

Arbitrary imitation, Pattern Completion and the Origin and Evolution of Human Communication

Monica Tamariz (monica@ling.ed.ac.uk)

Language Evolution and Computation, PPLS

The University of Edinburgh

3 Charles Street, Edinburgh EH8 9AD, UK

Abstract

Existing accounts of the origin of human communication assume a pre-existing behavioral system shared among members of a social group. This paper is concerned with the origin of that system; specifically, it explores its characteristics and functionality as well as the circumstances under which it could have appeared. A number of agent-based computer simulations test whether the capacities for arbitrary imitation and pattern completion can lead to a behavioral system that could be co-opted for communication. The results show that arbitrary imitation and pattern completion may indeed generate a population-wide shared behavioral system whose structure reflects the structure of the environment, and therefore could easily have been co-opted for communication. This system may have paved the way for other biological capacities widely believed to be necessary for communication, such as shared intentionality and symbolicity, to co-evolve.

Introduction

Much human socially transmitted culture arguably depends on arbitrarily copying learnt patterns whose function or origin is often unknown to the learner. This is illustrated by Gergely and Csibra's (2006) Sylvia's ham recipe story: Sylvia always cut the end of the ham when she cooked it because that is the way her mother did it; when the mother saw her do that, and asked why, Sylvia told her: "Because that is the way you always did it". The mother explained that her pan was too small to hold a whole ham, and that was why she had to cut off the end. Children engage in mindless imitation of elaborate actions even when these are irrelevant for the desired goal (Horner & Whiten, 2005; Lyons, Young & Keil, 2007; Whiten et al., 1996), or even for no apparent goal, as illustrated by the personal observation (which partly inspired this research) of a 24-month-old who, after seeing his mother clap her hands to try and catch moths on multiple occasions, interrupted his playing to clap when he saw a moth in the room, even if the moth was so far he could not possibly catch it – and

he did not even attempt to. Mindlessly imitated behavior, especially if it is not costly, may entrench the use of non-functional patterns in a community; this increases the degree of structure (reduces randomness) in the cultural environment, which therefore becomes easier to process and learn. This paper explores the role of arbitrary imitation in the origin of communication in our species.

Recent approaches to the evolution of human communication propose that humans evolved a unique biological adaptation, the socio-cognitive capacity for theory of mind (Tomasello, 1999) or symbolic reference (Deacon, 1997). This adaptation presupposes and relies on non-communicative behavior being already in place (Tomasello et al. 1997), for instance the behavior sequences observed in chimpanzees resulting from ontogenetic ritualisation (Tomasello & Call, 1997). During the transition to communication, non-communicative behavior would come to be understood as a reflection of others' meanings or intentions. In this study I explore an alternative kind of pre-communication behavior system caused by the evolution of the general cognitive capacity to *decouple means from ends*, instantiated as the abstraction of form from function and the abstraction of behavior from the producer of behavior.

Arbitrary imitation for pattern completion

Imitation has been proposed as one capacity that preceded and may have afforded the evolution of communicative behavior (Donald, 1991; Zlatev, 2007). Tomasello (1999) claimed that only human cultural behavior involves true imitation, or replication of the *means* that another individual employed to obtain an end or function, as opposed to emulation, or achieving the same *function* as the other individual, regardless of the means employed. I focus on *arbitrary imitation*, or replication of behavior irrespective of whether it can be identifiable as a means to an end or not.

Possible motivations for arbitrary imitation include a conformity bias (Asch 1955) and a pattern completion (hereafter, PC) bias. I focus on the latter. PC is the activation of a complete representation upon exposure

to a partial representation and is invoked as the fundamental cognitive mechanism in approaches to learning and cognition such as Sign Gestalt Expectancy (Tolman, 1932) and connectionism (Bishop, 1995; Ripley, 1996).

Consider the following pattern as experienced by one individual: the recurrent correlation of a number of stimuli, including objects and individuals in the environment and behavior produced by oneself or by others. If the behavior component is missing from an (incomplete) instance of this pattern, the individual may satisfy the PC bias by producing the missing behavior, and thus complete the pattern. If the behavior produced is a copy of a behavior observed in another individual, and if the behavior itself has no other function than to complete a current pattern, we have an instance of arbitrary imitation for PC.

