Towards Explaining the Evolution of Domain Languages with Cognitive Simulation

David Reitter (reitter@cmu.edu) and **Christian Lebiere**
Carnegie Mellon University
Department of Psychology, 5000 Forbes Avenue
Pittsburgh, PA 15213 USA

Abstract

We simulate the evolution of a domain language in small speaker communities. Data from experiments (Garrod et al., 2007; Fay et al., 2008) show that human communicators can evolve graphical languages quickly in a constrained task (Pictionary), and that communities converge towards a common language even in the absence of feedback about the success of each communication. We postulate that simulations of such horizontal evolution have to take into account properties of human memory (cue-based retrieval, learning, decay). We implement a model that can draw abstract concepts through sets of non-abstract, related concepts, and recognize such drawings. The knowledge base is a network with association strengths randomly sampled from a natural distribution found in a text corpus; it is a mixture of knowledge shared between agents and individual knowledge. In three experiments, we show that the agent communities converge, but that initial convergence is stronger when communities are structured so that the same pairs of agents interact throughout. Convergence is weaker in communities when agents do not swap roles (between recognizing and drawing), predicting the necessity of bi-directional communication in domain language evolution. Average and ultimate recognition performance depends on how much of the knowledge agents share initially.

Keywords: Alignment; Language Evolution; Domain Languages; Microevolution; Cognitive Architectures, Multi-Agent Simulation

Introduction

Languages evolve: like biological systems, they undergo mutation and selection as they are passed on between speakers and generations. Similar to its biological counterpart, human communication evolves under environmental constraints. Fitness of a communication device (software) is a function also of the cognitive hardware: cognitive facilities constrain the language system. In this paper, we use an independently motivated cognitive memory architecture to constrain an evolutionary process that produces a communication system.

Recent models of dialogue describe how interlocutors develop representation systems in order to communicate; such systems can, for instance, be observed using referring expressions that identify locations in a maze. Experiments have shown that referring expressions converge on a common standard (Garrod & Doherty, 1994). Pickering & Garrod's (2004) Interactive Alignment Model suggests that explicit negotiation and separate models of the interlocutor's mental state aren't necessary, as long as each speaker tends to adapt to themselves and their interlocutors, as they are known to do on even simple, linguistic levels (lexical, syntactic).

Some evolutionary models (vertical models) see the transmission of cultural information as a directed process, in which information is passed only from the older to the younger generation. Horizontal models explain the emergence of language as a continuous process within generations. *Individualistic* models of language evolution assume that innate learning and processing systems set a prior, towards which language converges. Interaction and the cultural environment do not leave marks in the resulting language. *Collaborative* models, on the other hand, accept that language mutates and converges within generations as well. They claim that meaning-symbol connections spread between collaborating agents and ultimately converge on a predominant one. It is the dichotomy between individual and community-based learning that motivated the experiments by Garrod et al. (2007) and Fay et al. (in prep.), which serve as the basis for the model presented here.

In the horizontal society of cognitive agents in our study, agents adapt their communication system collaboratively to environmentally shaped and cognitively constrained needs of each individual. With our model, we aim to use a cognitive framework – specifically a memory model – to reflect processes in the individual that give rise to emergent convergence and learning within the community. By this, we acknowledge the fact that cultural evolution is constrained by individual learning; each agent learns according to their cognitive faculty (cf., Christiansen & Chater, 2008). The possibility of cultural language evolution has been supported by computational simulations (e.g., Kirby & Hurford, 2002; Brighton et al., 2005).

