# Luc Steels, the Talking Heads Experiment and Cognitive Philosophy

A tutorial accompanying the presentation in Current Issues

PART II - The Talking Heads Experiment

Gideon Borensztajn, Vidhi Trehan

UvA, April 2006

# 1 The Talking Heads Experiment Explained

### 1.0.1 A note about terminology

In describing the TH experiment, we will use the term *meaning* of an object in stead of *concept*. In stead of property we will often use *feature* or *category*. It will later become clear that *meaning* of an object can be expressed as the set of categories, that uniquely distinguishes the object from the rest. Also, *image segment*, *referent* and *object* are used interchangeably.

### 1.0.2 Experimental Setup

The experimental setup of the TH experiment consists of a population of virtual agents embodied in robots, situated in real world (figure 1). The robots are equipped with moveable cameras and are connected to a computer for cognitive processing, and to devices for audio in- and output.

The shared environment consists of a magnetic white board, on which various shapes are pasted: colored triangles, circles, rectangles, etc (see figure 2). The environment is altered between games, and new, previously unseen shapes are constantly added.

### 1.0.3 The semiotic square

The problem setting can be represented in simplified form by using a semiotic square, which is the analogue of the meaning triangle mentioned in the introduction (see figure 3).

In some cases, a computer simulation using an artificial world (the so-called GEOM world) will be sufficient for studying the dynamics. In this case, the referent and the image collapse into eachother, and we have a semiotic triangle.

Figure 1: Setup for the talking heads experiment

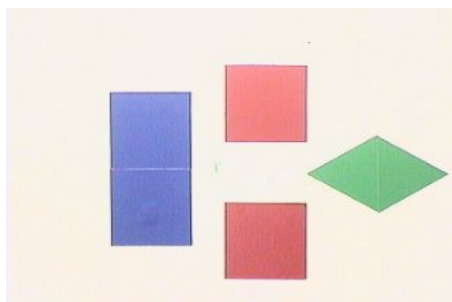

Figure 2: White board with various shapes

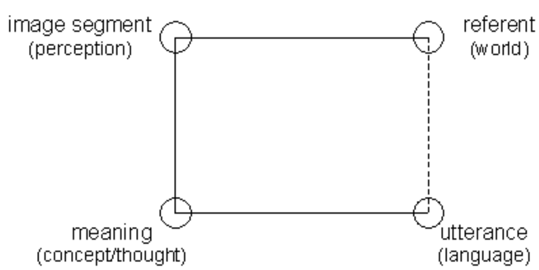


Figure 3: Semiotic square

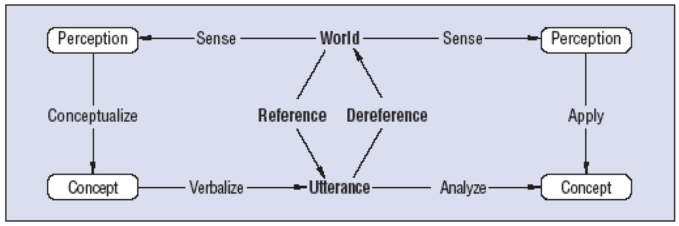


Figure 4: Two joint semiotic squares and interplaying processes

Based on this simple division, we can study the following processes, which are the essential stages in communication: The speaker *perceives* an object in the scene, *conceptualises* it into a meaning, and *verbalises* the meaning into a word/utterance The hearer then *interprets* the utterance to find a meaning in his own conceptual space, and *applies* the meaning on the scene to pick out a referent/object.

### 1.0.4 The model

The TH experiment is an integrated model of perception , conceptualisation and language. For the model to comply with the semiotic square, the architecture of the agents parallels the semiotic square, so (in the simple case) it consists of 3 layers :

- Perceptual layer (for segmenting the image - output are sensory channels)
- Conceptual layer (for making categorical distinctions - output are categories)
- Lexical layer (output are utterances)

The goal is to study the dynamics between those layers, and thereby find so-called 'engineering solutions' to some important philosophical questions. We will of course not resist the temptation to discuss the significance of those engineering solutions for philosophy.

For example, the Gavagai problem, mentioned in the introduction, now formulated in terms of semiotic squares: If you join together the semiotic squares of 2 agents, you can visualize how agents communicate meanings (see figure 4). They can exchange utterances, or point to objects in the world, but they have no direct access to each other's meanings. Neither can the *hearer* skip the meaning of the *speaker* if he wants to know what object the utterance of the *speaker* refers to, because the meaning stands between the utterance and the referent. Therefore, the *hearer* will have to try to guess the meaning of the *speaker*, which is a difficult task, because every agent conceptualises the world autonomously, according to his own personal experience with the world, and his conceptualisation is unknown to other agents The TH model will demonstrate how language helps to coordinate the conceptualisations of different agents.

### 1.0.5 The guessing game

The dynamics are studied in the context of the *guessing game*. This is a game played among two agents, in which the *speaker* chooses a topic from the white board, conceptualises it, and communicates it verbally to the *hearer*. The hearer then has to guess which topic was chosen.. If he fails, speaker and hearer repair their conceptual layers and lexical layers. The process is repeated, and multiple agents act in turn as speakers and hearers, while the elements on the white board are constantly changing. The guessing game is assumed to model the real world situation, in which children learn a language through communicative interaction with other language speakers (e.g. that elephant?).

If in the end a successful communication system emerges, that would show that some kind coherence has developed between the conceptual layers and the lexicons of the different agents. In the dynamic process of playing games, the agents have co-ordinated their meanings and lexicons. It will be shown that both disambiguating cases in the real world and language help in establishing a shared conceptual and lexical system among the agents.

