Chapter 13
Emerging Symbols

Stefan Leijnen

Abstract Using a neural network simulation of a series of language training experiments with chimpanzees, the difference between indexical and symbolic interpretation is explored. From the results of the simulation follows a discussion about the systemic requirements for crossing the symbolic threshold and how the primacy of icons applies to computational models.

13.1 The Meanings of *Symbol*

In a study aiming to test the linguistic abilities of chimpanzees, several experiments are devised and conducted to demonstrate how different learning strategies produce different uses of language (Savage-Rumbaugh & Rumbaugh, 1978). The study shows how their learning curves can be understood from the way these chimps acquire language, allowing for a behavioral operationalization of language acquisition. The results are embedded within a larger semiotic theory of symbolic interpretation, distinguishing between three types of signs (icons, indices and symbols) that describe how an object can be related to a referent by an interpreter (Buchler, 1955; Hookway, 1985; Chandler, 2002).

Several other language training studies (Gardner & Gardner, 1977; Premack, 1976; Rumbaugh, 1977) show that apes can acquire large vocabularies. The subject has to point to one or more lexigrams on a board in order to express its thoughts or desires. Researchers stimulate the apes to use the correct lexigrams and apply appropriate grammar rules. However, even though their sentence construction capability can be trained to be more or less flawless, their learning strategy appears to differ from the way humans would approach such a problem. Although they appear to use lexigrams as representations of the objects they stand for (like humans do) their pointing behavior is a trained response to the presented stimulus.

S. Leijnen (✉)
Institute for Computing and Information Sciences, Radboud University Nijmegen, 6525 AJ Nijmegen, Netherlands
e-mail: stefan@leijnen.org

T. Schilhab et al. (eds.), *The Symbolic Species Evolved*, Biosemiotics 6,
DOI 10.1007/978-94-007-2336-8_13,

The often implicit assumption that these apes use lexigrams as representations for something else is not to be easily overlooked. For us to talk about apes using language and having a vocabulary, evidence is required that – indeed – these apes use *linguistic skills* to solve a problem, instead of *associative skills* to merely discover a correlation between stimuli and responses leading to a reward. The difference between these two skills is subtle but crucial, especially considering the principal reason for doing ape language studies is finding out if they are actually capable of learning a language.

So how are we to make this distinction clear? We find two contrasting definitions of symbols in which the difference is expressed (Deacon, 2003):

(S1) A symbol is one of a conventional set of tokens manipulated with respect to certain of its physical characteristics by a set of substitution, elimination, and combination rules, and which is arbitrarily correlated with some referent.

(S2) A symbol is one of a conventional set of tokens that marks a node in a complex web of interdependent referential relationships and specific reference is not obviously discernible from its token features. Its reference is often obscure, abstract, multifaceted, and cryptic, and tends to require considerable experience or training to interpret.

The chimpanzees in the Savage-Rumbaugh and Rumbaugh study are subjected to a training program that causes the disparity between these two kinds of symbols to become salient, demonstrated by a significant difference in performance results. In one experiment, the chimps learn to distinguish lexigrams for four objects (*banana*, *orange*, *coke* and *milk*) and two verbs (*give* and *pour*). The chimpanzees are required to use the correct verb with each noun by arranging them in a sentence. Producing accurate sentences like *give orange* or *pour milk* is rewarded; producing incorrect compounds like *pour banana* or *coke milk* is discouraged.

Once the chimps have learned to associate pairs correctly, a follow-up experiment shows that their symbol use is, in fact, non-symbolic. As the researchers introduce new edibles and liquids to the experiment, the amount of trials needed to learn to embed these words into sentences grows. Instead of using the web of relations to which the lexigrams refer – the chimps know that edibles are given and liquids are poured, but they don't apply this knowledge to the construction of lexigram sentences – they memorize each verb-noun correlation as a rule. The chimps use lexigrams as

> [...] a set of events which come to precede the receipt of a desired action or object. [...] errorless trials, though given in a fashion which closely approximates that of the final choice, do not lead to symbolic learning even in simple tasks such as food names (Savage-Rumbaugh & Rumbaugh, 1978).

The apes have learned to use symbols as defined by *S1*, but not according to the more strict definition *S2*. The relations between the lexigrams are arbitrary as the chimps fail to notice the analogy with the relations between objects and actions. *S1* is a rather shallow, computational definition of symbols that doesn't capture the way humans use symbols as expressed in *S2*. Hence, phrased in semiotic terms, the chimpanzees have learned to use lexigrams as indices. An index pairs two things together based on their co-appearance, like a thermometer (number and temperature) or a windsock (position and wind direction). In this case, a noun lexigram is paired with a verb lexigram.