It is because adaptation according to experience is determined by human learning behavior that simulation in validated learning frameworks is crucial. Griffiths & Kalish (2007) for instance model language evolution among rational learners in a Bayesian framework; the purpose of the present project is to simulate the evolution of a communication system using an architecture with an accurate account of memory access and a concrete experimental design. We will introduce a cognitive model that simulates a participant in the experiment; multiple models interact as a community of participants. The purpose of this paper is to observe how a compositional language system is created between collaborating agents in a computational, cognitive simulation. We will show that the model demonstrates learning behavior similar to the empirical data. We assume these agents share a common reference system initially, display cooperative behavior and adopt mixed roles as communicators. Therefore, we explore different scenarios that test the necessity of our preconditions, in particular the initial common ground and the fact that each agent can be both on the sending and the receiving end of the communications.

The Task

The Pictionary experiment (Garrod et al., 2007) involves two participants, a *director*, who is to draw a given meaning from a list of concepts known to both participants, and a *matcher*, who is to guess the meaning. Director and matcher do not communicate other than through the drawing shared via screens of networked computers; the matcher is able to draw as well, for instance to request clarification of a part of the picture. Each trial ends when the matcher decides to guess a concept. Garrod et al.'s set of concepts is divided into five broad categories (e.g., actor, building); the concepts within each are easily confusable (e.g., drama, soap opera). Each game involves several trials, one for each concept on the list, in randomized order. The director is not informed of the guess made by the matcher, and neither participant receives feedback about whether the guess was correct. Participants switch roles after each trial. Participants to play many games so that the emergence of consistent drawings can be observed.

We implement the experiment in a form applied by Fay et al. (in prep., 2008), where 16 concepts (plus 4 additional distractors) were used in a design with two conditions. In the *isolated pair* condition, participants were split into fixed pairs. They played seven rounds of six games each with the same partner. In the *community* condition, participants changed partners after each round. Each community consisted of eight participants. The pattern of pairings was designed so that after the first round, four sub-communities existed, after the second round, two sub-communities. After round four, the largest separation between partners was 2 (i.e., each agent has interacted via another one with every other agent); it was 1 after round seven. Fay et al. evaluated the iconicity of drawings, showing that isolated pairs developed more idiosyncratic signs, while the signs emerging within communities were more metaphoric (i.e. deducible) and easier to understand for new (fictitious) members of the language community. As idiosyncracy increases with each drawing-recognition cycle, but resets (to some degree) when communication partners change, communities may end up evolving similar idiosyncracy once every pair of participants played the same number of games.

The simplest measure and the one crucial for the evaluation of models like ours is *identification accuracy*. Fay et al. found that their participants generally converged quickly to a common meaning system. Convergence reached a ceiling of around 95% in both community and isolated-pair conditions. Changing interaction partners from round to round, as in the community condition, reduced accuracy during the initial changes; however, the community reached good ID accuracy after just a few rounds. We will use the development of ID accuracy as one way to evaluate the model.

The Model

ACT-R (Anderson, 2007) is an architecture for specifying cognitive models, one of whose major components is memory. ACT-R's memory associates symbolic chunks of information (sets of feature-value pairs) with subsymbolic, activation values. Learning occurs through the creation of such a chunk, which is then reinforced through repeated presentation, and forgotten through decay over time. The symbolic information stored in chunks is available for explicit reasoning, while the subsymbolic information moderates retrieval, both in speed and in retrieval probability. The assumption of rationality in ACT-R implies that retrievability is governed by the expectation to make use of a piece of information at a later point. Important to our application, retrieval is further aided by contextual cues. When other chunks are in use (e.g., *parliament*), they support the retrieval of related chunks (*building*).

A single ACT-R model implements the *director* and *matcher* roles. As a director, the model establishes new combinations of drawings for given target concepts. As a matcher, the model makes guesses. In each role, the model revises its internal mappings between drawings and target concepts. Table 1 gives an example of the process. The model is copied to instantiate a community of 64 agents, reflecting the subjects that took part in the Pictionary experiments.

Our model uses a scalable and efficient re-implementation of ACT-R called *ACT-UP*, letting us underspecify model elements such as the production-rule system, which would neither introduce nondeterminism nor carry explanatory weight in this particular model.

