CREATIVE COMPOSITIONALITY FROM REINFORCEMENT LEARNING IN SIGNALING GAMES

MICHAEL FRANKE

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Amsterdam, 1098XH, The Netherlands
m.franke@uva.nl

Compositional language use shows in creatively associating hitherto unencountered meanings and forms in systematic ways. I submit that compositionality, as a key feature of human language, is no reason not to see a continuum between human speech and animal communication. Basic forms of compositional creativity presuppose surprisingly little cognitive sophistication. If changes in agents' behavioral dispositions are susceptible to similarities between different meanings and, independently, to similarities between different forms, creative compositionality can emerge in a signaling game model with reinforcement learning.

1. Introduction

A decisive step in the evolution of language was the transition from a holophrastic term language to a compositional language (Jackendoff, 1999). A holophrastic language consists of simple expressions that are individually meaningful, but are not combined in meaningful ways. In contrast, a compositional language has structured linguistic expressions which are built up from simpler individually meaningful parts. The meaning of a complex expression is related in a systematic way to the meaning of the parts that it comprises. Human language can be used holophrastically, but is compositional. Evidence for holophrastic communication in animals is known for long (c.f. Seyfarth, Cheney, & Marler, 1980). Animals also combine signals into sequences with novel meanings (c.f. Arnold & Zuberbühler, 2006; Ouattara, Lemasson, & Zuberbühler, 2009). Language-trained primates even creatively produce short sequences of meaningful elements to express new meanings (c.f. Marks Greenfield & Savage-Rumbaugh, 1990).

A compositional language has many advantages over a non-compositional one: it can convey more with less means, is therefore less susceptible to noise, can be learned from fewer examples, and much else. But in order to understand how the transition from a holophrastic to a compositional language might have been possible, it is unsatisfactory to simply point to a potential evolutionary advantage of compositionality once it is there (contra Nowak & Krakauer, 1999).

The relevant question is rather by what mechanism early forms of compositionality could have arisen in the context of a holophrastic system.

Many learning mechanisms are capable of linking a structured meaning space with a structured space of potential expressions, and so provide potential answers to the *how?*-question we are after (Batali, 1998; Kirby, 2002). It is good to know that it is possible for rather sophisticated agents to learn, and even generate, a compositional language. But once we know it, the key question becomes what the *minimal* cognitive abilities are that could lead to the transition in question.

Skyrms (2010) addresses this question in a game theoretic setting and suggests to see the beginning of compositional language in a model first introduced by Barrett (2007). The following paragraphs will introduce this model, together with the relevant background on signaling games. I proceed to argue that the Barrett-Skyrms model misses a key feature of compositionality, namely that it is a flexible and potentially creative ability to associate novel expressions with novel meanings. But rudimentary forms of creative compositionality do not presuppose much sophistication. Agents who perceive similarities between world states and (unrelated) similarities between signals can evolve a disposition to creatively exploit existing associations between states and signals. This can be demonstrated by a simple signaling game model using Roth-Erev reinforcement learning with two defensible amendments: (i) a *spill-over* mechanism that distributes accumulated rewards also to non-actualized contingencies proportional to how similar they are to the successful actual contingency (c.f. O'Connor, 2013), and (ii) a small amount of *lateral inhibition* (c.f. Steels, 1995).

Signaling Games. Signaling games were invented by Lewis (1969). In the simplest case, an unbiased random process selects one out of two possible world states. The sender knows the selected state, but the receiver does not. The sender sends one out of two signals. The receiver perceives the signal and chooses one out of two acts. If the chosen act matches the world state, the game is a success for both players, otherwise a failure. If the sender uses one signal consistently in the first state, and another in the other, and if the receiver chooses the appropriate act after each signal, sender and receiver will always play successfully. Such behavior of sender and receiver, as it were, bestows a meaning on the signals: each signal comes to be associated with a unique world state and its corresponding action.