For the ape subjects to use the lexigrams as symbols (according to *S2*) a reference is required to the network of relations for which the lexigrams stand. Evoking such a reference is exactly the goal of the next experiment in the chimp language training program. It is set up in almost the same way as in the previous ones, but this time the apes' attention is drawn towards the food and drink dispensers by increasing their saliency with light and sounds signals. The apes now notice the dispensers opening, also when they're empty. This causes some of the apes to pair their understanding of objects and actions with their understanding of lexigrams, and transfer knowledge between these networks. Instead of memorizing each and every lexigram combination as an index these chimps have created a symbolic link, which offers them a more efficient way of storing information in the long run.

13.2 Simulated Learning

The chimp language training research supports the claim that symbolism is not intrinsic to a word, lexigram or object, but is dependent on the interpretation itself. Interpreters can be iconic, indexical and symbolic, and some of the apes where capable of all three of these skills while others could only reach the indexical level. In order to explain this gap, it would be insightful to take a peek inside a chimp's head, study how signals travel between neurons and how eventually a lexigram sentence comes about. In a meticulous study of the chimp's interpretation process, the differences that cause the symbolic shift could be unveiled. Of course, the sheer complexity and size of the brain would result in far too many parameters for us to make sense of. As an alternative, computer simulated models of smaller, less complex brains can be used in order to discover the systemic requirements for symbolic interpretation.

For our experiments we will use an artificial neural network: a three-layer perceptron (McCulloch & Pitts, 1943) with full connectivity (Fig. 13.1). The nodes in the hidden and output layer are implemented with a step activation threshold function (1) (cf. Table. 13.1).

$$y_j = \begin{cases} 1 \; if \; \sum_{i=1}^{n} (w_i x_i) \geq \theta \\ 0 \; if \; \sum_{i=1}^{n} (w_i x_i) < \theta \end{cases} \tag{13.1}$$

By varying the connection weights between neurons different network architectures are generated, each with a potentially different behavior (i.e. returning a specific output in response to a certain input). After a set of random weight

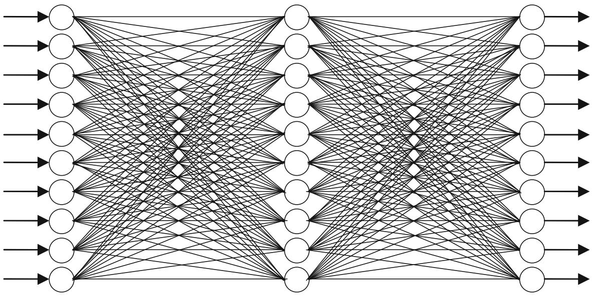

Fig. 13.1 Three-layer perceptron with full-connectivity

Table 13.1 Step activation threshold parameters

y_j	Base value for output connection j
x_i	Base value for input connection i
w_i	Weight of input connection i
n	Number of input connections
θ	Threshold parameter (0.85)

Table 13.2 Parameters of the genetic algorithm

# Children per generation	50
# Elites per generation	10
# Maximum generations	30000
# Learning runs	100
P (mutation) per bit	0.01

configurations has been selected, each of their input layers is activated with trial data and propagated as an activation wave through the network. Weight configurations are stored in a binary array. A score is awarded to each network based on the percentage of desired output values in a series of training sessions. The highest scoring networks (the elites) are then recombined using cross-over and mutation to form a new generation of network configurations, and so on. Due to the similarity with biological evolution and the storage of information in gene-like data arrays, this method is formally known as a *genetic algorithm* (Holland, 1975). The parameters of this particular GA are given in Table 13.2.

13.3 Experiments

Using the computational tools described above, the difference between indexical and symbolic interpretation is shown in a series of experiments. The two types of chimps (symbolic and non-symbolic) of the original language training research are modeled as neural networks. Objects, actions and lexigrams are replaced by binary strings of input and output data. The genetic algorithm acts as a training program, forwarding input data into the networks and evaluating the results.