Arbitrary imitation for PC requires the two abstractions mentioned above: first, it requires decoupling behavior from its iconic or primary utility function. Here, behavior's functionality resides in being the missing bit that completes the current context pattern. Second, individuals must assume that behavior produced by oneself is equivalent to behavior produced by another individual, and that both are equally good completions of a pattern.

Hypothesis

This paper is concerned with the idea that arbitrary imitation for PC evolved in hominins prior to the appearance of communication. Specifically, I test the hypothesis that *arbitrary imitation for PC in a population can result in a behavioral system that both (a) reflects the structure of the population's environment, and (b) is shared by all members of the population, in the absence of any intention or expectation of communication.*

Behaviors are cumulatively perceived and processed by individuals in the community. We assume that individuals in a social group are exposed to similar patterns; therefore imitation for PC should increase the level of structure in the environment by increasing the frequency of the imitated behaviors (through individual behavior "downloading" information onto the common environment, as in Clark, 1997). This more structured environment can then be more easily exploited by other individuals' arbitrary imitation for further PC activity. Moreover, this should result in an emergent social coordination of the associations or mappings between behaviors and other aspects of the environment among the individuals in the population. The only missing requirement for communication involves a dyadic social dimension, most likely related to theory of mind, that is absent from, but that could have evolved on top of, arbitrary imitation.

If arbitrary imitation for pattern completion had a foundational role in communication, and it is anecdotally observed today in examples such as Sylvia's story, it would suggest there is scope to further test the present-day role of this cognitive bias in the maintenance and shaping of communication systems such as language.

Methods

The hypothesis explained above is tested with a computer simulation experiment where agents learn about their environment and produce behavior. Unlike other models of language evolution (e.g. Kirby, 2002; Steels, 2002), the present simulations do not include communicative function or intention. Arbitrary imitation is implemented by letting the agents have access to ("be able to observe") the behavior productions of others. Behavior production is either guided by a heuristic that maximizes the systematicity of the agents' internal representation of the world (which optimizes the possibility of correct PC), or is produced randomly. Finally, the environment structure is also manipulated.

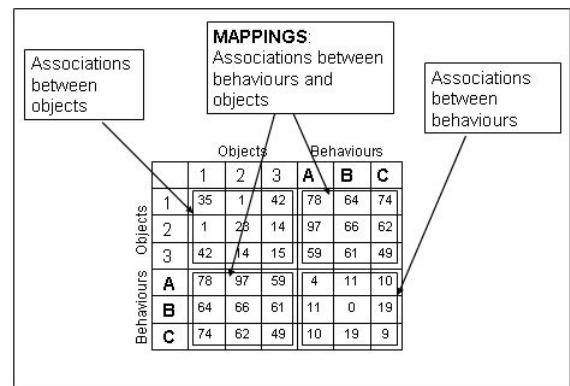


Figure 1. An agent's memory or cooccurrence matrix. In the simulation run that generated this matrix there were 3 objects and 3 behaviors; the complete matrix and the object and behavior matrix subsets are symmetrical; the two mapping matrix subset are not (they are mirror images of each other).

Agents do associative learning based on co-occurrence: the level of association between two items for an individual is proportional to the frequency with which they have co-occurred in the same context in the individual's experience. Their memory is a symmetric square matrix storing the cooccurrence counts between every pair of stimuli (see Fig. 1).

Each simulation consists of a number of interactions where the agents observe and interact with the environment thus:

1. A proportion of the agents in the population are randomly selected as participants (*observers*) in the current context; a proportion of the observers are randomly selected to be also *producers* of behavior.
2. A current environment is constructed by randomly selecting objects (which may be repeated) from the object set.
3. Each producer in turn observes the current environment and selects a behavior from its behavior repertoire applying either a PC heuristic or randomly (see below). After all producers have made their selections, the behaviors are added to the current context.
4. Each observer agent increases the cooccurrence frequency count between each and every other elements in the current context it has access to in its memory matrix.

Independent variables

As already mentioned, three independent variables are manipulated in the simulations: (1) whether there is imitation or not (2) whether behavior selection is guided by PC or is random and (3) the environment structure.