Maintaining a communication system

The simplest form of keeping a communication system in ACT-R memory *chunks* is a set of signs. Each sign pairs a concept with a set of drawings. Competing signs can be used to assign multiple drawings for one concept, this would create *synonyms*; multiple concepts can also combine with the same drawings, creating *homonyms* and ambiguity.

To create new concepts, we need to introduce a subsymbolic notion of relatedness. We use ACT-R's spreading activation mechanism and weights between concepts to reflect relatedness. Spreading activation facilitates retrieval of a chunk if the current context offers cues related to the chunk. Relatedness is expressed as a value in log-odds space (S_{ji} *values*).

When the model is faced with the task to draw a given concept such as *Russell Crowe* (one of the concepts in the experiment) that has no canonical form as a drawing, a related but drawable concept (*drawing*) is retrieved from declarative memory. Similarly, we request two more concepts, deferring any desire of the communicator to come up with a distinctive rather than just fitting depiction of the target concept. The case of a model recognizing a novel combination of drawings is similar; we retrieve the concept using the drawings as cues that spread activation, making the target concept the one that is the most related one to the drawings.

After drawing or recognizing, the target or guessed concept, along with the component drawings, is stored symbolically in memory as a chunk for later reuse (*domain sign*). These signs differ from the pre-existing concepts in the network, although they also allow for the retrieval of suitable

Director	Matcher
Fails to retrieve domain sign for A. Retrieves related concept: ⇒ component drawings 123 Draws components 1, 2, and 3 Learns domain sign A-123	 Requests related concept with cues 123 ⇒ concept B Guesses B Learns domain sign B-123
Retrieves domain sign for target concept B ⇒ component drawings 345 Verifies that B is retrieved when drawings 345 are activated Draws components 3, 4 and 5 Learns domain sign B-345	 Requests related concept with cues 345 ⇒ concept B Guesses B Verification: Requests domain sign for B ⇒ domain concept B-123 345 spread more activation to B than do 123, thus, learns domain sign B-345

Table 1: A protocol of two model instantiations, first failing to communicate concept A through three related drawings 1, 2 and 3, then successfully communicating concept B via drawings 3,4 and 5. The Matcher first adopts B-123 as a domain sign, then revises it to B-345.

drawings given a concept, and for a concept given some drawings. When drawing or recognizing at a later stage, the memorized domain signs are preferred as a strategy over the retrieval of related concepts. The system of domain signs encodes what is agreed upon as a language system between two communicators; they will be reused readily during drawing when interacting with a new partner, but they will be of only limited use when attempting to recognize a drawing combination that adheres to somebody else's independently developed communication system.

Knowledge

Agents start out with shared world knowledge. This is expressed as a network of concepts, connected by weighted links (S_{ji}). The distribution of link strengths is important in this context, as it determines how easily we can find drawing combinations that reliably express target concepts. Thus, the S_{ji} were sampled randomly from an empirical distribution: log-odds derived from the frequencies of collocations found in text corpus data. In a corpus comprising several years worth of articles that appeared in the *Wall Street Journal*, we extracted and counted pairs of nouns that co-occurred in the same sentence (e.g., "market", "plunge"). As expected, the frequencies of such collocations are distributed according to a power law. We found that the empirical log-odds resulting from these that form $S_{ji} = \log(P(J|I)/P(J|notI))$ (Anderson, 1993) (J and I being the events that J and I appear) can be approximated by a Generalized Inverse Gaussian-Poisson distribution (given in Baayen, 2001).

Such knowledge is, however, not fully shared between agents. Each agent has their own knowledge network resulting from life experience. This difference is essential to the difficulty of the task: if all agents came to the same conclusions about the strongest representation of target concepts, there would be little need to establish the domain language. We control the noise applied to the link strengths between concepts j and i for agent M (S_{Mji}) by combining the common ground S_{ji} (shared between all agents) with a random sample N_{Mji} in a mixture model: $S_{Mji} = (1-n)S_{ji} + nN_{Mji}$. Then, n [0;1] sets the proportion of noise. For Experiments 1 and 2, the noise coefficient is set to 0.2.