For the indexical learning model, the objects, actions and lexigrams are coded according to the method displayed in Table 13.3. There are a couple of things

Table 13.3 Binary encoding examples for the indexical experiment

Network input	Binary string	Correct output	Binary string
banana + bias	1000000001	banana lexigram + give lexigram	1000000010
coke + bias	0100000001	coke lexigram + pour lexigram	0100000001
orange + bias	0010000001	orange lexigram + give lexigram	0010000010

that should be noted about this encoding. First, it disregards iconic interpretation processes by translating multifaceted entities into easily discernable icons. The chimpanzees are required to make distinctions between, bananas, yellow lexigrams, cans of coke and acts of pouring, but the neural network simply uses a ten bit binary string as input and output of the indexical process. This ensures that the neural network learns to create indexical associations, instead of a mixture of icons and indices: marginalizing the role of iconic interpretation isolates the indexical interpretation process which facilitates the study of its features. Also, in order to allow for a fair comparison with the symbolic network, a bias unit is added to the input vector.

The neural network is trained by the genetic algorithm to output the correct binary string, given a certain input string. For the input string, the leading eight bits indicate the presence of a particular edible or liquid, the ninth bit is always zero and the tenth bit is always one. The output string uses the leading eight bits to signify the use of a food or drink lexigram. The trailing two bits denote the use of an action lexigram.

Once the first pairing has been learned (i.e. *banana* with *give banana*), a second pair is added to the dataset. The learning continues with the same network and a training set of two possible input strings. This process is repeated until all eight objects have been associated with correct output sentences. The time it takes the network to learn each additional object is displayed in Fig. 13.2a.

The chimps that learn to manipulate lexigrams as symbols are induced to adopt a new learning strategy by the food and drink dispensers. These dispensers make them reconsider the relation between the lexigram buttons and obtaining a reward. They notice a systemic similarity between the system of lexigrams and the system of objects and actions (Deacon, 1997) and use their existing knowledge of the object domain to produce correct lexigram sentences.

For the symbolic learning model we use the same approach as for the indexical simulation, with the exception of the domain knowledge being available in the input string. In other words, the subject already knows that a banana is given (not poured) and takes this knowledge into account when it constructs a sentence. The additional information helps to predict the correct outcome, as actions and action lexigrams are correlated. The training data is shown in Table 13.4, the resulting learning curve in Fig. 13.2b.

A comparison between the learning curves of the indexical and symbolic models is somewhat biased. Just as the chimpanzees were at some point required to learn that bananas are given and milk is poured, so should the symbolic network, one

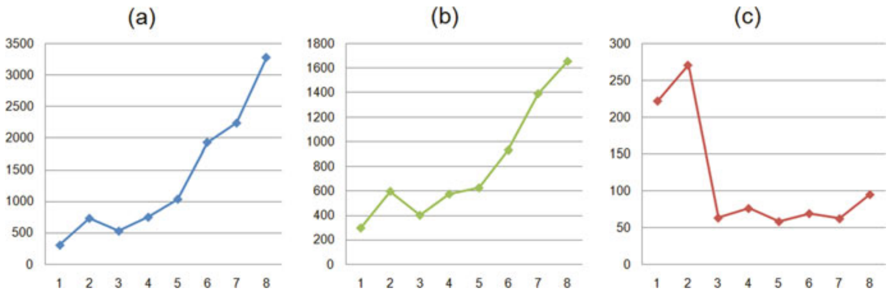

Fig. 13.2 Learning curves for the indexical task (**a**), the symbolic task (**b**) and the domain task (**c**). The y-axis indicates the number of generations it takes for each additional object (x-axis) to be learned

Table 13.4 Binary encoding examples for the symbolic experiment

Network input	Binary string	Correct output	Binary string
banana + give	1000000010	banana lexigram + give lexigram	1000000010
coke + pour	0100000001	coke lexigram + pour lexigram	0100000001
orange + give	0010000010	orange lexigram + give lexigram	0010000010

Table 13.5 Binary encoding examples for the domain experiment

Network input	Binary string	Correct output	Binary string
banana + bias	1000000001	give + bias	1000000001
coke + bias	0100000001	pour + bias	0100000001
orange + bias	0010000001	give + bias	0010000001

could argue. The goal of these experiments is to test the difference between indexical and symbolic learning; to exclude learning the domain knowledge would be a bias. Therefore, a third experiment is carried out. A neural network learns to associate objects with corresponding actions, using the same method as in the previous experiments. Table 13.5 contains the training data, the resulting learning curve is displayed in Fig. 13.2c.

