Symbol Grounding and Transcendental Logic

Erich Prem
The Austrian Research Institute for Artificial Intelligence
Schottengasse 3, A-1010 Wien, Austria
Tel. +43 1 5336112, FAX +43 1 5320652
erich@ai.univie.ac.at

Abstract

Symbol Grounding tries to answer the question as to how it is possible for a computer program to use symbols which are not arbitrarily interpretable. Whereas the signs in conventional programs are just "parasitic on the meaning in our heads", grounded symbols should possess at least some "intrinsic meaning." This paper gives a brief overview of what Symbol Grounding is and summarizes some of today's connectionist Symbol Grounding models. Instead of concentrating on cognitive linguistics, we try to present an alternative view of Symbol Grounding. Our analysis reveals that Symbol Grounding is in fact the endeavour of automated model construction. Although it originated in a somewhat anti-formal spirit it is (necessarily) full of parallels to classical symbolic logic. We present our view that Symbol Grounding is in fact a connectionist version of transcendental logic, which is the basis for generating formal models of non-formal domains. Such formalizations are inherently logical, though not only based on formal but also on material truth conditions.

1 The Symbol Grounding Problem

One of the best known criticisms of what is now often termed "the symbolic approach to AI" has been introduced by John Searle. His *Chinese Room Argument* [Searle 80] is one of the most debated arguments in the philosophy of AI (and will therefore not be repeated here). Searle points us to the fact that all semantics in an AI program is but pure syntax. The central construct of such programs, the *symbol*, has no other meaning than that which the programer has given to it through connecting it to other, however again, symbolic descriptions. The reason for this lies in the very fact that generally *symbols are arbitrary* and that AI programming symbols are especially arbitrary since their concrete form implies no restriction on how they are processed. This diagnosis has been the starting point for Stevan Harnad, who in his 1990 paper "The Symbol Grounding Problem" defines a research problem for the AI community. This research problem is stated as the quest for a method which grounds symbols in something else but yet another symbolic description.

Before we turn to Harnad's proposal of how to solve this problem it is worth pointing out some subtle differences in Harnad's and Searle's accounts of the problem. In accordance with his previous work [Searle 83] Searle is more concerned with the problem of *intentionality* than with *meaning.* His main argument is that symbol structures in a program do not (and indeed cannot) implement the right causal powers for a computer program to *refer,* i.e. to have original intentionality. His example, the chinese language "understanding" program serves to make this point clear: In any such program there is a separation of causal powers manipulating symbols and those producing intentionality.

Harnad, on the other hand, is more concerned with how (language) symbols acquire their meaning, i.e. the problem is restated in terms of a problem for cognitive science. For him, this problem can be reformulated as how we can construct symbols which are not totally arbitrary in their interpretation, but are in some sense causally connected to the environment or the experience of an intelligent system. Such symbols would then, according to Harnad, no longer be "parasitic on the meanings in our heads" but possess original meaning.

This reformulation is, of course, in much closer correspondance to cognitive linguistics than Searle's problem is. Especially, it is close to language acquisition in children, if we assume

that words are – cum grano salis – some kind of symbols. (That they are not is discussed in [Dorffner et al. 93].) Harnad's (and others') idea to create these symbolic structures consists in paralleling a somewhat naive view of language learning [Harnad 93]. This view says that symbols (words) stand for objects (resp. classes of objects). Their meaning is acquired through two processes. (i) Categorization, i.e. the process of forming object classes. (ii) Symbolization, i.e. labeling such a class with an arbitrary name.

Neural networks seem particularly suited for fulfilling these requirements for the following reasons: They are able to form categories or clusters in an unsupervised fashion. They can also be trained to label inputs in a supervised way. And they seem suited for directly dealing with immediately grounded sensory input, i.e. input which comes directly from a measurement device. Before we begin our analysis of Symbol Grounding (SG) principles, we briefly overview a few SG implementations.[1]

2 Today's Connectionist Symbol Grounding Models

Two classes of models form the major part of connectionist SG endeavours: (i) pure object models, (ii) verb models.

(Models which try to ground other aspects of natural language, like e.g. Terry Regier's model of grounding spatial terms [Regier 92] will not be treated here.)

2.1 Object models

One group of connectionist proposals addresses the problem of grounding object categories in perception. Typically, the models possess two types of input: one for (simulated) sensory data, another one for symbolic descriptions of the data. The networks are usually chosen so that the sensory input is categorized through un- or selfsupervised categorization algorithms like Kohonen networks. The resulting category representations are then associated with some symbolic description of the static input.