Adaptation pressure

Notably, participants in the experiment converged to a common sign system fairly quickly. This happened even though there was no evident, strong pressure to do so. Agents received no explicit feedback about the quality of their guesses or drawings. The only weak clue to the success of a set of drawings was whether the partner made a guess quickly. A helpful strategy for the matcher is to assume consistency between matching and drawing.

Invariably, the model will mistake a set of drawings for a reference to the wrong target. Lacking a feedback loop in this experiment, the model has no choice but to acquire even flawed domain signs and boost their activation upon repetition. Under these conditions, there is little pressure to converge. It is difficult to see how interaction partners could ever agree on a working communication system, given that there is no benefit for a model in choosing the concept-drawing associations of its interaction partner. However, the model does leverage *consistency* as proposed in Grice's maxims of manner, "Avoid ambiguity" and "Avoid obscurity of expression" (Grice, 1975). To do so, it assumes that a given set of drawings is associated with only one target concept, and, conversely, that a given target concept is associated with only three drawings. Suppose, for example (Table 1), that the model associates concept B with drawings $1,2,3$ (short: B-123). Later on, it comes across drawings $3,4,5$ as another good way to express B. In fact $3,4,5$ serve as convincingly stronger cues to retrieve B than do $1,2,3$. Thus, the model not

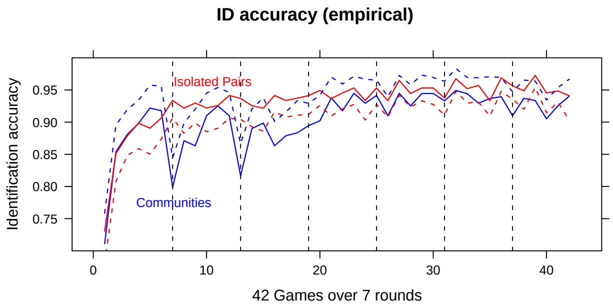

Figure 1: Identification accuracy for isolated pairs and communities (human data) as provided by Fay (p.c.). One-tailed 95% confidence intervals are given (upper bounds for communities, lower bounds for pairs), based on standard error (normality assumption).

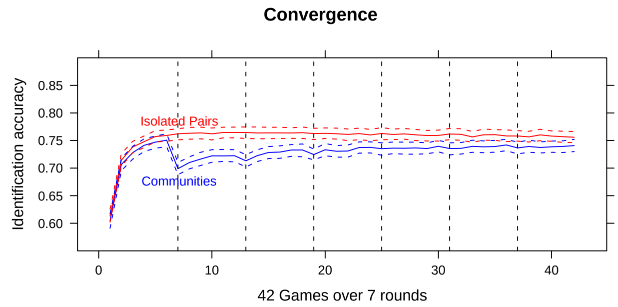

Figure 2: Mean identification accuracy in model simulations: As in the human data, both community pairs and isolated pairs gain most of their ID accuracy in the first game, but community pairs lose much accuracy when switching partners. 95% C.I., bootstrapped. 100 runs.

only correctly recognized B, but also learns the new preferred combination B-345. In the following rounds, B-345 will likely shadow the alternative in a winner-take-all paradigm, since B-345 is newer than B-123 and, thus, has stronger activation due to activation decay (noise and reinforcement may keep B-123 as a winner for longer). The decay mechanism counteracts the creation of synonyms.

In evolving the domain language, the model will avoid creating homonyms as well. Suppose a concept C is to be drawn, and 345 are retrieved as closely related and highly active drawings. Here, the model attempts to verify that 345 cannot be understood as any other concept than C. As the most strongly active concept for 345 is B, these drawings are ruled out to express C. With this mechanism, the model is able to cheaply modify the system of signs without extensive reasoning about the optimal combination every time a concept is added.