13.4 Conclusion

A neural network model is used to simulate two different learning strategies in a series of three experiments. A genetic algorithm operates on a population of networks to train them in producing the desired output string. To generate a training dataset with input and output patterns, eight objects, two actions and ten lexigrams that were also used in the chimpanzee trainings tasks are encoded into binary patterns. For each of the experiments this results in a learning curve, showing the average number of generations needed by the genetic algorithm to find a working network configuration when a new object is inserted into the training dataset. The first experiment (indexical task) simulates how much learning time is required to map objects to lexigram sentences. In the second experiment (symbolic task) both the object and the action are part of the input. Finally, a third experiment (domain task) is added to avoid a possible bias. In comparing the indexical and symbolic task the learning time required for the domain knowledge task is added to the learning time for the symbolic task. This gives four learning curves, as shown in Fig. 13.3.

Several conclusions can be drawn from these curves. The domain knowledge task takes considerably less time than the other tasks, which can be attributed to the required output containing only one variable (either give or pour) instead of two. Also, there is an overall decrease in learning time after the third object is added. Once the two possible output patterns have been learned, the network has created a tendency to produce the right kinds of output patterns in the future. This holds for the indexical and symbolic tasks as well as for the domain task; however, due the steep learning curves of the former two this effect is not as significant.

The chimpanzee experiment claims that the apes that adopted a symbolic approach required more training time and made more errors during training, but once they had crossed the symbolic threshold they were able to produce better sentences and learn new symbols faster. Figure 13.3 shows that this also holds for the simulated interpreters. Requiring less time to learn the first objects, the indexical learning curve grows steeper than the symbolic learning curve in the long run.

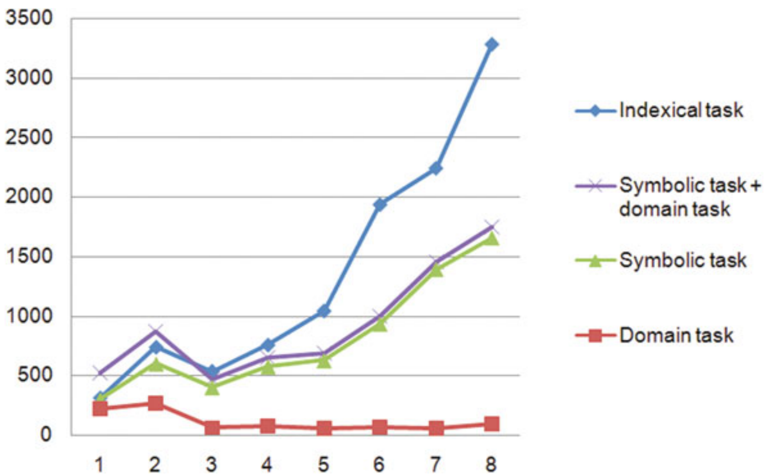

Fig. 13.3 Learning curves for all three tasks compared. The y-axis indicates the number of generations it takes for each additional object (x-axis) to be learned

13.5 Discussion

We have set up the neural network experiments in order to investigate the differences between indexical and symbolic learning. Although such a difference can be shown to exist in our models, the experimental findings do not prove the accuracy of the models used nor do they validate the conversion from the chimpanzee language training program to the simulation. It should be noted that too many simplifications and assumptions had to be made to call these networks either indexical or symbolic interpreters. In order to reduce the complexity and tractability of the learning task, a relatively straightforward neural network is used. Also, even though a bias is avoided by adding the domain task, it is unclear how exactly the learning curve of the domain task and the symbolic curve ought to compare to the results of the indexical task. One should therefore be prudent with generalizing the particular model and approach used in these experiments.

However, when the results are projected onto the semiotic theory (similar to the approach taken by Savage-Rumbaugh and Rumbaugh), they do allow for interesting conclusions to be drawn. The learning curves help to identify the mechanisms that underlie the shift to symbolism. The findings show that this shift serves a practical purpose as it allows the subject to off-load memory from one domain to another, thereby avoiding duplication of information. With selection pressure favoring language use, this gives an advantage to symbolic over non-symbolic systems. The findings also indicate that for a symbolic shift to take place, the different domains (e.g. the domain of objects and actions and the domain of lexigram relations) are required to be mapped onto each other by the interpreter. Understanding how this mapping takes place is an important step towards a more accurate simulation of the interpretation process and the role of symbols herein.

Recall our two definitions of symbols, *S1* and *S2*. In the case of *S1* a lexigram would point directly to a referent (i.e. an index). According to the second definition *S2*, the symbol would also have a pointing relation to its referent, albeit a more obscure one which is embedded in a web of interdependent referential relationships. In the chimp experiments, the relations that exist among objects and lexigrams are also embedded in a web that spans both the lexigram domain and the object-action domain. A lexigram can be an index for another lexigram: their simultaneous use will likely lead to a pointing relationship from one to another (*banana lexigram* is usually followed by *give lexigram*, hardly ever by *pour lexigram*). The realm of objects and actions has a similar system of pointing relations (coke is always poured and never given). Therefore, a symbolic relation is, as one might say, a *higher-order pointing relation* from one domain to another.