A typical and early example is the model of Cottrel et al. [Cottrell et al. 90] where faces of a small group of people (i.e. video images) are mapped onto their corresponding names. The goal is to ground the names in perception. Figure 1 depicts the model. For each modality (faces and names) autoassociative networks exist, which extract the corresponding class features by means of error backpropagation. Each network, names and faces, is trained separately. In a second training phase associations from faces to names and vice versa are trained.

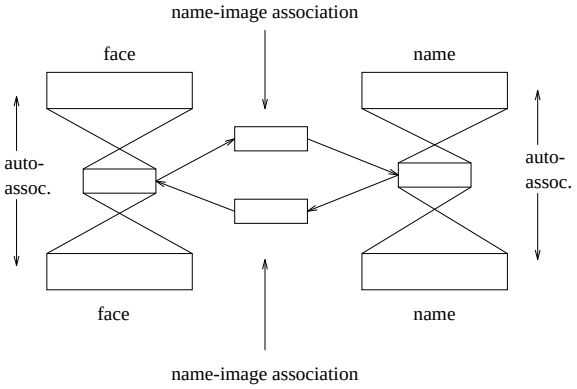

Figure 1: Symbol Grounding Model (Cottrell et al.)

Architectures which are highly similar in their functional principles have also been proposed by [Chauvin 89], [Schyns 91] and [Dorffner 92]. All three authors perform Symbol Grounding as a model of concept (and word) acquisition in children. In all three approaches, the naming phase (associating input pictures and labels) is separated for psychological plausibility. I.e. in a first phase concepts are formed through unsupervised category formation. These concepts are labeled with a relatively arbitrary name in a second phase. However, only Dorffner's model is explicitely designed so as to avoid that similarities in the input are reflected in similarities in the names. We call this effect "Symbol Blending."

Connecionist models for grounding dynamic aspects of language in sensory experience do not exhibit such strong similarites like the ones presented in the previous section. This is mainly due to the fact that processing input sequences is still a novel research issue and comparatively difficult. Models for dynamic aspects of language have e.g. been presented by Cottrell [Cottrell et al. 90] or by Nenov and Dyer [Nenov & Dyer 88]. Cottrell's model is schematically depicted in figure 2.

[1] For an account of SG viewed from the perspective of historical AI criticism see [Prem 94b].

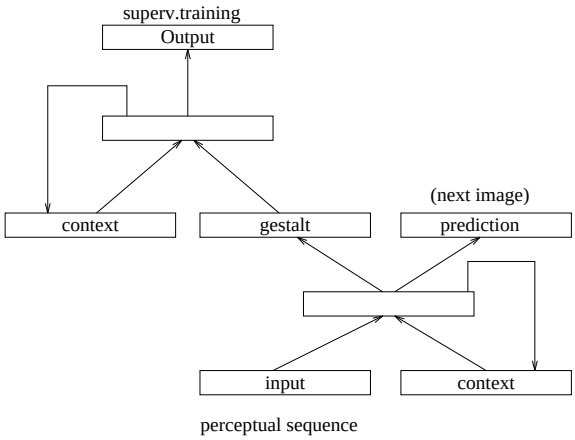

Figure 2: The *movie description network* (Cottrell et al.)

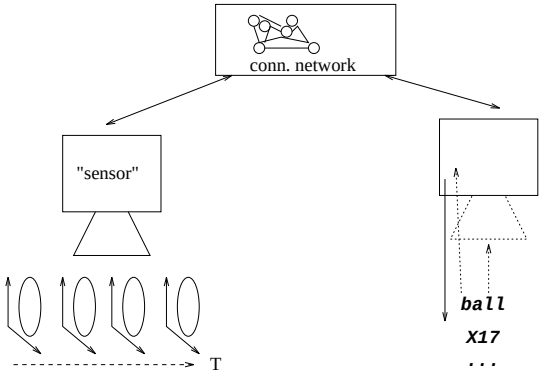

Figure 3: A generic SG architecture.

These models are capable of learning simple sentences like e.g. "ball bounces right." Their main difference to the object models is twofold: (i) The model does not only connect sensory inputs to a symbol, but it also learns to produce a sequence of such symbols. (ii) The input consists in some sort of time sequence, not only of a static input image.