### 1.0.6 Methodology of the paper

Since this is a very complex problem, and different processes interact with each other, we will break up the guessing game into sub-games, and simplify wherever possible. First there is the problem of sensory perception, and segmentation of the image into objects, that will be dealt with in the following section. Then, assuming that segmentation has been solved, we deal with the problem of conceptualization, by means of the discrimination game. Finally, assuming meanings exist and are shared and agreed upon, lexicalization is treated in the context of the naming game. Subsequently, everything is integrated together, and we will look at the interactions and feedback connections between the layers.

## 1.1 Perception

We shall start by discussing the perception layer.

### 1.1.1 Segmentation

The world does not present itself neatly segmented into objects, but as a continuous stream of analogue signals. In order to segment the image of the scene into discrete objects, techniques known from machine vision are used in the TH experiment, such as edge detection, finding junctions, patches of same color, textures, etc. This is the preprocessing stage. Similar processes have been found to be hard-wired in the brain, e.g. orientation-sensitive cells detect edges, center-surround cells detect patches, etc.

As soon as the segments are extracted from the image, the sensory characteristics of each segment are determined. In the TH experiment, the information about a particular dimension is represented by a sensory channel. For the sake of simplicity, we will only use 3 sensory channels here: HPOS, VPOS and GRAY. The value of a sensory channel represents the continuous information, that is made available (to

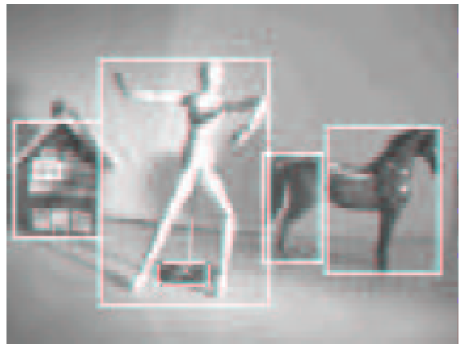

Figure 5: Extracting segments

cognition) by the senses. Here is an example of the sensory channels of an object O1:
O1: [HPOS: 0.89] [VPOS: 0.02] [GRAY: 0.37]
These values are output to the conceptualisation layer.

#### 1.1.2 Context-sensitivity

The values of the sensory channels are scaled relative to the context of the scene in order to obtain for every channel a value between 0 and 1. For example, the value GRAY for each segment is divided by the largest value of GRAY in the scene.

#### 1.1.3 Saliency

The saliency of a channel is defined as the distance (after sensor-scaling) between the perceived values of the topic, and the nearest value of one of the other segments in the scene. The most salient channel is the channel by which it is easiest to distinguish the topic from the context. Agents use saliency to determine what channels to pick for determining a distinctive category set (meaning) of the topic. Saliency helps to restrict the attention to channels that are most effective in communication.

#### 1.1.4 Remark

In the current implementation of the TH experiment the pre-existence of objects/segments is taken for granted. Therefore hearer and speaker can figure out the referent through pointing. But in real life it is not obvious where the boundaries between objects are. For the sake of the philosophical argument it would be desirable not to postulate that objects pre-exist in the external world, but let the agents construct them, just as they construct the categories (see later). This has in fact been done [6]: the segments can be constructed in a task-dependent way - by propagating higher level constraints downward to the sensory layer, telling it how to segment the image (but this is out of the scope of this paper).

## 1.2 Conceptualisation

The conceptual layer receives analogue input signals from the sensory channels of the perceptual layer. It outputs one or more conceptualisations (meanings) of the topic (together with their rankings) to the lexical layer, if it has found any.

In order to conceptualise the scene it is necessary to categorize the input from the sensory channels. The continuous values are transformed into discrete categories. This results in a compact representation of the segments in the scene.

Categorization is done by dividing up the ranges of the sensory channels into regions, and assigning a different category to each region. For example, a value 0.18 of the horizontal position (HPOS) sensory channel can be categorized as [HPOS 0.0-0.25], which means that it lies within the first quarter of the HPOS sensory channel. One might interpret this as the feature LEFT. Categories are equivalent to features. Categories may also be formed from for combinations of channels. For example, there could be a category ORANGE, which means that the sensory channels for RED, GREEN and BLUE lie within a certain range, or BIG, indicating a certain range for the sensory channels of WIDTH and HEIGTH. If it is needed to perform finer distinctions, the categories can be refined by subdividing the regions (splitting them into halves).

### 1.2.1 The discrimination game

The categories are formed under the pressure of a task. Just as in real life conceptualisation is driven by the need for survival, here conceptualisation is driven by the need to distinguish between objects. The task is modeled by a discrimination game. At a later stage we will replace the task by a communicative task, which is modeled by the so-called guessing game. The discrimination game is the first stage of the guessing game.

The discrimination game is played by a single agent. The agent perceives the scene, which consists of various colored shapes pasted on the white board, and chooses at random a topic from among the segments. (We assume segmentation was completed successfully by the perceptual layer). He then categorizes the objects in the scene by using his existing set of categories, and comes up with a category or conjunction of categories that distinguishes the topic from the other objects in the scene. For the agent, this set of distinguishing categories (or features) constitutes the *meaning* of the topic. The game is successful if the agent has found (a set of) distinctive categories. Otherwise the game fails, and the agent is urged to produce more fine-tuned categories (we'll see later how), and try again.

This illustrates the point made in the introduction, that meaning can be most efficiently represented by distinguishing features. The meaning of the topic is uniquely specified if it can be distinguished from all the other objects, and that is all that is needed for the task.