1. In the **imitation** conditions, agents can observe other agents' behaviors, and consider those behaviors as part of the current context. They therefore register the cooccurrence between their own and others' behaviors and the objects in the environment, as well as cooccurrences among objects. In the control (no imitation) condition, agents only register cooccurrence among objects and between their own behavioral productions and the objects.

2. In the **PC** conditions, producers select the behavior that is most likely to be activated in their memory matrix given the objects in the current context and given an expectation of regularity (expect the same object-object and object-behavior combinations to occur again and again). This is implemented by agents selecting the behavior that maximizes the systematicity of the memory matrix resulting from their own individual experience over successive contexts. The measure of systematicity (Syst) is based on *RegMap* (Tamariz & Smith, 2008), a formalization of Hebbian learning that quantifies the regularity of the mappings between two domains (e.g. between signals and meanings in a language). *RegMap* uses information theory's redundancy (Eqn. 1) and frequency.

$$(1) R = 1 - \frac{\sum_i p_i \log(p_i)}{\log(i)}$$

Redundancy is 1 minus the entropy (H), and measures the structure of a system, its departure from randomness or its predictability. *Syst* is measured as follows: For an object-behavior (O, B) pair, first a

matrix is created by computing the cooccurrence frequencies x_{ij} between each object i and each behavior j . *Syst* is then calculated for B given O (Eqn. 2) and for O given B (Eqn. 3), which are finally combined to obtain *Syst*(O, B) (the measure that is maximized in the PC heuristic) using Eqn. 4.

$$(2) Syst(B|O) = \frac{\sum_i R(x_{ij}) \sum_j (x_{ij})}{\sum_{ij} (x_{ij})}$$

$$(3) Syst(O|B) = \frac{\sum_j R(x_{ij}) \sum_i (x_{ij})}{\sum_{ij} (x_{ij})}$$

$$(4) Syst(O,B) = \sqrt{\text{RegMap}(O|B) \times \text{RegMap}(B|O)}$$

In the control (no PC) heuristic condition, speakers select a behavior at random from the behavior repertory.

3. Three conditions of **environment structure** are tested: In the *random environment*, the probability of an object appearing in the current context of an interaction is equal to that of, and independent of the appearance of, other objects. In the *frequency-based environment* condition, the object set has an exponential frequency distribution: $\forall x \in O \rightarrow p(x) = 2^x$. In the *dependency environment* condition, the presence of an object in the context is determined by the presence of another object. Context construction in the simulations below (context size = 8; four object types O = {1,2,3,4}) consists of random selection of one object o_i , and inclusion in the context of four tokens of o_i and four of o_{i+1} . In this way, each object o_i only ever co-occurs with object o_{i+1} .

Dependent variables

Two variables are measured at the end of each run of the simulation, based on the agent's memory matrices.

1. The first one is the level of **coordination of the mappings**, measuring the degree to which individuals in the population have reached similar mapping matrix subset states. This is quantified as the average of the coordination of the mapping matrix subset for each pair of agents in the population. For a pair of agents (A, B), the coordination of their mapping matrices is given in Eqn. 5. It equals the square root of the correlation (Pearson's r) between their mapping matrices multiplied by the average variance of the same mapping matrices.

$$(5) Coord(A,B) = \sqrt{(r(A,B)) \times \left(\frac{Var(A) + Var(B)}{2} \right)}$$

2. The second dependent variable is related to the potential of a behavioral system in the population to be co-opted for communication. Mappings that could be co-opted for communication need to be potentially apt

for one-to-one mappings for both production and comprehension, therefore the **systematicity for production and comprehension** values are measured using Eqns. 2 and 3 applied to the mapping subset of each agent's memory matrix. $Syst(B|O)$ relates to production and measures how confident a speaker can be, given its mapping matrix subset (its knowledge about object-signal mappings), about which behavior to select to express an object. $Syst(O|B)$ relates to comprehension and measures how confident an observer can be of which object a behavior refers to.

Results

The simulations are run with 10 agents. At each interaction, 6 observer agents are present, of which 3 are also producers; the context includes 8 object tokens selected from a repertory of 4 object types. The agents have a repertory of 4 behaviors to select from. Each simulation is run for 600 interactions (when the systems have reached stable values of the dependent variables), at the end of which the average coordination and systematicity of the agents' mapping subset of their memory matrices are recorded.