3 SG and Model Construction

SG is pursued with two rather different motivations: One is the development of a model of aspects of natural language. The other is the quest to enrich formal symbol structures with some kind of sensor-based information. The main assumptions with which these goals are pursued are:

- Symbols are entities similar to words for sensory qualities.
- Semantics grows from a historical coupling of signs to sensory experiences.

The general picture arising from this background of SG models is depicted in figure 3.

Figure 3 shows that SG is implemented today as a method for the automatic development of categories over some sensor space and the automated labeling of these categories with presented shapes. The goal is that after training, the SG system is capable of "saying" the right symbols in a given situation. This means that in SG models the only purpose of symbols is their reference to external objects, resp. to the sensory data generated by them. A theory of SG which largely adopts such a conception can only result in a specifically scientific (i.e. descriptive) model which is designed to reflect natural phenomena for epistemic purposes. I would like to suggest that this process is highly similar to the automated development of a formal model of a natural system. ("Natural" because we are still talking about grounded systems, which are connected to the world through measurement devices.) Some kind of encoding procedure maps "objects" of the environment (usually trough employing meters) onto symbols in a formalism. Implications in the formalism are then used to predict what natural law does to the natural system. This situation is depicted in figure 4.

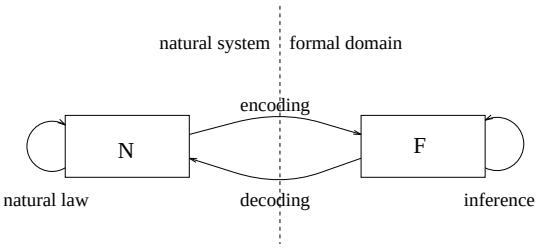

Figure 4: The modeling relation.

That SG systems do a bit more than just labeling inputs, i.e. that they even implement basic descriptive or predictive rules will be described below. Before we begin to investgate the relations of SG to some aspects of formal logic, a brief summary of transcencental logic shall be given below.

4 Transcendental Logic

For the convenience of the reader who may not be familiar with transcendental logic, some of its basic principles shall be outlined in this section. However, only those aspects which are important for the parallels with Symbol Grounding can be mentioned here and this account will necessarily be simplifying.

Transcendental Logic originated with Immanuel Kant's (1724–1804) work on epistemology. His endeavour can be approached from two points: The first tries to find an answer to the question "How are synthetic judgements a priori possible?", i.e. it searches for common principles in all theoretical sciences. If these principles are identified with the human process of knowledge acquisition, this question turns into "What are the conditions under which we construct our knowledge of the world?". At the basis of this process of construction lies the question as to how concepts are acquired. A set of such principles would be called "transcendental logic".

Transcendental logic can be called logic because it searches for those conditions under which propositions are true. Today, of course, logic has mainly been reduced to purely formal logic. This development towards a purely formal account of logic, founding truth only on the shape of symbols in a proposition, has already been visible to Kant. In fact, his *transcendental* logic has been developed so as to *overcome* such a purely formal approach. Instead, the development of a logic which does not "abstract from all content of experience" must be based on the subject's way of experiencing the world. This, in correspondence to English empiricism is a central belief of Kant: Only that cognition is legitimate which is based on experience.

In his "Critique of Pure Reason" Kant formulates the principles of such a transcendental logic, the final set of tools of reason. At their heart lies a set of categories which enumerates all the principles which are necessary to bring "the manifoldness of experience into the unity of concepts." This list parallels Aristotle's categories and his traditional forms of reasoning, a fact which has subsequently interested Hegel in his "Science of Logic" [Hegel]. We shall now investigate the fact that the basis of these categories (i.e. the classical forms of judgement) are visible again in SG systems.

5 Grounding Signs

Let us approach logical aspects of SG by asking if it is really symbols which are grounded in the above models. In order to do this, we need an acceptable definition of what a symbol is. One definition which is widely accepted in the AI community has been given by C.S.Peirce. He characterizes the symbol as a special type of sign, which is connected to its object (its referent) through the idea of the symbol user, without which no such connection would exist. I.e., a sign looses the property of being a symbol without being interpreted. What makes symbols so useful for AI is that their actual shape is totally arbitrary. Since it is only the idea of the symbol user which guarantees connection of sign and object, this idea can negotiate between any object and any form.