### 1.2.2 Categories and features are imposed, and task-driven

We'll discuss later how new categories are created. For now it is important to realize that categories are only preserved if they are useful for distinguishing objects in the scene, so only if they are meaningful for the task at hand. So if the agent happens to

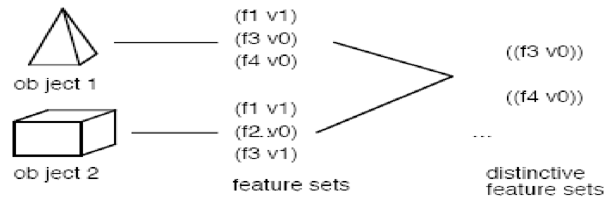


Figure 6: Distinctive feature sets

encounter a scene in which there is a distinction between a small topic, and all large objects in the background, this is enough to create (or rather strengthen) a category LARGE. LARGE is only meaningful in opposition to SMALL. If all the objects the agent will ever encounter in his life are the same size, the agent will never create the categories LARGE and SMALL, or, if they already existed, delete them, because they have become redundant for the task.

An important consequence is that categories (features) and concepts (distinctive feature sets) are actively constructed; the distinction is imposed on the scene instead of recognized. Categories are creations of the agent, driven by a particular task. The categories do not pre-exist in the external world, they are not a-priori, they are not learned, but imposed upon the scene. The segments do not have intrinsic features, but rather the features emerge as a side-effect of the discrimination task. This is precisely the theory of meaning outlined in the introduction.

In the introduction we already mentioned examples from neurobiology that make the same point: it has been shown that the rat brain develops sensitivity to distinguish fine grained categories within a certain modality (e.g. loudness, frequency) depending both on the external stimuli the rat receives during rearing, and on whether or not their distinctive capacity was rewarded in the form of food [3].

### 1.2.3 Selectionism in the TH experiment

We still need to implement a mechanism that governs category formation and pruning. A model for conceptualisation has to deal with the paradox of learning new concepts, mentioned in the introduction. There is the chicken and egg problem formulated by Fodor: how can one acquire a concept before one has some hypothesis about it? For the TH experiment, the question can be restated as a question about the formation of new categories.

We argued that a purely *bottom-up* empiricist approach cannot account for language or concept acquisition. In the first place, it suffers from the aforementioned problem, that the new category needs to be accomodated by the existing categorizations, which requires interaction with *top-down* input. Moreover, with learning by induction, categories become stable and learned only after exposure to a large number of cases, while in a real situation one observes that categories appear instantaneously.

We promised in the introduction that the selectionist approach can deal with these

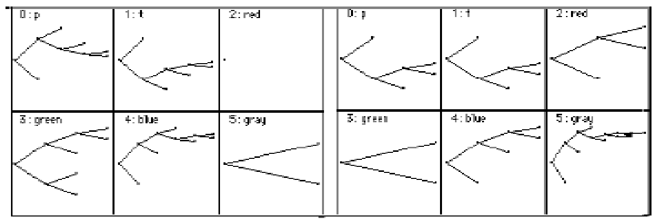


Figure 7: Discrimination trees available for two agents

problems. We will now show how selectionism is implemented in the TH experiment. As explained before, selectionism is characterized by the following principles:

- A growth process, which generates categories randomly, even in the absence of input.
- A mechanism for preservation of categories, which allows for gradual build-up of more complex and refined categories. (→ *scores, maintain history*)
- A selectionist force, which keeps the random growth process under control, by determining which categories are maintained, and which are pruned. (→ *environment*)

In order to categorize, agents grow discrimination trees (see figure 7), a separate tree for every sensory channel or conjunction of sensory channels that might yield a distinctive feature set. A discrimination tree is a kind of decision tree, which categorizes the input from the sensory channel.

When continuous input from the perceptual layer comes in, it is filtered along the discrimination trees. Starting from the top, the input trickles down the tree, until the value ends up in a single most specific category, while at every node of the tree it is decided whether the value belongs to the category or its opponent. Nodes are also called categorizers, because they force the categorization of the input value. If for a particular discrimination tree the value of the topic ends up in a different node than the values of all the other segments in the context, then the discrimination tree is said to have discriminated the topic from the context, and the meaning of the topic is identified with the node it ended on. Of course, there might be more than a single discrimination tree that successfully discriminates the topic from the context, in which case the categories enter a competition for the claim of being the correct meaning of the topic.

The selectionist growth process is implemented by letting a single discrimination tree grow in a random fashion, each time a discrimination game has failed. When the tree grows, the range of a terminal node is divided into two subranges, creating two new subnodes. The categorizers (nodes) that were effective in discriminating in the recent past have precedence, so the process is not completely random.

The agent starts with a single top level categorizing node for each discrimination tree. Initially there will be an explosive growth of the trees, because there is a high failure rate, since the agents have not yet developed enough categories to discriminate. (cf. explosion of synaptic connections in young child). Once discriminative success is steady, the discrimination trees will stabilize.

### 1.2.4 No learning, no induction

We see that categories may form without any inductive learning: in fact a category may be created through random growth without the agent ever seeing an example. This explains the single instance learning of concepts often seen with children. The environment is imperative only for preserving and strengthening categories, not for creating categories. We shall discuss the role of the environment shortly.

It follows also that categories (and concepts) need not be innate; they are initially created by the random growth process. Selectionism may be brought up as a counter argument against the Poverty of Stimulus (POS) argument, often produced by generative grammaticians to defend the innateness of a Universal Grammar. The POS argument says that it would be impossible with the small amount of input children hear, to learn the grammar of a language in such a short time. However, a random growth process of hypothetical grammar 'rules', would shorten the learning track significantly.

As remarked in the introduction, this is a fitting model of the growth process in the brain. Synapses in the developing brain form many spurious random connections. Only after the initial connections have already been made, the environment starts exerting its influence on the brain, by causing neurons to be pruned if they don't receive any trophic support through electrical stimulation from input signals. The process of growth of discrimination trees thus describes quite accurately the plasticity observed in the developing brain.

### 1.2.5 No conceptual coordination before language

Every agent grows his own discrimination trees, and these may be different due to different developmental histories (sensual experiences) of the agents. That means that before there is any feedback from language, individual agents conceptualise the world differently, and their conceptualisations are not coordinated (yet).