Reinforcement Learning. Players could arrive at meaningful signal use in manifold ways. Much depends on their cognitive abilities. From an evolutionary perspective, and in line with the kind of methodological minimalism advocated above, it is interesting to assume that players are rather unsophisticated, incapable of rational decision making and possibly even unaware that they are playing a game. Basic forms of reinforcement learning (RL) are relevant for this purpose and have been studied well in this context. Basic RL assumes that each player keeps

an implicit record of the past successes associated with each action choice, the so-called *accumulated rewards*. The sender keeps a record for each state-signal pair, the receiver one for each signal-act pair. Whenever a play was successful, agents add a reward to the pair that was actually used. Initially, all accumulated rewards are 1. Accumulated rewards inform the agents' dispositions to act. In the simplest case, the probability that the sender selects signal m in state t is given by the *relative accumulated rewards* for that pair, i.e., the accumulated rewards for t and m divided by the sum of all accumulated rewards for t and all possible signals: $p(m \mid t) = {ar(t,m)}/{\sum_{m'} ar(t,m')}$. Similarly for the receiver.

Argiento et al. (2009) proved that this form of RL will eventually settle into a communicative constellation for the basic signaling game sketched above. If there are more states or signals, or if states are not equiprobable, things change. E.g., with three equiprobable states and three signals basic RL leads to a fully communicative state in ca. 95% of simulation runs (Barrett, 2007).

The Barrett-Skyrms Model. The meaningfulness that arises in Lewis-style signaling models is holophrastic. But a slight extension of the model raises hopes that very rudimentary forms of compositional meaning can also be traced in this way. Barrett (2007, 2009) studied signaling games with two signals, four states and corresponding actions. Instead of one sender, there are two. If each sender can send one out of two signals, it is possible to communicate exactly which world state obtains. Simulations of RL show that this situation almost always ensues. In the fully communicative situation, each sender's signal conveys one bit of information about which of the four world states is actual. The receiver puts the necessary information together and "infers" what the right choice of action is; or at least this is how it looks from the outside.

Skyrms (2010) suggests a variant of the multiple-sender model as a step towards understanding the origins of compositionality. The crucial observation is that the two-sender case is formally equivalent to a set-up with one sender who may send one out of two signals twice in sequence. Nothing else changes. RL still frequently leads to fully communicative signal use. But we have complex signals now, made up of two parts. Each part communicates one bit of information about the state. Skyrms therefore suggests to see here "a simple kind of compositionality" because "[t]he information in a complex signal is a *function* [my emphasis] of the information in its parts" (Skyrms, 2010).

I partly agree, partly disagree. Although we *can* describe the situation as one where the meaning of a complex signal is a function of its parts, there is no justification for doing so. A simpler description is that the receiver has simply learned to respond to four signals in the right way. Nothing hinges on the fact that these four signals are composed of two individual signals to us. The dispositions of the receiver to react to complex signals do not depend on their composition. Similarly, the sender has simply learned to emit one out of four signals that are implemented

as a bit-string of length two. There is no indication in the model that the agents have learned to apply a *function* of the meaning of the basic signals of which the complex signal is composed. They never actually use the simple signals. If they would, they might use them in ways unrelated to their "meaning contribution" to the complex signals. An explanation of basic forms of compositionality (or basic conjunctive inferences for that matter), requires an explanation of how agents acquire a *rule-like disposition* that shows when applied to novel stimuli. Only then is there a justification *within* the model to assume that they use the meaning of basic signals to arrive at the meaning of a complex signal.

Creativity and Spill-Over RL. Compositional signal use should show in creativity in the application of behaviorally acquired meaningfulness. Only then are we justified in describing behavior as following a functional combination of meaning. But creativity in this sense is at odds with basic RL. We would like to see whether RL-learners can be creative when confronted with a novel stimulus. But basic RL does not influence choice dispositions in non-actualized contingencies. So basic RL-learners will make uniform random choices in novel situations.

Variants of RL in which rewards are accumulated also for non-actualized contingencies exist (c.f. O'Connor, 2013). I submit that creative use of acquired meaningfulness is possible if rewards spill-over to non-actualized contingencies proportional to their similarity with the actualized one. Suppose that in state t signal m has led to payoff $x \geq 0$. (I focus on the sender from here on; the receiver part is analogous.) Basic RL adds x to the accumulated rewards $ar(t, m)$ only. In contrast, *spill-over* RL adds x to all $ar(t', m')$ in proportion to how similar the pair $\langle t', m' \rangle$ is to $\langle t, m \rangle$. Concretely, if similarity between pairs is a number between 0 and 1, then we add x times the similarity of $\langle t, m \rangle$ and $\langle t', m' \rangle$ to $ar(t', m')$.