The other two types of signs on which Peircean semiotics is based are icon and index. The icon (earlier called "simile") refers to its object because of an "inner similiarity." Take as an example the picture of a house. Because of their similiarity in appearance, i.e. in their gestalt, the house and its image are connected. If we try to make more excplicit what an "inner" similarity is, we find that this means a similarity in an essential property, i.e. one that allows us to identify the object as that special object.

The index, on the other hand, stands for an object because of some natural (or "outer") property. The index is also a sign without anyone interpreting the sign. One of Peirce' own examples is the polar star being an index for "north" (or the thermometer for temperature).

If we now go back to our SG systems we see that what is actually generated there is some inner similiarity in the model according to which the sensory input maps onto a sign. However, in the models which we have described above (except maybe Dorffner's example) the mapping cannot be performed between any arbitrary choices of sensory patterns and "sign" patterns. The features of the connectionist networks employed ensure that similar inputs are mapped onto similar outputs. This has the consequence that similar sensory vectors will necessarily map onto similar symbols. Not only is this a cognitively implausible effect, but it is also inconsistent with what we would expect symbols to do, i.e. either to refer or not refer to their object. In addition to this, SG systems are, of course, computer models which (after training) will always

work predictively based on natural law. This means that grounded symbols also exhibit aspects of indexes.

Obviously, we find that the process of grounding symbols in experience destroys at least some of its purely symbolic character for it generates icons which also have some properties of indexes. It is only the fact that in the usual SG model we can connect many symbols to many instances of sensory vectors that it is still justified to speak of grounded symbols. However, recalling that the symbol is originally only connected to its object because of an idea alone, we should not be surprised that the effort of basing it on automated (indexical) generation of similarity relations (icons) destroys some of its arbitrary (symbolic) character.

However, this result is mainly an argument about how to call SG systems. What is more important instead, is that this threefold structure of semiotic reference is tightly connected to three forms of reasoning. In fact, the three types of signs *represent* the three classical inference methods.

6 Symbol Grounding Models and Forms of Reasoning

At a first glance it may seem strange to construct a connection between SG and logic. This impression arises from the fact that logic nowadays nearly always means *formal* logic. However, if we go back to C.S. Peirce, we find a definition of logic as *the science of the conditions which enable signs to refer.* If one assumes now that symbols already refer to objects in the world, this science naturally reduces to syllogistic and set theoretic considerations. If we ask instead, how it comes that a given sign means this or that, we are engaged in what is traditonally termed *transcendental* logic.

6.1 Signs and Inferences

We shall now ask on what basis are we entitled to say that sign X means Y. We will do this by again using the Peircean characterization of signs. The reader should be warned that if he insists on a view that is inherently formal all the following forms of reasoning will seem to be mere variants of deduction only. What will be done subsequently is, however, more in correspondance to Hegelian logic and, as Zeidler claims in [Zeidler 92], also in correspondance with the original intention of Aristotle. This original intention was *not* only to describe completely formally the process of reasoning, but to enable the natural scientist to generate *new* knowledge. According to Aristotle then, his second figure "in a certain sense" proofs the confirming and his third figure makes generalization possible [Aristotle, 62b].

Because it is so familar to us, we begin with the *icon*. We have seen that in this case some common "inner" quality allows us to identify sign and its referent. It is the possession of a common (essential) property P which negotiates between the two. By reference to P, the icon represents *intension.* In more formal terms, let M be the set of object states, Q be the set of the sign states. Then it is the fact that both, M and Q, possess (belong to a set) P which makes them stand for each other. (M is P. Q is P. $\rightarrow$ M is Q.) This negotiation takes exactly the form of a logical inference, in fact, of abduction or analogy. C.S. Peirce has called this process *hypothesis, abduction* or *retroduction.* In this case a rule (e.g. "All beans in this sack are white.") and a result ("These beans are white.") give a concluding case ("These beans are from this sack.") [Peirce, 3:325f]. Such a conclusion always happens without a guarantee, but this is necessary to discover new knowledge.[2]

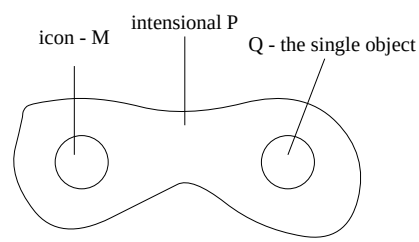

Figure 5: Negotiation through an essential property.