The same phenomenon is observed with young children. Before they have mastered adult language they often have a very original way to conceptualise the world. But as soon as they become fluent in language they loose their original perspective. Experimental work on this subject, focused on conception of tense and space, has been done by Slobin in [5].

### 1.2.6 Competition between meanings

All the way, use and success of the (conjunctions of ) categories associated with every discrimation tree are monitored. (In fact, meaning is a more appropriate term than category, so we'll use that in the following). Every meaning maintains a score for how many times it has been used in the game, and another score proportional to its success in the discrimination game. If the game was successful, the score of the winning meaning is increased by a small amount. If there were other competing meanings, that also successfully discriminated the topic from the context, their scores are decreased - this is called latent inhibition. The meaning that has the highest score wins the competition, so it will get an even higher score. This winner-takes-all process sets in motion a positive feedback loop: the winning meaning obtains higher and higher scores, while the scores of the loosing meanings decrease more and more, until eventually they are 0, after which the category is pruned. This way conventional meanings entrench, like the emergence of a broad walking path in the bushes.

In fact, at a later stage, when we are going to couple the conceptual layer to the lexical layer, there is no direct winner after conceptualisation: all the possible conceptualisations (meanings) will be offered to the verbalisation process, together with their ranking. Now the fact whether or not the meaning can be verbalised will play a role in deciding which is the winning meaning. In that case, the main selectionist force comes from language. For the meantime, we will not consider the effect of language in our model.

### 1.2.7 Example of a discrimination game

Refer to figure 2: There are four objects; the top square (object-2) is selected as the topic. Following is the automatically generated comment on the game:

```
   Game 22
a1 segments the scene into 4 objects:
object-0, object-1, object-2, object-3
a1 chooses object-2 as the topic
a1 categorizes the topic as [VPOS 0.0-0.5] (score 0.57)
The discrimination game succeeds.
```

The game succeeds because the agent was able to find a distinctive value. If the top square would have been yellow, then YELLOW would have also been an option, and the meaning with the highest score would have been chosen.

### 1.2.8 The role of the environment

So far, we have observed that categories and meanings form driven by a task, the discrimination game. But their is also the influence from the environment, which in a more indirect way drives category formation. This is the selectionist force. In order for a meaning or a category to increase its score, or not to be pruned, it needs disambiguating cases from the environment, so it has the opportunity to win the game. For example, if a sensory channel has the same data for every possible segment the agent ever sees, the agent is not going to find a distinctive category based on that channel, so the tree for that channel will not develop, and eventually even the randomly generated categories for that channel will die out: for example, if all objects would have the same color, all the color categories would eventually die out. (e.g. see what happened in figure 7 with RED)

### 1.2.9 Incremental refinement of concepts

The selectionist force exerted by the environment is responsible for preservation and pruning of the discriminations trees. It is also the environment that in the end determines to what extend meanings (categories) are refined. If the environment offers cases that can only be disambiguated on the basis of a slightly different value of a sensory channel, then very specific categories may develop. But this will only happen if the coarsest categories are formed first: the finer categories need to grow out of the coarse categories in the tree. So, one first needs to acquire a category [HPOS 0.0-0.5] before one can acquire a category [HPOS 0.0-0.025].

### 1.2.10 Cognitive development

We see the same with children: they need to develop rough concepts before they can grasp finer concepts. They need to learn what is a tree before they can learn what is a branch. The metaphor of the growing discrimination tree demonstrates how ever finer distinctions can be bootstrapped, how the learning paradox can be solved: the only way it will work is by starting with very crude concepts, and subsequently incremental fine-tuning of concepts.

There are theories in developmental psychology that say exactly this. In the words of William James [1] , the new-born baby perceives the world as one 'booming buzzing confusion'. The developmental psychologist Daniel Stern argues in [7] that the first rough concepts of the newborn infant are of the mother (which encompasses the entire external world, both dangerous and giving) and of the self (which encompasses the drives and needs of the body). Concepts are differentiated out of eachother. Out of the concept of the mother, concepts of other people, objects and spaces are differentiated gradually and incrementally, while out of the concept of the self, the concepts of body parts and individual needs are differentiated. The TH experiment thus proves to be valuable also when it comes to modelling cognitive development.

## 1.3 lexicalization

Until now, we have discussed the implementation of the perceptual layer (paralleling the side of the semiotic square handling the relation between the real world and the perceived image) and the conceptual layer (dealing with the side of the semiotic square that relates the perceived image to the conceptualisation or meaning). Now we need to look at the final side of the semiotic square and the implementation of the lexical layer: the relation between *meaning* and *utterance*. This will be done in the context of the *Naming game*.

### 1.3.1 Goals of the Naming game

How do agents lexicalize meaning to produce an utterance? And, how do they interpret an utterance to produce a meaning? The agents need to have a way of relating form and meaning, and using these to produce and understand others' utterances.

But, how can an agent understand another agent's utterance? To simplify, at this stage we will assume that all agents share the same meanings, thus avoiding the

'Gavagai' problem. This assumption will be taken out in the next chapter, where we couple the Naming game directly to the Discrimination game. Associations between meanings and word forms are stored in the so-called *lexicon* of the agents. The purpose of the naming game is to demonstrate how it is possible that a *shared* lexicon emerges so that agents can engage in co-ordinated, social interactions.

It is not trivial to develop a shared lexicon because there are multiple agents and each can invent words for the same meanings or concepts. This would result in unavoidable *synonymy* (multiple words for one meaning) and *ambiguity* (multiple meanings for one word). If there was not some sort of dynamic interplay - learning and changing what you already know, based on information from the environment (other agents) - each agent would contentedly continue creating their own words to express some meaning, resulting in total synonymy and ambiguity. No one would understand each other.