The experiment design exhaustively combines two imitation conditions (imitation and no imitation), two behavior selecting heuristics (PC and random) and three environment structures (random, frequency-structured and dependency-structured). 130 simulations were run with each factor combination of the $2 \times 2 \times 3$ design.

Coordination of the mappings in the population

The level of coordination of the mappings among the agents in the population at the end of the simulation runs was measured in all possible condition combinations (Fig. 2). Coordination is significantly positively affected both by imitation and PC, and there is a strong interaction between the two: two-factor ANOVA tests in every world structure returned $p < 0.001$ for each factors and for their interaction in all three environment structure conditions. Mapping coordination is also significantly affected by the environment structure, with one-way ANOVA tests run in all four imitation x pattern-completion combination conditions also returning $p < 0.001$.

No PC, i.e. random selection of behavior (top and bottom left plots), generates random mappings, which, naturally, tend to be different for each agent (low coordination). In the dependency and, especially, the frequency-structured environments, however, agents are more likely to converge on similar mappings. When a PC heuristic is in place alone (top right plot) coordination is increased: a non-random heuristic generates more non-random mapping matrices; because of this, the Pearson's r component of the coordination measure becomes less obscured by the variance factor

(Eqn. 5). Additionally, when agents interact with a structured environment, mappings are better coordinated.

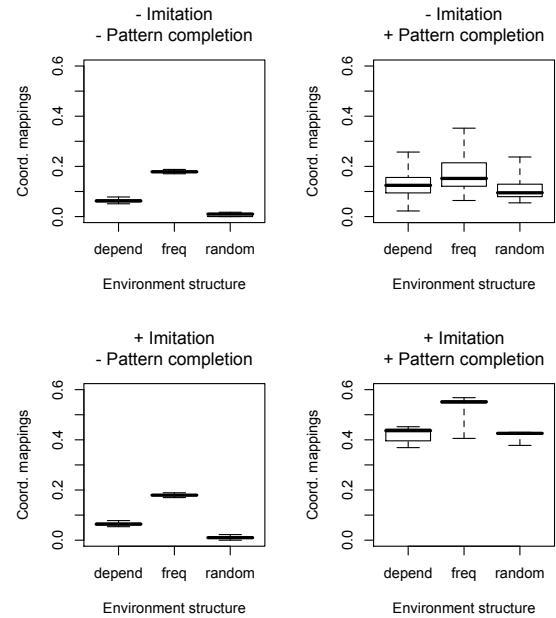


Figure 2. The degree of mapping coordination in the agent population in the 12 combinations of imitation x PC levels x environment structure. Each box-plot shows sample minimum, lower quartile, median, upper quartile and sample maximum.

When PC is combined with imitation (bottom right plot) coordination is much higher in the three environments. This is driven not only by heuristic and the initial environment structure, but also by the fact that the behavior that each agent produces (in a non-random way, following a PC heuristic) becomes part of the environment that is observed and learned by the population, thus increasing the level of structure in the environment.

The effect of environment structure is apparent in all four plots, with frequency structure returning the highest coordination values, followed by dependency structure and finally by random structure. In the frequency-structured environment, the proportions of objects in the interactions will be skewed, importantly, in the same direction for all agents. Dependency structure results in contexts that are more similar between agents than random structure, but because the dependencies are between the objects, each agent will experience a more unique frequency distribution of objects in the contexts it is exposed to than in the frequency-structured environment condition.

Systematicity of the mappings

If the mappings are to be co-optable for communication, they must be systematic both for production and for comprehension.