The position of the terminus medius in this process suggests that such an inference corresponds to Aristotle's second figure [Aristotle, 26b/27a]. The terminus medius has the position of the predicates, it is what is "said" about both subjects. It is clear, of course, that in a

[2]H. Simon also states that this is the place in reasoning, where the "real" inventions are made [Simon 65, p.186].

purely *formal* approach, abduction is only possible with a negative conclusion. The classical example goes like this:

> All stars glow by their own light.
> This object does not glow by its own light.
> ⇒ This object is not a star.

In order to accept our view that the icon represents abduction, it is necessary to see that the role of intension corresponds to the role of the general in classical reasoning; that the object comprises the singular and that the sign represents the special. Terminus medius of iconic representation is the icon's intension. Fig. 6 represents a part of a hierarchical concept tree.

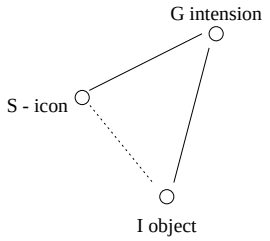

Figure 6: Part of arbor porphyriana. G-general, S-special, I-singular. Straight lines show relations given, dotted line represents conclusion.

Identifying sign and object empirically will therefore succeed if both are recognized as being essentially of some property P.

The *index* is a bit more problematic, mainly because we are not so much used to working with it. We remember Peirce' hint that the index is also index without anyone interpreting it. Therefore, what is characteristic about it, must lie in the object itself. This is what is originally meant by the "naturalness" of the index or, in other terms, that the connection exists because of natural law. It is the object alone which informs about its quality of being an index. Viewed logically, this process is similar to induction, which informs about the total set of objects with a specific property. Generally, in induction we are given a set X with elements x_i and a subset S of X. We know now (e.g. from experiments with some s_i) that for all s in S there holds a property $P(S)$. Induction means to generalize from $(\forall s)P(s)$ and $S \subset X$ to $(\forall x)P(x)$.

In the case of the index the single objects negotiate their meaning, they are the elements of induction, which are measured by a meter (the index, e.g. a thermometer) and which are instances of a general property, like temperature. Figure 7 depicts this relation.

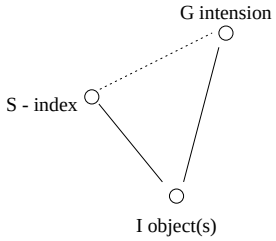

Figure 7: Part of arbor porphyriana. G-general, S-special, I-singular. Straight lines show relations given, dotted line represents conclusion.

In Aristotelian induction the terminus maior is proven for the medius by means of the terminus minor [Aristotle, 68b]. In Aristotle's third figure the terminus medius is the subject of both predicates: A is B, A is C; i.e. in the hierarchical concept tree, we find the medius at the bottom (cf. fig. 7).

Finally, the symbol is a general representation of an object. It is *the sign alone,* which negotiates its meaning. In other terms, the symbol represents intension and extension, it alone ensures that one is in connection with the other. In a psychological interpretation this twofold representation becomes a picture of (i) a subjective interior world and (ii) an objective outer world. In Aristotelian terms, again, this perfect double reference of the symbol corresponds with the double role of terminus medius of Aristotle's first figure (deduction) as subject *and* predicate. This observation lies at the bottom for viewing symbols as a means of communication. The symbol here is injected between the singular (the object) and the general (intension). This is why symbolic negotiation is similar to deduction. The clear boundary between different symbols which can be found in some SG models is nothing but the consequence of a set theoretic consideration. Objects possess a property or they don't. Symbols refer or don't, tertium non datur. (See figure 8.)

We have seen in the previous section that SG systems are not only grounding symbols, but the

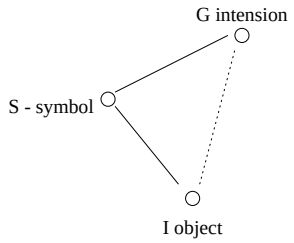

Figure 8: Part of arbor porphyriana. G-general, S-special, I-singular. Straight lines show relations given, dotted line represents conclusion.

signs with which we are dealing show at least some aspects of icons and indexes, too. The neglect of these differences has lead to the "Symbol Blending" effect in many SG systems. Additionally, it has been shown in this section that Peirce' three types of signs are representations of three classical forms of reasoning. Therefore, it should be possible to identify the concrete procedures in SG systems with these building blocks.