Algorithm

Directing The model is given a target concept A to convey. It uses *domain signs* and general knowledge to decide about a sign. At the end, the composed concept is committed to declarative memory as a domain sign. Domain knowledge is explicitly accessible and overrides subsymbolically derived compositions. As a consequence, the model acts with consistency: once a combination has first been used to convey a concept, the model will be more likely to use it. The director proceeds with the following algorithm.

1. Attempt to retrieve a domain sign for A of form $A - \alpha\beta\gamma$. If successful, verify by retrieving a domain sign B for the same three drawings $\alpha\beta\gamma$ is retrieved ($B - \alpha\beta\gamma$). Only if $A = B$, accept the domain sign $A - \alpha\beta\gamma$ and continue with step 3; otherwise choose another domain sign.
2. If no acceptable domain sign is found, use subsymbolic knowledge to combine concepts to express related target meanings. Using the target meaning as cue, retrieve three drawings $\alpha\beta\gamma$. The most active drawings are retrieved preferentially.
3. Draw $\alpha\beta\gamma$.
4. Learn $A - \alpha\beta\gamma$ (ACT-R buffer clearing action, repeated multiple times during the drawing process).

Matching Recognizing a drawing takes place in a similar fashion: domain knowledge is preferred over associative guesses. The model is given three drawings $\alpha\beta\gamma$. It proceeds with the following algorithm.

1. Attempt to retrieve a domain sign for $\alpha\beta\gamma$, resulting in $C - \alpha\beta\gamma$. If successful, verify by retrieving a domain sign of form $C - \delta\epsilon\zeta$. Only if $\alpha, \beta, \gamma = \delta, \epsilon, \zeta$, accept the domain sign $C - \alpha\beta\gamma$ and continue with step 3.
2. If no acceptable domain sign is found, retrieve a concept C using cues $\alpha\beta\gamma$ (spreading activation).
3. Guess C.
4. Learn $C - \alpha\beta\gamma$ (ACT-R buffer clearing action, repeated multiple times during the drawing process, but less often than during directing.)

ACT-R memory parameters were set to values consistent with the literature (transient noise 0.2, base-level constant 1.0, base-level learning and spreading activation enabled, retrieval threshold 1.0).

Experiment 1: Learning and Convergence

In the first experiment, we evaluate whether the model exhibits similar learning and convergence behavior, and whether there are differences in learning between the isolated-pair and community condition, as observed in Fay et al.'s experiment. The model uses the same number of concepts, trials and simulated participants as in the experiment.

Results

As shown in Figure 2, the learning behavior differs in the two conditions. *Isolated pairs* and *Community pairs* show a learning effect, i.e. they converge in their communication systems. However, unlike isolated pairs, community pairs dis-

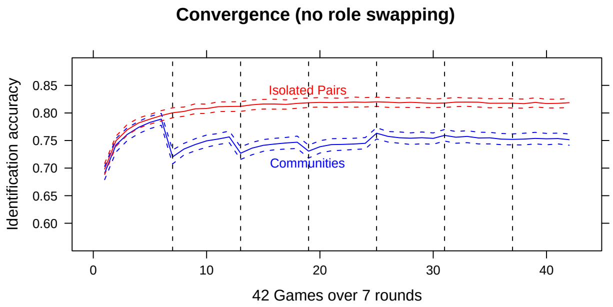

Figure 3: As in Figure 2, but without swapping roles.

play lower ID accuracy after the 7th game (game 1 of round 2), i.e. after switching partners.

We fitted a linear model to test some of the predictions more explicitly. The linear regression model treating round, game and condition (isolated pairs vs. communities) as independent variables, predicting log-transformed ID accuracy showed expected effects for round ($\beta = 0.03, p < 0.0001$) and game ($\beta = 0.02, p < 0.0005$), indicating improving accuracy with each game and round. An interaction of round and game ($\beta = -0.0046, p < 0.0005$) showed that the convergence leveled off in later rounds (as expected). There was no main effect of condition ($p = 0.45$), but an interaction of condition (isolated pairs) and round in the predicted direction ($\beta = -0.008, p < 0.05$), suggesting that convergence continued on for longer in the communities condition, and leveled off sooner in the isolated pairs condition. (All β in log space.)