The whole idea of the Naming game is to show how these sources of incoherence - synonymy and ambiguity - can be restrained. Even if it can be shown that a shared lexicon can emerge without any prior design or global co-ordination, can it persist in the population and be transferred from generation to generation? Can the framework account for forever changing sets of meanings - how are these understood and how do they proliferate? Can new agents enter the population and acquire the existing language, and can agents leave a population without destabilizing the whole language? The current section will show how all this can be done, with a very simple architecture.

### 1.3.2 Inventing a Lexicon

The relation between meaning and utterance is discussed and reconstructed in the form of the Naming Game. The Naming game in fact makes up the second part of the guessing game (the discrimination game constitutes the first part).

In the Naming Game, played by *speaker* and *hearer*, the *speaker* selects a meaning from the shared repertoire of meanings, looks up a possible word for this meaning in its own lexicon and transmits this word to the *hearer*. The *hearer* interprets the word by looking it up in its own lexicon and transmitting the meaning he himself obtained. The game succeeds if this meaning matches the one intended by the *speaker*, otherwise it fails. If it fails, then the *speaker* communicates the meaning directly - the *hearer* adds this new form-meaning pair to his lexicon or creates a new one if the word does not yet exist in his lexicon.

### 1.3.3 Representing lexical associations

Every agent has a private lexicon, which is stored in *associative memory*. The lexicon is made up of *form-meaning pairs*. Since the agent is going to be unsure about the meaning of a certain form at some stage, agents must also store alternative meanings for one form and different forms for one meaning. In this respect they can also handle *ambiguity* and *synonymy*, respectively.

The agents are thus equipped with a sort of lexical matrix (e.g. table 1.3.3). These hold form-meaning pairs. From table 1.3.3, you can see that the rows represent synonymy, and the columns, ambiguity.

| | coba | zapo | bila | pama | wabidu | limiri |
|---|---|---|---|---|---|---|
| [DARK] | 0.3 | 0.2 | 0.1 | 0.8 | - | - |
| [LARGE] | - | - | - | 0.5 | 0.3 | 0.6 |

Table 1: Lexical matrix

Each form-meaning pair has an associated *score* (between 0 and 1; table 1.3.3). The scores are specific to an agent, and only based on local interactions with other agents, highlighting the principle that no agent has a complete overview of another's lexicon.

### 1.3.4 Using knowledge from other agents: updating scores

A speaker can only transmit one word for a particular meaning - it chooses the one with the highest associated score. Each agent monitors how successful a particular word-meaning association is using the score. The more a word is used, the higher the score. The score on one association is increased if it was successful; at the same time, the score on competing associations are decreased - *lateral inhibition*. Success of a form-meaning pair causes the *speaker* to decrement competing associations that use another form for the same meaning (*synonyms*); the *hearer* at the same time decrements competing associations that relate another meaning to the same form (*ambiguous* items). This way, *synonymy* and *ambiguity* are dampened. If the game fails, the scores on the tested associations go down.

### 1.3.5 Inventing and acquiring words

Agents are allowed to *invent* new words, if they do not have a word for a meaning they want to communicate. They can also add new words to their lexicon if they hear a word that they have not come across before. Both these happen with a certain probability, getting across how 'free' the agent feels to extend the lexicon (word creation rate) and how critical they are in accepting other agents' linguistic influence (word absorption rate).

### 1.3.6 The group lexicon

The lexicons are different in every agent, since after many rounds of the Naming game (involving multiple agents) they have had different experiences and encounters. Form-meaning associations propagate through the population as more and more games take place. The *group lexicon* can be observed by inspecting the internal states of all the agents of the population. It lists for every meaning the dominant word form in the population, and the frequency with which the dominant association is preferred by the agents within the population. As a form-meaning association spreads in the population, eventually, its frequency will reach100%. Once all form-meaning associations in the group lexicon have a frequency of 100%, a *consensus* is reached.

The group lexicon is neither known by the individual agents, nor stored anywhere in the system. But, on the macroscopic level, the group lexicon is what characterises the language. This is in line with the idea, mentioned in the introduction, that language should be studied at the level of the population and not at that of the individual: language is a system that is dynamic, adaptive, and heterogeneous, rather than static and homogeneous.

When the consensus is reached, the group lexicon does not change any more (unless new agents or meanings enter the population; see later). There are many associations in the individual lexicon of an agent that did not make it into the shared lexicon (e.g. ones that were created and not picked up, or ones whose score became 0 because another word became dominant) - perhaps a pruning mechanism could be implemented here, which would take these out of memory. A shared lexicon has emerged.

### 1.3.7 Measuring the success of the communication system

The easiest measure for the coherence of the lexicons of the agents is *communicative success*, which is the average proportion of successful Naming games. When the population (of both the meanings and the agents) is larger, it takes longer to reach total *communicative success* (success reaches 100% when a shared lexicon emerges). This is measured across consecutive sets of games.

*Lexical coherence* is another measure used to determine to what extent the agents share the same lexicon. It is the average of the frequencies of all form-meaning pairs in the group lexicon. If all agents prefer the same form-meaning pairs for all meanings, lexical coherence is 100

Even if communicative success is total, this does not imply that the agents have the same lexicon. The hearer has to associate the same form used by the speaker for the same meaning to reach success - this does not mean that the hearer prefers to use the same form for the same meaning, thus synonyms can occur.

### 1.3.8 Testing the system in growing populations

The architecture proves to handle variations in the set of meanings when new meanings are introduced into the population. It can also handle new members in the population - even if the new members create new words, these are not likely to survive, especially if certain associations are entrenched in the population already - the new agents will adopt the words used abundantly in their environment. This can be seen in children, who first pick up the words they hear more frequently. An even more illuminating example is that of twins - it is often seen that they develop their own language to communicate with each other, yet when continuously in the company of other individuals, they will also adopt the language spoken in their environment.