6.2 SG-architectures and Forms of Reasoning

In SG models, we find the following forms of classical inferences, hidden in connectionist approaches to the problem.[3] Although it is not always very clear where to put one or the other technique, the following system is suggested.

In the case of the indexical aspects of SG models, it seems natural to identify them with the phase of unsupervised categorization. The actual formation of what amounts to a specific object usually happens without a teacher. Only the objects themselves inform the system about the categories to which they belong. Figure 9 shows the regions in which we find such empirical exemplars of the inductive process.

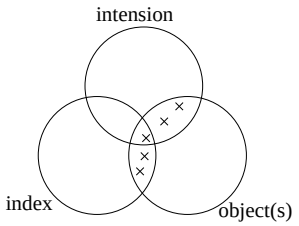

Figure 9: The inductive basis of SG.

[3] Here we depict the relations like it is often done in classical logics (e.g. [Quine 64]) with Venn diagrams.

In connectionist models the inductive process is based on vectorial similarity (e.g. the Euclidean distance) and on relative frequencies of input vectors.It seems, however, natural to extend this process also with aspects of reinforcement learning. This way, the *utility* of a concept, instead of its mere vectorial similarity could play a role in categorisation.

Today's SG models perform an extreme simplification of this step, because the "index and its referent," i.e. the sign and the object are already separated in separated parts of the system. All SG systems posses one modality for the symbols, another one for the objects. In real life, however, this separation is not so simple.

Secondly, there are the abductive aspects. Figure 10 shows the possible area for empirical evidence of grounding icons. I.e. in sections which are marked by a cross (x) empirical evidence for the fact that *this* sign stands for *that* object can appear. Of course, we are mainly interested in those cases, where the crosses are concentrated in the intersection of all three circles. Only in these cases will our generalization be correct.

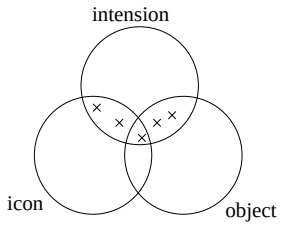

Figure 10: The abductive basis of SG.

We therefore need strategies to generalize correctly. In SG architectures, to actually identify the objects encountered with the signs presented, the most important strategy is supervised learning. In those cases where SG is performed with language learning in mind this is supposed to model timely coincidence of the sign and its significatum in the process of teaching a word to a child. The problem we are facing is, however, not trivial. In the general case it would mean to find the part in the input picture to which the sign refers.[4]

[4] The reader may compare this to Quine's example of "radical translation," where we try to find out what the native's exclamation "Gavagai!" means. Is it the rabbit running over the field, a part of it, materialised rabbitness, etc.? Cf. [Quine 60]

Finally, it must also be guaranteed that *symbols* are grounded. I.e. that we implement some sort of *tertium non datur* principle. Dorffner achieves this through so-called C-layers, which implement a variant of competitive learning so as to ensure that only concepts which are sufficiently stable are mapped onto symbols.

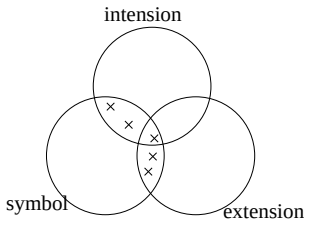

Figure 11: The deductive basis of SG.

If the symbol is to fully negotiate between extension and intension, there is not very much room left for empirical evidences as shown in figure 11.

7 Automated Model Construction

7.1 Logical Principles and Time

It has been proposed above to replace the view of SG as a psycholinguistic program by a view which suggests that SG is some kind of automated model construction of a natural system. But for this view the entailment rules in the formalism are still missing. All we have achieved so far is the formalisation of some environment states into symbols. The extension of this discussion to space and time is quite natural and has also been performed by Kant and his successors. Although these parallels still form a research question (currently under investigation at our lab) I shall briefly overview the argumentation here. The three forms of reasoning can be identified with three principles of logic, which are listed below: (From [Zeidler 92].)

1. *A is B.* At the very basis of abduction lies the idea of "sameness". This is what makes analogy such a difficult form of reasoning (even in purely formal domains): that sameness depends on what is considered to be important (or essential, as we have termed it above).

2. *If B then Non-B.* This principle is argued to lie at the basis of induction, being expressed as "separation."

3. *A is either B or Non-B.* It has already been argued here that the principle of the excluded middle lies at the basis of deduction. It can be characterised as "opponentship."