Basic and spill-over RL differ in their assumptions about the learner's secondary-dispositions. Secondary dispositions are a learner's dispositions to change his (primary) dispositions to act given feedback about success or failure. Spill-over RL presupposes that agents' secondary dispositions are sensitive to similarity of choice-point/action pairs. Basic RL does so, too, but also assumes that all pairs are maximally distinct. Depending on what kind of stimuli and similarities are at stake, spill-over RL presupposes *less* cognitive sophistication. The spill-over may be due to an inability to distinguish sharply.

Model. The simplest non-trivial case where spill-over RL might lead to compositional creativity is a signaling game with six states and six signals. Three states and three signals are simple, three of each complex. If A and B are simple states (signals), then AB is a complex state (signal) built from A and B. Obvious ways of thinking about complex states and signals are meaning conjunction and signal sequencing, but other ways of combination are conceivable. We can remain entirely abstract here. Each state/signal has similarity 1 to itself. Similarity is a

symmetric notion, and complex state/signal AB bears a similarity $0 \leq s \leq 1$ to both A and B, and 0 to all others. The similarity of pair $\langle t, m\rangle$ and $\langle t', m'\rangle$ is the similarity of t and t' times the similarity of m and m'. If $s = 0$, we obtain basic RL. Notice that agents' secondary dispositions are sensitive to the similarity among states, and the similarity among signals, but not to similarities between a state and a signal, or even similarities between associating this signal with this state and that signal with that state. So, our assumptions about similarities do not smuggle in a particular cognitive ability at all; on the contrary.

This set-up promises to help explain spontaneous and creative compositional signal use. Suppose the sender has signals for states t_{A} and t_{B}, namely m_{A} and m_{B}. Suppose the sender is in state t_{AB} for the first time. t_{AB} bears marks of both t_{A} and t_{B} but is identical to neither. Perhaps, so the intuition goes, accumulated rewards of using m_{A} and m_{B} successfully in the past in states t_{A} and t_{B} have spilled over sufficiently to build a strong association between the hitherto unseen state t_{AB} and the hitherto unused signal m_{AB}. Maybe this association is strong enough for m_{AB} to be the most likely action choice in t_{AB}. This would then truly be a creative compositional use of a complex signal based on what its parts mean.

Lateral Inhibition. Unfortunately, not all values of s achieve this. The pair $\langle t_{\mathrm{AB}}, m_{\mathrm{A}}\rangle$ is at least as similar to $\langle t_{\mathrm{A}}, m_{\mathrm{A}}\rangle$ as $\langle t_{\mathrm{AB}}, m_{\mathrm{AB}}\rangle$ is. If $0 < s < 1/2$, the accumulated rewards of $\langle t_{\mathrm{AB}}, m_{\mathrm{AB}}\rangle$ will be strictly lower than those of $\langle t_{\mathrm{AB}}, m_{\mathrm{A}}\rangle$ (see below). Spill-over RL alone is not enough to evolve strong dispositions for creative compositionality for small s.

Things change if we add the possibility of *lateral inhibition*, which is another standard variation on classical RL models (Steels, 1995). If $0 \geq i \geq 1$ is a parameter for lateral inhibition, then, if $\langle t, m\rangle$ is part of a successful play, we subtract i from the sender's accumulated rewards for all $\langle t, m'\rangle$ and $\langle t', m\rangle$ (where $t \neq t'$ and $m \neq m'$), if the result is non-negative, and otherwise set the accumulated rewards to 0. (Likewise, for the receiver.) Positive i helps acquire a compositional language, although it is not strictly necessary either. It helps because, intuitively speaking, lateral inhibition does not affect the association of $\langle t_{\mathrm{AB}}, m_{\mathrm{AB}}\rangle$, but diminishes the association of $\langle t_{\mathrm{AB}}, m_{\mathrm{A}}\rangle$ and $\langle t_{\mathrm{AB}}, m_{\mathrm{B}}\rangle$.