Discussion

The results demonstrate, first, that agents converge both when retaining partners and when interacting with changing partners. Second, the results show that partner switching results in a setback in performance, but that agents continue to optimize their communication systems. This demonstrates that different dyads indeed converge on different signs for the same concepts. Notably, the setback appears to be smaller for rounds 3 through 7, i.e., through repeated partner switching, agents converge to a more common language.[1]

Overall, the model behaves similarly in many ways to the empirical data; however the initial and final accuracy achieved by the model is consistently lower than the approximately 70% and 95% accuracy (respectively) achieved by human subjects in the Pictionary experiments.

Experiment 2: Director and Matcher roles

Garrod et al. (2007) compared the performance of their participants in a comparable Pictionary task when a single director remained in that role throughout the experiment (single director, SD condition), vs. when participants swapped roles after each round (double director, DD condition). Identification accuracy was slightly higher for the role-swapping, double-director condition than in the single-director condition (significantly so only in the final rounds 5 and 6). This condition is similar to the *isolated pairs* condition in our model. Our model can not only simulate the role-swapping conditions, but also predict contrasts between isolated pairs and communities. The general question here is whether unidirectional communication would be sufficient to develop a community language. So, in this experiment, agents did not switch roles after every concept conveyed, i.e. they remained either director or matcher throughout the game. (Note that, unlike Fay et al.'s experiments and our simulation, Garrod et al.'s study involved feedback about the guesses.)

100 instances of Fay et al's experimental design were run.

Results

Identification accuracy for isolated pairs converged to a higher level than in Experiment 1. Interestingly, communities failed to achieve the same level of accuracy when director and matcher roles were not swapped (Figure 3).

Discussion

This experiment showed that turn-taking is essential for the development of a common community language. Isolated pairs benefit from uni-directional communication (as in Garrod et al's data), presumably converging towards the director's chosen language system. Communities are predicted by the model to require bi-directional communication to converge towards a similarly reliable communication system.

Experiment 3: Noise in Common Ground

A crucial assumption of the compositional semantics in this model is that the agents start out with common knowledge. For instance, both director and matcher need to accept that ambulances and buildings are strongly related to the concept *hospital*. However, the strength of the links between those concepts may differ without precluding the matcher from making the right inference.

The model allows us to test the importance of this assumption and predicts the results of a lower overlap between the knowledge bases of each agent.

Results

Figure 4 shows that mean identification accuracy (7th round, all games) decreases with increased levels of noise in the sub-symbolic knowledge state common to the agents. The model appears to deal reasonably well with noise levels of up to 0.3 (coefficient in the noise mixture) for both isolated pairs and communities configurations. This generally holds when taking all rounds into account. (At high noise levels, the initial acquisition of domain signs still works, but agents fail to converge further beyond the initial game or beyond a lower ceiling.) Further work should reveal whether further learning

[1] Note that Figure 2 suggests an effect of condition on the ceiling that is achieved; the regression analysis does not support this. We believe it is due to randomization of the concept order; further work is needed here. Note that in these initial experiments, we simulated only the same number of subjects and communities as in the experiments.