If one follows this suggestive summary, it is easy to see how such a system finally leads to formal models. Fig. 12 tries to suggest that, intuitively, *sameness* creates the permanency of objects in time. *Separation* serves as the (inductive) source of creating causal relations (Kant: "The real whereupon something follows"). And *opponentship* puts the "objects" into relation with each other. Our idea is now to also identify sameness with the notion of a state in a formal model, separation with the timely and causal sequence of states. However, this suggestive terminology is still under evaluation.

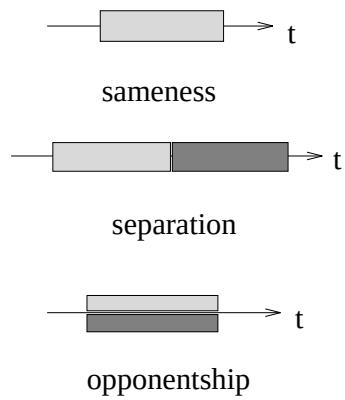

Figure 12: Schematic representation of three transcendental principles.

7.2 An Implemented Autonomous Agent

Finally, to give this view some AI practicality we are studying these principles by means of a simulated autonomous agent. (Further aspects of practicality are discussed in [Dorffner & Prem 93].) This agent actively explores its environment and forms concepts on the basis of (simulated) environmental interaction. We are not pursuing this research in order to establish yet another model of SG. Instead, the idea is to investigate automated model construction as a form of transcendental reasoning

which can be achieved by using the principles which have been outlined here.

There are also some new aspects of pure SG in our project, e.g. we are dealing with the utility of concepts. The agent learns to distinguish situations based on reinforcement signals. Another rather novel feature of our architecture (as far as SG and autonomous agents are concerned) is that we are using recurrent networks. This enables the agent to use information about a whole sequence of environmental states and his actions. Very often such models are only capable of classifying single input vectors, comparable to some kind of "slide show" which the agent observes. The most important practical aspect of this research is to make the agent describe what it is experiencing by means of signs that are arbitrarily selected by a supervisor. This way, the agent develops some kind of "subjective" world-model, which is, however, expressed in formal terms (symbols). By using the principles which have been outlined in the last section we hope to make the agent not only refer to single objects in its environment, but to also develop descriptive (and predictive) rules. A simple example for this would be an agent that learns to distinguish between distinct sequences of perceptual features (like odour sources in a tunnel) that influence the agent's decision to turn left or right. If the agent becomes capable of describing such a rule, it would have constructed knowledge about its environment, which by means of supervised teaching of symbols would also be meaningful to human observers of the agent. This would be the ultimate goal: an agent which discovers and describes its environment in terms of symbols which are not totally arbitrary in their interpretations.

8 Conclusions

The following two steps are the basis for our considerations here:

1. According to Kant, transcendental logic does not abstract from all contents of experience. Instead, it actually asks what the procedural prerequisites for (the generation of) the objects of experience are.

2. Peirce regards logic as a scientific endeavour which clarifies the conditions that enable signs to refer to objects.

As a consequence of these conceptualizations we conclude that transcendental logic and Symbol Grounding are based on the very same principles of logic. Since SG systems try to form concepts and label these concepts, they try to implement this view of transcendental logic. This implementation comes together with a search for principles that will make these systems work. In SG papers these principles are sometimes called "invariants" of sensory experience. SG model builders are therefore in search of rules which will allow the system to find the object in its visual field when given a name for it. In Kantian terms this means to search for the transcendental basis of cognition.

But as we have seen above, this basis is not genuinely simple. It seems that everything we can construct will crucially depend on how we implement the abductive step of finding general intensions. Following Hegel we can even say that the three forms of reasoning seem to condition each other. For the AI researcher, this means that some principle must be missing here. For the transcendental philosopher it is, of course, the lack of consciousness, which reflects the threefold structure of reasoning.

9 Acknowledgements

The author wishes to thank G. Dorffner, A. Flexer, H. Trost and K.W. Zeidler for their comments on the manuscript. Especially, I am indebted to K.W. Zeidler for first having pointed me to the commonalities between transcendental and formal logic. The continuous support by R.Trappl is also gratefully acknowledged. The Austrian Resarch Institute for Artificial Intelligence is supported by the Austrian Federal Ministry of Science and Research.

References

[Aristotle] Aristotle *Anal.pr. (Organon III).*

[Chauvin 89] Chauvin Y.: Toward a Connectionist Model of Symbolic Emergence, in Proceedings of the Eleventh Anual Conference of the Cognitive Science Society, Lawrence Erlbaum, Hillsdale, NJ, pp.580-587, 1989.