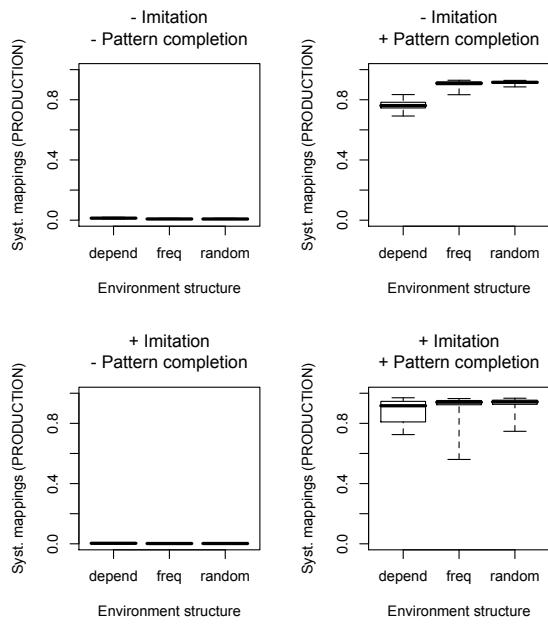


Figure 3. Systematicity of the mappings for the purposes of production in all condition combinations.

Fig. 3 shows that the PC heuristic strongly increases the systematicity of the agents' mappings for **production**: it all but eliminates the uncertainty for behavior selection for production. When arbitrary imitation is also present, mappings are even more highly systematic in all environments. In the random environment, production is always very unambiguous (each agent produces always the same unique behavior); in the frequency-structured environment, systematicity is more variable, but with high values; in the dependency-structured environment there is more ambiguity, and also a wider range of possible systematicity outcomes, with a clear bimodal distribution (not apparent in the figure): in half of the cases, a single behavior is produced every time, as in the random environment, but in the other half, distinct behaviors are produced for distinct objects. This indicates that PC strongly contributes to mapping systematicity for production in all environment structures.

The pattern is very different for systematicity in **comprehension** (Fig. 4). Most obviously, even the highest values (in the conditions including PC) are clearly lower than those for systematicity for production. This is due to the fact that agents in the PC conditions are maximizing the systematicity of their memory matrices by *producing* behavior; their productions accrue to the environment, which becomes

cumulatively more systematic *for production*. Unambiguous comprehension, on the other hand, is only possible if the environment is structured, and the agents cannot alter this in the present simulations. The results for systematicity for comprehension in the structured environments are nevertheless highly significantly different from those in the baseline random environment in all cases.

It is evident in Fig. 4 that the structure of the environment alone (even in the absence of imitation PC) confers a degree of unambiguous comprehension. PC only affects systematicity for comprehension in the dependency environment. In this environment, comprehension systematicity is consistently higher than for any other environment in the absence of imitation. When imitation is present, however, comprehension systematicity also has a bimodal distribution (not apparent in the figure): in about half of the runs, the heuristic makes no difference; in the other half, systematicity is much higher than without imitation, overlapping with that in the no imitation condition (top right), if statistically different from it: 2-tailed t -test, $t_{27} = 6.36$; $P < 0.001$.

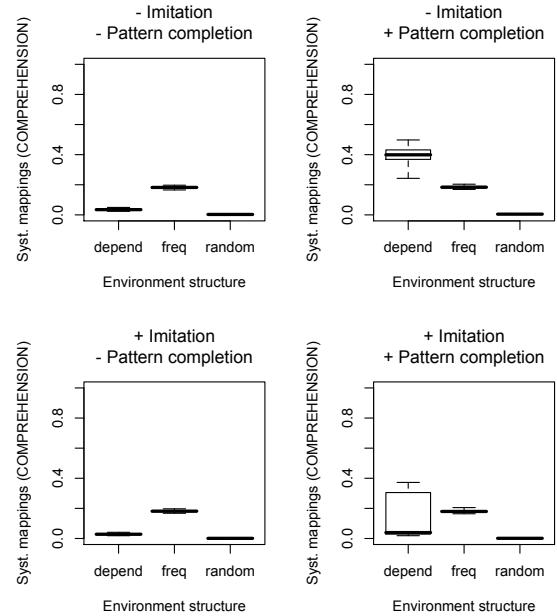


Figure 4. Systematicity of the mappings for the purposes of comprehension in all condition combinations.

Potential suitability for communication

The ideal behavioral system to be potentially co-optable for communication should have all of (1) high coordination, (2) high systematicity for production and (3) high systematicity for comprehension. These three requirements are only met in 50% of the runs in the +imitation, +PC, dependency-structured environment condition.