The maintenance of the lexicon in increasing populations indicates that the lexicon can be preserved from one generation to the next. In real language populations, as well as entering a population, members also leave it (migration). The model again deals with this - the removal of agents does not break down the system. The transmission of the lexicon is done entirely culturally. There is no prior knowledge of the new members of the population.

### 1.3.9 Winner-takes-all

One appealing characteristic of this model is that there is no central supervisor to reach a consensus. Rather it is a subtle interaction between language use and each agent's adaptation to the language heard in the environment that results in a shared lexicon. If an association becomes preferred by a group within the population, then its frequency of use goes up and its score goes up across the group. The associations gaining impetus create a snowball effect as it proliferates through the rest of the population. It is a group co-ordination process.

This collective behaviour and self-organisation is also abundant in nature. Complexity arises from a simple set of mechanisms and behaviours. For example, the formation of paths in an ant society also has the characteristics of no central planning, no critical dependence on a single element, resilience to influx or outflux of elements, and adaptation to changing circumstances.

### 1.3.10 Spatially distributed naming games

Language is as such a complex adaptive system. Dynamics of language are related to dynamics of the underlying population. If populations change too quickly, the language can destabilize (e.g. English is said to have lost its case system during the Black Plague). Also, human populations mix and impact each other's languages (take on structures of other languages e.g. English being influenced by the Viking invasions).

Implementation of the naming game that takes into account spatial distributions shows that language game models may be used to study characteristics of language contact and change as just described. Agents closer to each other develop a stable language different from that of other subgroups, yet they still develop a second, weaker language - an interlingua, that is shared among all the clusters. This illustrates how dialects come about, dependent on rich communication within communities, and less communication between communities

So, with a simple set of behaviour rule and associative memory, a shared repertoire can arise, coming into an equilibrium state and getting transmitted from one generation to the next. It happens without a central co-ordinator or prior knowledge (not innate). The population also copes with new and obsolete meanings (expansion and contraction of the lexicon) and migration of population members. Generative grammars look at single speakers in a homogeneous language community, whereas this approach, which looks at the evolution of language in inhomogeneous communities - can study language change and contact.

## 1.4 Coupling the layers

### 1.4.1 The guessing game

The guessing game couples the naming game to the discrimination game. This time, there is no direct feedback anymore about successful discrimination, but only about successful communication. It is played among two agents. The purpose of the guessing game is for the hearer to guess the meaning of the utterance produced by the speaker. Speaker and hearer take turns. The speaker selects a topic from the white board; he conceptualizes it, and looks up the words associated with the possible meanings in his lexicon. He then transmits the word with the highest score to the hearer.

The hearer, in his turn, looks up the possible meanings of the word in his lexicon, and then filters the objects in the scene according to these meanings (applies the meaning to the scene), in order to pick out a single referent. If he finds a unique referent, he points at it. If the game fails, e.g. the hearer pointed to the wrong object, then the speaker gives feedback by pointing to the correct topic. In that case, both the hearer and the speaker repair their lexicon and their categories, in order to make a better guess in the next game.

### 1.4.2 A simple test simulation

As before, we will work from a simple to a complex situation: we shall assume here that perceived images of both agents are the same. So we replace the real world with an artificial image from the GEOM world, thereby skipping the perceptual layer. In the next section, when we deal with grounding, this assumption is taken away.

The first test is done with only two agents, which consider only the most salient channel for conceptualization. This results in a situation in which both agents make the same conceptualizations, and create the same lexicons, so there is no synonymy and no ambiguity (the explanation will be given later). In this simple case it can be seen that communicative success increases. Initially, only the most basic distinction are lexicalized by both agents, but after 500 games more and more refined categories are created by both agents, until it stabilizes when communicative success reaches 100%.

### 1.4.3 No direct feedback, reinforcement learning

In contrast to the discrimination game, there is no direct feedback anymore about meanings. In the guessing game the sensory input has to traverse two layers before the speaker can know whether his conceptualization is successfull or not.

To maximize success in the guessing game, the speaker will prefer not simply the category which has the highest score, but he will also prefer categories that are already lexicalized, and he will also consider the scores of the associations in the lexicon. At the level of the conceptual layer the speaker still doesn't know what is the most appropriate conceptualization from the viewpoint of language success; he has to know the outcome of the lexical layer. Which conceptualization is finally considered as the best one will depend on the lexical layer. Therefore, the speaker must transfer all alternative conceptualizations from the conceptual layer to the lexical layer.

The hearer has a similar problem: he can only know which meaning was intended by the speaker after traversing from the lexical layer via the conceptual layer to the scene: the only feedback the hearer gets is from pointing, so only after he has applied the meaning to pick out an object from the scene, he will know whether his interpretation of the word was correct or not. So, the context of the scene helps the hearer to disambiguate words and meanings. Disambiguating cases, in particular, will help the hearer to coordinate his conceptualizations with those of the speaker. From the different alternatives, the hearer will select only those meanings that pick out a unique referent. Like the speaker, the hearer will consider both the scores of the association, and of the category.

If there is ambiguity in the scene, more than one way to conceptualize reality, the hearer can possibly guess the wrong meaning. This is the now familiar Gavagai problem: 'does Gavagai mean rabbit, fur, or moving thing? everything could fit...' This explains why children make over- and under-extensions. For example, 'orange' could be a ball, a doorknob - for a child it can refer to any round thing.. Until, at a certain age, the child has encountered enough disambiguating cases, so that they have coordinated their concepts with the rest of the language community.

### 1.4.4 Updating the scores

The scores are updated in the same manner as in the discrimination game and in the naming game - through lateral inhibition. But now there is also feedback from language on the scores of the categories, because communicative success determines which the winning category is. When a category is used as part of the communication its use counter goes up. Successful use of a word in the guessing game causes both the score of the association in the lexicon, and the success score of the associated category to increase. This gives rise to a coordination between conceptualization and lexicalisations within the agent, because categorizations will be preferred that lead to successful language games.

