the relative proportions of competing tokens to be affected by a random walk. In both cases, however, the probability of a particular mapping undergoing total deletion is inversely proportional to its relative frequency. If the pressures exerted by basic reinforcement models or obverter production cause the majority of mappings to gravitate towards an optimal system, then random sampling is enough to remove all competitors and lead to one-to-one mappings.

What, then, are the crucial elements which determine whether a population will construct optimal signaling? The next section will discuss the underlying qualities shared by all models with this property.

Discussion

Reliable transmission of information between agents is not by itself enough to lead to the emergence of an optimal signaling system: there must be competition between homonyms, leading to a situation where each signal maps unambiguously to a single meaning. The opposite directionality of simultaneously strengthening signals in one meaning-space while decrementing them in another is a self-reinforcing, rich-get-richer process. Models which use lateral reinforcing and closed-group obverter models, this does not happen. While both processes contain an implicit bias against homonymy, without some
form of deletion this is not strong enough, leading to ambiguous states which are semi-stable. In the absence of deletion, the weight of stored exemplars serves both to preserve ambiguous mappings and inhibit moves towards optimality. Deletion can be either active, such as in negative reinforcement, or it can arise through passive processes of random memory deletion or intergenerational sampling.

The factors, then, which determine whether a population will reliably construct optimal signaling are:

1. Speakers have to convey information – at least some of the time – about how they associate signals and meanings.
2. Information associating a signal to a meaning must bias the receiver against associations with other meanings.
3. Information must be lost: this may be via deletion, forgetting or intergenerational sampling.

In reinforcement learning, information rewards communicative success and optionally punishes failure. The information provides an inherent bias against homonymy. Similarly, the same bias is packaged into obverter production, which maximizes the chance of successful comprehension. In observational and feedback models on the other hand, the lateral inhibition of homonyms encapsulates both the bias and the deletion.

**Conclusion**

Self-organization of learned communication systems results from both individual and population-level behavior as well as their interactions. This generality explains the seemingly opposed interpretations and conclusions seen in modeling approaches: the relevant factors that guarantee convergence can be implemented in many ways. In fact, all of the proposed models may be partially accountable for the emergence of shared communication systems in humans. This has implications for both modeling and experimental approaches. When a certain set of conditions leads to a system of agreed signaling conventions, those conditions cannot be assumed to be the sole cause of the phenomenon. Instead, the conditions may simply fulfill the necessary requirements outlined above.

**References**