Discussion

Imitation and PC together enhance the coordination of the mappings between objects and behaviors in the population to a much higher degree than either separately. PC enhances systematicity for production because the agents simply seek to have a memory matrix that allows unambiguous selection of behavior for production, and are under no pressure to produce behaviors that will be understood by others.

The frequency-structured environment enhances systematicity for comprehension. Here, some objects are so frequent that the mapping matrix ends up reflecting the object frequency and a strategy of interpreting a behavior as referring to the most frequent object is better than chance. The dependency-structured environment also has an impact, in the cases where agents converge on the same systems (high coordination) and the association between each object and a unique signal is further enhanced over time by the agents' coordinated productions.

Arbitrary imitation improves systematicity for production, but seems to have negative effects for systematicity for comprehension in the dependency environment. However, in this condition we have, again, a bimodal distribution, including very low comprehension systematicity values, but also some very high ones, when two or three behaviors are produced, which can be unambiguously associated to specific objects, a crucial advantage if communication is to co-opt the system. These results indicates that the structure of the environment, especially if it includes dependencies between objects, can result, in half of the cases, in significantly-higher-than-chance systematicity of the mappings both for production and comprehension.

Therefore, this study suggests that the most propitious conditions for the emergence of a behavioral system that could be co-opted for communication include arbitrary imitation for PC and a dependency-structured environment. Moreover, this result supports the idea that a cognitive bias for imitation may have been instrumental for the origin of communication (Donald, 1971; Zlatev, 2007). Additionally, it proposes a mechanism whereby arbitrary or blind imitation (Gergely & Csibra, 2006) for pattern completion results in the coordination of cultural conventions in a population and thus can play a role not only in the origin but also in the ongoing spread and maintenance of symbolic cultural conventional systems including language.

Acknowledgments

This research was funded by a Leverhulme Early Career Fellowship. The author also acknowledges helpful comments on this work from Dan Dedić.

References

Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, 193: 31-35.

Bishop, C.M. (1995). *Neural networks for pattern recognition*. Oxford: Oxford University Press.

Clark, A. (1997). *Being there: Putting brain, body and world together again*. Cambridge, MA: MIT Press.

Deacon, T. (1997). *The symbolic species*. London: Allen Lane (The Penguin Press).

Donald, M. (1991). *Origins of the modern mind*. Cambridge, MA: Harvard University Press.

Gergely, G. & Csibra, G. (2006). Sylvia's recipe: The role of imitation and pedagogy in the transmission of human culture. In: N. J. Enfield & S. C. Levinson (Eds.), *Roots of Human Sociality*. Oxford: Berg.

Horner, V. & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Animal Cognition*, 8, 164-181.

Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In Briscoe, T. (Ed.), *Linguistic evolution through language acquisition*. Cambridge: Cambridge University Press.

Lyons, D.E., Young, A.G., & Keil, F.C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences*, 104: 19751-19756.

Ripley, B.D. (1996). *Pattern recognition and neural networks*. Cambridge: Cambridge University Press.

Shannon

Steels, L. (2002) The Evolution of Communication Systems by Adaptive Agents. *Adaptive Agents and Multi-Agents Systems* 2002: 125-140

Tamariz, M. & Smith, A.D.M. (2008). Quantifying the regularity of the mappings between signals and meanings. In A. D. M. Smith, K. Smith and R. Ferrer i Cancho (Eds.), *The Evolution of Language: Proceedings of the 7th International Conference on the Evolution of Language*, 315-322.

Tolman, E.C. (1932). *Purposive behavior in animals and men*. New York: Century.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M. & Call, J. (1997). *Primate cognition*. Oxford: Oxford University Press.

Tomasello, M., Call, J., & Gluckman, A. (1997). The comprehension of novel communicative signs by apes and human children. *Child Development*, 68, 1067-1081.

Whiten, A., Custance, D.M., Gomez, J.C., Teixidor, P. & Bard, K.A. (1996). Imitative learning of artificial fruit processing in children and chimpanzees. *Journal of Comparative Psychology*, 110(1): 3-14.

Zlatev, J. (2007). Embodiment, language, and mimesis. In T. Ziemke, J. Zlatev and R. Frank, (Eds.) *Body, language and mind: Vol 1: Embodiment*. Berlin: Mouton.