[Cottrell et al. 90] Cottrell G.W., Bartell B., Haupt C.: Grounding Meaning in Perception, in Marburger H.(ed.), GWAI-90, Springer, Berlin, 1990.

[Dorffner 92] Dorffner G.: A Step Toward Sub-Symbolic Language Models without Linguistic Representations, in Reilly R. & Sharkey N.(eds.), Connectionist Approaches to Natural Language Processing, Vol.1, Lawrence Erlbaum, Hove, East Sussex, United Kingdom, 1992.

[Dorffner & Prem 93] Dorffner G., Prem E.: Connectionism, Symbol Grounding, and Autonomous Agents, *Proc. of the 15th Ann. Meeting of the Cognitive Science Society*, Boulder, CO, p. 144–148, 1993.

[Dorffner et al. 93] Dorffner G., Prem E., Trost H.: Words, Symbols, and Symbol Grounding, to appear in: *Proceedings of the 16th Wittgenstein Symposium,* Vol.2, 1993.

[Dreyfus 84] Dreyfus H.: *Why Computers Can't Think.* Macmillan 1984.

[Dreyfus 90] Dreyfus H.: *Being-in-the-world.* MIT Press, Cambridge, MA, 1990, p.115–127.

[Frege] Frege G.: Über Sinn und Bedeutung, *Zeitschrift fuer Philosophie und philosophische Kritik,* 100,25-50, 1892.

[Harnad 90] Harnad S.: The Symbol Grounding Problem, *Physica D*, 42 (1990), p.335–346, 1990.

[Harnad 93] Harnad S.: The Origin of Words: A Psychophysical Hypothesis, in Durham W., Velichkovsjy B. (Eds.), Muenster:Nodus Publ., 1993.

[Hegel] Hegel G.W.F.: *Wissenschaft der Logik*, (Suhrkamp, Frankfurt/Main, 1990) 1816.

[Nenov & Dyer 88] Nenov V.I., Dyer M.G. DETE: Connectionist/Symbolic Model of Visual and Verbal Association, in Kosko B.(ed.), *Proc.IEEE International Conference On Neural Networks,* San Diego, IEEE, Vol. II, pp.17-24, 1988.

[Peirce] Fisch M.H., Kloesel W. (Eds.) *Writings of Charles Sanders Peirce. A Chronological Edition.* Indiana University Press, Bloomington, 1982.

[Prem 94a] Prem E.: *Symbol Grounding. Die Bedeutung der Verankerung von Symbolen in reichhaltiger sensorischer Erfahrung mittels neuronaler Netzwerke,* PhD Thesis, University of Technology, Vienna, forthcoming.

[Prem 94b] Prem E.: *Symbol Grounding Revisted.* To be presented at the Workshop on Combining Symbolic and Connectionist Processing, 11th European Conf. on Artif.Intell., ECAI'94, 1994.

[Quine 60] Quine W.V.O.: *Word and Object,* MIT Press, Cambridge, MA, 1960.

[Quine 64] Quine W.V.O.: *Methods of Logic,* Revised Ed., Holt, Rinehart & Wilson, New York, 1964.

[Regier 92] Regier T.: Learning Perceptually-Grounded Semantics in the Lo Project, in *Proc. 29th Annual Meeting of the Association for Computational Linguistics,* Association for Computational Linguistics, pp.138-145, 1991.

[Schyns 91] Schyns P.G.: A Modular Neural Network Model of Concept Acquisition, *Cognitive Science,* 15(4), 1991.

[Searle 83] Searle J.R.: *Intentionality: An Essay in the Philosphy of Mind.* Cambridge: University Press, 1993.

[Searle 80] Searle J.R.: Minds, Brains and Programs, in *Behavioral and Brain Sciences,* 3,417–457, 1980.

[Simon 65] Simon, H.A (1965). The Logic of Rational Decision. In *The British Journal for the Philosophy of Science.* 63 (16), 169–186.

[Smolensky 88] Smolensky P.: On the Proper Treatment of Connectionism, in *Behavioral and Brain Sciences,* 11(88), p.1-74, 1988.

[Zeidler 92] Zeidler K.W.: *Grundriß der transzendentalen Logik,* Junghans Verlag, Cuxhaven, 1992.