Lateral inhibition is not an innocuous assumption. A positive i suggests that the agents tend towards one-to-one associations between choice points and actions. Some psychologists see evidence for a related tendency in language acquisition, the so-called *mutual exclusivity bias*: when learning new words children seem to assume that different objects must have different names and different names must refer to different objects (Clark, 2009).

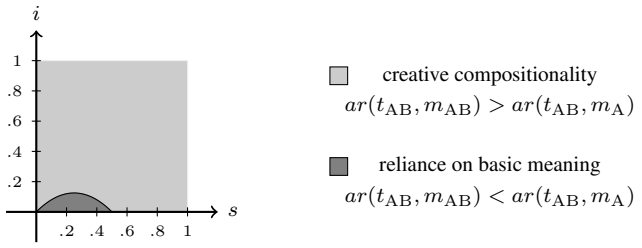

Figure 1. Comparison of accumulated rewards in the mean field limit when agents use a fully communicative language for simple states and signals. The plot shows the regions of the parameter space $i, s \in [0; 1]$ where the complex signal m_{AB} is the most likely choice in unfamiliar state t_{AB}.

Analysis. Suppose the sender uses m_{A} exclusively in state t_{A}, m_{B} in t_{B} and m_{C} in t_{C}. The mean field change in accumulated rewards is easy to calculate:

$$\dot{ar}(t_{\mathrm{AB}}, m_{\mathrm{A}}) = \dot{ar}(t_{\mathrm{AB}}, m_{\mathrm{B}}) = {}^1\!/_3 \max(s - i, 0) \qquad \dot{ar}(t_{\mathrm{AB}}, m_{\mathrm{C}}) = 0$$
$$\dot{ar}(t_{\mathrm{AB}}, m_{\mathrm{AC}}) = \dot{ar}(t_{\mathrm{AB}}, m_{\mathrm{BC}}) = {}^{s^2}\!/_3 \qquad \dot{ar}(t_{\mathrm{AB}}, m_{\mathrm{AB}}) = {}^{2s^2}\!/_3$$

In the mean field limit, the probability of the sender to choose the "compositional" message spontaneously converges to $p(m_{\mathrm{AB}} \mid t_{\mathrm{AB}}) = {}^{s^2}\!/_{2s^2 + \max(s-i,0)}$. If $i \geq s$, then this probability peaks at ${}^1\!/_2$. If $i > s - 2s^2$, the accumulated rewards of the creative compositional pairing $\langle t_{\mathrm{AB}}, m_{\mathrm{AB}} \rangle$ will be the highest for that choice point. The region of the parameter space where this holds is shown in Figure 1.

Simulations. Limit results are important, but do not inform us about short-term dynamics. Numerical simulations do. Figure 2 shows results from spill-over RL for different parameter values. Initially, accumulated rewards were 1. Agents first played with only simple states and signals for 10^4 rounds. In ca. 99% of trials a communicative code evolved. 100 of these were recorded for each parameter pair. Agents then continued to play with the full state and signal set. The plots show the average relative accumulated reward at choice point t_{XY} for options m_{XY} and m_{X}. With $X, Y \in \{A, B, C\}$ each plotted point is an average of 3 times 100 data points. We see that under favorable parameter values a dominant disposition for creative compositionality co-evolves quickly, together with the basic meaning association of simple states and simple signals.