### 1.4.5 The Sapir-Whorf hypothesis

This is how language influences category formation - this is the Sapir-Whorf hypothesis! Humans adjust not merely their language, but also their concepts, if they are concerned that they are unable to communicate them. It means that the concepts of an individual agent are shaped in accordance with the language that is spoken in the language community, and at the same time it explains why conceptualization (and not merely the lexicon) becomes coherent among agents of the same language community. The effect of learning a language on the mind of a child is that his concepts grow to be more similar to those of the adult language users. Language and concepts co-evolve. xxx if updating scores does not seem a very intuitive, think of it as strengthening and weakening the weights of the neural connections between language utterances and concepts.

### 1.4.6 Updating the lexicon

The Guessing game can fail for many different reasons. In one case the hearer doesn't know the word spoken by the speaker, so the game fails. The hearer then adds the word to his lexicon, paired with his conceptualization of the referent that the speaker pointed to, after the game failed.

Another reason for failure is that the hearer cannot guess a single unique referent from the meaning he found in his lexicon. Also in this case, the hearer conceptualizes the object pointed at, and associates his conceptualization with the word uttered by the speaker.

### 1.4.7 Scaling up

So far, we assumed only two agents. If we scale up the size of the population, allowing for more than 2 agents, synonyms start entering the language, because many agents may invent a new word for the same meaning not knowing that the word already exists in the community. This doesn't happen in a population of two agents, because every new word is immediately shared.

Even with a larger population, the simulations show that synonymy gets damped. The reason can be found in the dynamics of the system: the winner take all principle and lateral inhibition of the scores. After a successful game, the speaker increases the score of the winning lexicalization, and lateral inhibition pushes down alternative lexicalizations. Thus a single word becomes dominant for a certain meaning in the entire population, and synonymy is dampened. New words hardly get a chance to propagate in the language community because they are seldom selected to be used, and they therefore die out.

We further assumed that only the most salient sensory channel was considered for conceptualization, preventing ambiguity to enter the system. If we allow for other channels to contribute to conceptualization, ambiguity may enter the language, because no longer the hearer can guess with 100% certainty the intended meaning after the speaker points to the referent. Thus, multiple meanings will start circulating for the same word, and ambiguity enters.

Language helps to dampen ambiguity: if the speaker has two alternative conceptualizations for the topic, but only one of them is lexicalized, he will choose that one. This is another example of how lexicalization influences which conceptualization is preferred, the Sapir-Whorf hypothesis. Also in this case, lateral inhibition pushes down the scores of alternative meanings, resulting in dampening of ambiguity.

## 1.5 Semiotic dynamics

Altogether we see, that even in the case that we allow for more than two agents, and when we allow agents to consider all channels, synonymy and ambiguity will be damped, and the population will still bootstrap shared concepts and a shared lexicon.

Of course, we have only tried to make some aspects of the dynamical behavior of the system plausible. In fact, the system's dynamic behavior can be studied with some mathematical rigory, within the mathematical framework of complex non-linear dynamics. Non-linear dynamical systems are known from many other fields, e.g. ecology, and they are studied extensively. Such systems are known under some circumstances to show unpredicatable, chaotic behavior. The group of Luc Steels introduced the new field of 'semiotic dynamics', the study of the dynamic interactions between language, meaning, and referents. One of the research questions that is asked in semiotic dynamics, is what are the conditions that are needed for a shared language and shared concepts to emerge. Obviously, non-linear dynamics poses a much bigger challenge for practicioners of linguistics than the traditional formal logic framework, within which linguistics were usually studied. The added complexity is a result of the two-way interaction that is believed to exist between language, thought and perception, and of the interaction between language users. In contrast, formal logics and generative gram-

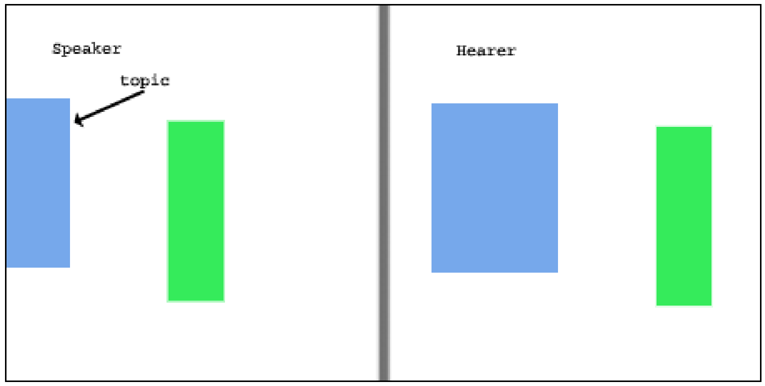


Figure 8: Example of scene causing perceptual ambiguity

mar present a too simplified picture of a one-way interaction, and a static language, and static meanings. Therefore their mathematical models may be much easier, but theynot adequate.

### 1.6 Grounding

Up until now, all the agents in the game have had the same 'picture' or view of the scene - so they segment the scene in the same way, finding the same salient channels. This is not the way it would happen in the real world. Depending on distance, relative location and angle of view, agents would perceive the scene in different ways.

Now the cognitive architectures of the Talking Heads need to be grounded in the real world. This means that instead of interacting in the artificial GEOM environment, where the images are given, they are situated in the real world environment (in front of a whiteboard of shapes), thus everything - features, concepts, language - must be related to the real world.

The previous section discussed coupling the conceptual layer to the lexical layer, and skipped the perceptual layer; now the final piece is to link the perceptual layer to the other layers of the agent.

This introduces additional difficulties, and additional sources of ambiguity. To see why, consider the game in figure 8.