For the interpreter to create this kind of relation, it needs to find domains that can be mapped onto one another. Not every pairing of indexical systems is viable, there has to be a correlation between them that makes linking them purposeful. The input data presented in the symbolic task has some redundancy in it, so it makes sense for the interpreter to correlate the system of lexigrams with the system of actions and objects (cf. Table 13.4). It is exactly this redundancy or *system iconicity* (redundancy implies a lack of difference) in the topology of the systems that makes a symbolic relation advantageous (Deacon, 1997). A symbol, therefore, is a triadic relation that requires two systems of indices with topological redundancy, resulting in a higher-order index between two loci in those systems. The recognition of this redundancy, the *insight* that two domains are alike, is prerequisite for the symbolic shift to occur in an individual.

We can take this deconstruction of the sign one step further and consider what an index, being the constituent of symbols, is itself composed of. A pointing relation always points from one thing A to another B, which may in turn point to a third C and so on. The index from A to B is activated by the recognition of A (which is an iconic process). By virtue of their indexical relationship, A causes B to become active (as though B has been recognized). Suppose for example that A is smoke and B is a fire. The thought of a fire may cause a new thought C, no matter whether the fire was perceived directly (icon) or thought of after perceiving smoke (index). Consequently, what is caused by an index is also an icon.

The pointing relation itself is caused by a recurring appearance of signal and referent, being in close proximity to each other in one or more dimensions (i.e. spatial or temporal). Recognizing B frequently after recognizing A causes the interpreter to make a prediction about the future occurrences of B after A. The commonality of these situations is the simultaneous occurrence of signal and referent. Once the signal appears again, the interpreter recognizes the state as one of those situations where both signal and referent occur together. This *recognition* is itself a higher-order icon, because it classifies the signal-referent relation as one of many that have occurred before. Hence, an index is a relation between two icons that exists by virtue of a higher-order icon: their regular co-occurrence.

As an index is solely composed of icons, and a symbol is a particular configuration of indices, it follows that icons are the primary building blocks for all three types of interpretation. This conclusion does not imply that every iconic interpreter is also an indexical or symbolic interpreter. As the ape language training tasks as well as the simulation experiments show, a specific configuration is required for symbolic interpretation. Some apes were clearly unable to do symbolic interpretation even though they had indexical capacities. The neural networks that were trained to learn indices clearly show a behavior that differs from symbolic networks. Likewise, indexical interpretation requires a specific setup of iconic skills in order to induce the formation of a higher-order icon.

This conclusion *does* imply that iconic interpretation is a fundamental skill for interpretation. The firstness of icons is argued for in semiotics (Peirce, 1894) but also by the proficiency of simple neural network models in classification tasks, where their robustness allows them to deal with distorted data (Kohonen, 1982; Harnad, 1990). The potential of these computational models for recognition and classification tasks makes them a good starting point for further investigations into associative and symbolic models of interpretation.

References

Buchler, J. (1955). *Philosophical writings of Peirce*. New York: Dover.
Chandler, D. (2002). *Semiotics: The basics*. London: Routledge.

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain*. New York: W. W. Norton & Co.

Deacon, T. W. (2003). Universal grammar and semiotic constraints. *Language Evolution, 7*, 111–139.

Gardner, R. A., & Gardner, B. T. (1977). Comparative psychology and language acquisition. *Psychology: The state of the art, 309*, 37–76

Harnad, S. (1990). The symbol grounding problem. *Physica, D42*, 335–346.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.

Hookway, C. (1985). *Peirce*. London: Routledge & Kegan Paul.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43*, 59–69.

McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 7*, 115–133.

Peirce, C. S. (1894). What is a sign? *The essential peirce: Selected philosophical writings, 2*, 4–10. Indiana University Press.

Premack, D. (1976). *Intelligence in ape and man*. Mahwah, NJ: Lawrence Erlbaum Associates.

Rumbaugh, D. (1977). *Language learning by a Chimpanzee: The Lana project*. New York: Academic.

Savage-Rumbaugh, S., & Rumbaugh, D. M. (1978). Symbolization, language and chimpanzees: A theoretical reevaluation on initial language acquisition processes in four Young Pan Troglodytes. *Brain and Language, 6*, 265–300.