Discussion. Even unsophisticated agents can acquire a disposition to creatively use signals in a new environment in a basic compositional way. The main ability needed for that is to have secondary dispositions that are suitably sensitive to any perceived similarity between states and any perceived similarity between signals. For rudimentary forms of compositionality, it is not necessary that agents, as it were, look for structural similarity between states and signals. If agents could track more information about similarities, they would presumably evolve more

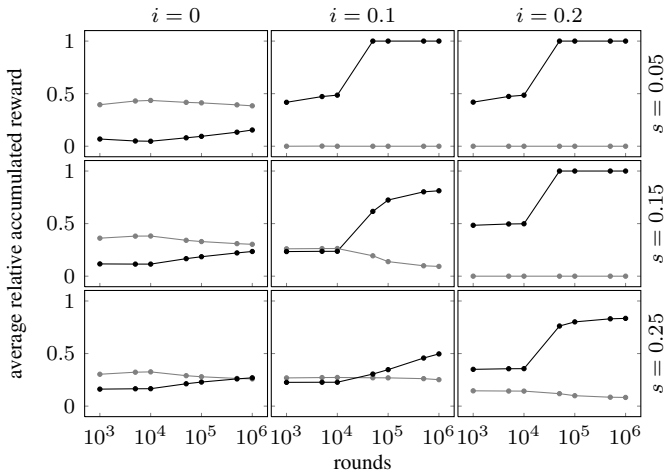

Figure 2. Results of numerical simulations of spill-over RL when there is no initial simple language in place. For different values of parameters the plots show average relative accumulated rewards in state t_{XY} for using m_{XY} (black) and m_{X} (gray).

elaborate compositional systems. It is tempting to speculate that a further step towards a human-like compositional language would involve recognizing similarities in the dynamically shifting patterns in the way a set of signals is used. But this is beyond the scope of this contribution, and irrelevant for a demonstration that basic forms of compositionality can arise already if agents merely track similarities among states and, independently, similarities among signals.

Evolving creative compositional dispositions is not a practical certainty. Only some parameter constellations readily allow for it. Low values for i and s seem most reasonable for unsophisticated agents. But it is then that creative compositionality is unlikely to evolve. This may explain why we have seen only little direct evidence of it in animal communication systems so far. Still, the model presented here makes clear that a continuous transition from holophrastic to compositional coding is possible.

It might be objected that spontaneous compositional language use, albeit it possibly the most likely choice, is never certain, i.e., $p(m_{\mathrm{AB}} \mid t_{\mathrm{AB}})$ is at most $1/2$. I believe that this is a good prediction that again excludes the presumably erroneous prediction that creative compositionality should be much more wide-spread in the animal kingdom than current evidence suggests. In sum, I propose that the model demonstrates that a path from holistic language to creative compositionality exists also for non-sophisticated RL-learners, but that the likelihood of finding this path is naturally upper-bounded.

References

Argiento, R., Pemantle, R., Skyrms, B., & Volkov, S. (2009). Learning to signal: Analysis of a micro-level reinforcement model. *Stochastic Processes and their Applications*, *119*, 373–390.

Arnold, K., & Zuberbühler, K. (2006). Language evolution: Semantic combinations in primate calls. *Nature*, *441*, 303.

Barrett, J. A. (2007). Dynamic partitioning and the conventionality of kinds. *Philosophy of Science*, *74*, 527–546.

Barrett, J. A. (2009). The evolution of coding in signaling games. *Theory and Decision*, *67*, 223–237.

Batali, J. (1998). Computational simulations of the emergence of grammar. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Evolution of language: Social and cognitive bases.* Cambridge, UK: Cambridge University Press.

Clark, E. V. (2009). Lexical meaning. In E. L. Bavin (Ed.), *Child language* (pp. 283–300). New York: Cambridge University Press.

Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences*, *3*(7), 272–279.

Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 173–204). Cambridge University Press.

Lewis, D. (1969). *Convention. a philosophical study.* Cambridge, MA: Harvard University Press.

Marks Greenfield, P., & Savage-Rumbaugh, E. (1990). Grammatical combination in *Pan Paniscus*: Process of learning and invention in the evolution and development of language. In S. Parker & K. Gibson (Eds.), *"Language" and intelligence in monkeys and apes* (pp. 540–578). Cambridge University Press.

Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *PNAS*, *96*, 8028–8033.

O'Connor, C. (2013). The evolution of vagueness. *Erkenntnis*. (To appear)

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys use affixation to alter call meaning. *PLoS ONE*, *4*(11), e7808.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, *210*(4471), 801–803.

Skyrms, B. (2010). *Signals: Evolution, learning, and information.* Oxford: Oxford University Press.

Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life*, *2*(3), 319–332.