Because the speaker looks from a different position, his entire view is shifted to the left, and part of the left rectangle is out of view. The speaker conceptualizes the topic as [HPOS 0.0-0.25], because for the speaker the horizontal position is the most salient distinctive feature. In contrast, for the hearer [WIDTH] is the most salient distinction. As a consequence, the hearer will adopt a different meaning for the same word, and subsequently both meanings will enter a competition.

### 1.6.1 Motor action and speech

Not only (imperfect) perception influences category formation, as we just saw, but also motoric action influences conceptualization. For example, if the motor skills of the speaker are not well-tuned, his pointing will be not be so accurate. This is another source for ambiguity.

An even more important motoric behaviour is speech production (not implemented in the TH experiment). In a community of language learners poor speech production is yet another source of ambiguity. Altogether, we see that it is not enough to only consider passive perception of the world to explain conceptualization, but the way the agents act upon the world must also be taken into consideration.

### 1.6.2 Why do real world experiments with robots?

Both the extra difficulties that arise from perception, and the difficulties that arise from imprecise motor action illustrate the need for doing experiments with real agents (robots) in a real world. To really understand how the whole communication system is bootstrapped, we must also understand how the system can cope with noisy and imperfect situations (robustness). We also need to create the opportunity for the agents to act upon the world, because learning concepts is a matter of active interplay with the world. We call this grounded cognition, because cognition, concepts and language are grounded in the real world.

### 1.6.3 Embodied cognition

The only motor action implemented in the TH experiment is pointing, but pointing is only the tip of the iceberg. In fact, it is true that the pre-lingual child learns many concepts from non-verbal interaction with the environment.

When the pre-lingual child touches everything he finds, and puts it in his mouth, this is his way of learning concepts by actively interacting with the objects in his environment. Before he has language, only direct feedback will help him to define the meaning of things. You may have noticed how a small child drops a spoon several times on the floor: she is actively engaged in exploring how spoons behave upon her action. She might be looking for disambiguating cases. For the child, the meaning of the objects is determined solely by how they react to the child's actions. This is so, because her physical interactions with the object, the way she can use the object and operate on it, are the only things that matter from the point of view of adaptation and survival. The bodily interaction with the objects defines how they are conceptualized [2]. As a consequence, concepts are very much tied to the body, and to the motor-actions that the child knows to perform at a certain age. This is called embodied cognition.

### 1.6.4 Coordination co-evolves together with perceptual concepts and with language

At the same time that the child learns to perceive distinctions, he learns to coordinate his actions in accordance with those finer distinctions. Again, this is a two-way process:

without fine motor coordination, including coordination of the eye muscles, the child cannot develop fine-tuned concepts, and without fine perceptive distinctions, the child will not be motivated to fine-tune his coordination.

This suggests that, together with conceptualization, also coordination must be bootstrapped, and that coordination co-evolves with conceptualization and (later) speech. This makes for a single adaptive dynamic system involving motor coordination, concepts and language.

### 1.6.5 Motor concepts

There are parallels between what we shall call 'motor concepts' and perceptual concepts: A motor concept is a plan for coordination of a motor action . Like the way perceptual concepts are organized and grow in discrimination trees, also a fine-tuned motor concept must grow/differentiate out of a crude motor concept, and also motor concepts will be stored in some kind of tree structure.

The other parallel is that, in the same way that you can form sentences by linearly combining simple words, and in the way that complex concepts can be combined out of simple concepts, complex motor behaviour, like reaching for something, or playing tennis, is constructed by chaining together simple motor concepts, like small arm movements.

### 1.6.6 Piaget and the sensorimotor stage

This was already acknowledged by Piaget, in [4]. Piaget distinguished 4 main stages in cognitive development, of which the sensorymotor stage (age 0-2 yrs) is the first one, coinciding with the pre-lingual stage.

Piaget explains that pre-verbal, sensory-motor intelligence originates from bodily interaction with the environment. He also notes that increasingly complex motor behaviour is differentiated out of primitive motor behaviour. The simplest coordinations are reflexes; in the following sub-stage the infant begins to chain its reflexes together into primary circular reactions (movements originating and directed to the self), which are then chained into secondary circular reactions, etc. [4]

Evidently, Piaget considered motor concepts to be an integral part of intelligence, required to develop perceptual concepts, and he realized that concepts must develop in an incremental way in order to be bootstrapped.

### 1.6.7 Does language have a monopoly on giving feedback on concepts?

From the TH experiment it may seem that language is the only means by which concepts can be coordinated and shared among agents in a community. One may rightly object, that pre-lingual children also have shared concepts to some degree, and so do some species of animals. There must are other factors that cause concepts to become shared. One, mentioned before, is the fact that all humans have the same body, which to some extend constrains the concepts that are formed, and makes them similar for every person.

Alternatively, non-linguistic conventions provide feedback, and thereby replace the role of language for the pre-lingual child, shaping the child's cultural concepts. For example, when a child kicks a ball he is encouraged to do so. But when he kicks an orange, his parents will probably punish him. This kind of indirect feedback, acts like language, in that it leads the child's attention to the difference between oranges and balls, and causes him to correct his conceptualization of an orange. In a similar way, the pre-lingual child learns the difference between a book and a notebook from feedback about how these objects are used in the culture: it is forbidden to draw with a marker inside a book, but drawing on a sheet of paper is allowed.

## References

[1] W. James. *The Principles of Psychology.* 1890.

[2] G. Lakoff and M. Johnson. *Philosophy in the Flesh.* Basic Books, New York, 1999.

[3] M. M. Merzenich. Basal forebrain stimulation changes cortical sensitivities to complex sound. *Neuroreport*, 2001.

[4] J. Piaget. *The Origins of Intelligence in Children.* Norton, New York, 1952.

[5] D Slobin. From thought and language to thinking for speaking. *?*

[6] Luc Steels. *The Talking Heads Experiment. Words and Meanings.* VUB, Brussels, 1999.

[7] D. Stern. *The Interpersonal World of the Infant.* 1985.