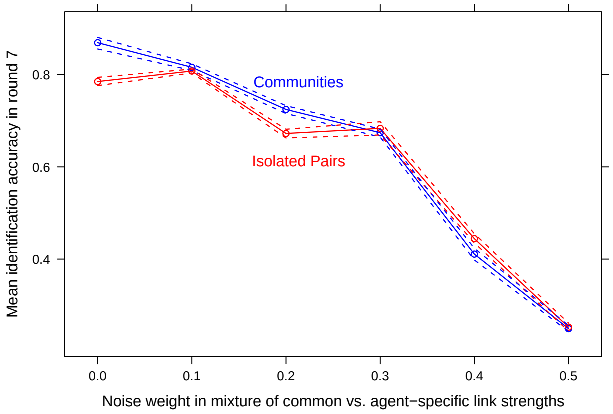

Figure 4: Mean identification accuracy at round 7 is reduced with noise between the knowledge bases of each agent. Bootstrapped 95% confidence intervals.

cycles can make up for the effect, i.e., medium noise levels lead to slower convergence and the failure to converge here is due to the limited number of games.

General Discussion

The model replicates several of the characteristics of the *communities* compared to the *isolated pairs* condition; specifically the set-backs after switching partners for the first few times and the ultimate convergence, despite very limited feedback. We also arrive at a clear prediction: bi-directionality is essential for linguistic convergence in communities.

At this point, we do not attempt to estimate optimal parameters in order to achieve a better fit to the empirical data. We believe that adaptation rates and the convergence ceiling depend both on the difficulty of the task, the specific materials (concepts) and the higher-level reasoning tools employed to optimize the language system. The task in Fay et al.'s experiment structured the list of concepts into a tree (e.g., there were four actors), making the job of drawing and guessing easier. Rather than just drawing what seems most closely related to the target concept, the experimental design invites them to choose a component concept that best disambiguates the drawing in the light of competing concepts (a head and a movie screen may be descriptive of Robert De Niro, but they do not distinguish him from Brad Pitt). Neither specific differentiation nor the precise choice of materials are modeled. Thus, we may overestimate the difficulty of the task. As a further simplifying assumption, our model always produced three component drawings before a guess is made. Garrod et al.'s (2007) design had participants give one another feedback about whether a drawing was thought to be recognized. However, our simplification is not expected to influence the character of the outcome.

Conclusion

We have demonstrated the use of validated, cognitively plausible constraints to explain an emergent, evolutionary group process via multi-agent simulation. Subsymbolic and symbolic learning within a validated human memory framework can account for rapid adaptation of communication between dyads and for the slower acquisition of a domain language in small speaker communities despite very limited feedback about the success of each interaction. Bi-directional communication is predicted to be necessary for a common language system to emerge from communities. The effects are robust against some divergence in prior common ground between agents.

Our model of the horizontal emergence of a common language in multi-agent communities is a first step to a computational, cognitive analysis of the learning processes involved in creating combined signs and acquiring links between them and arbitrary concepts, in order words, the evolution of language. Firm predictions can be drawn from this simulation only once robust convergence in much larger communities can be demonstrated, which will go beyond the empirical data that served as basis for this study.

Acknowledgments We thank Nicolas Fay and Simon Garrod for making their data and manuscripts available and Ion Juvina for comments. This work was funded by the Air Force Office of Scientific Research (FA95500810356).

References

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.
Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford, UK: Oxford University Press.
Baayen, R. H. (2001). *Word frequency distributions.* Dordrecht: Kluwer.
Brighton, H., Smith, K., & Kirby, S. (2005). Language as an evolutionary system. *Physics of Life Reviews*, *2*(3), 177-226.
Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(5), 489-509.
Fay, N., Garrod, S., & Roberts, L. (2008). The fitness and functionality of culturally evolved communication systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1509), 3553-3561.
Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (in prep.). *The interactive evolution of communication systems.* in preparation.
Garrod, S., & Doherty, G. M. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition*, *53*, 181-215.
Garrod, S., Fay, N., Lee, J., Oberlander, J., & Macleod, T. (2007). Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science*, *31*(6), 961–987.
Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech acts* (Vol. 3, p. 41-58). New York: Academic Press.
Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, *31*(3), 441-480.
Kirby, S., & Hurford, J. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (p. 121-148). London: Springer Verlag.
Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169-225.