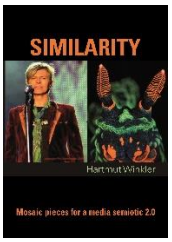
Taken from the book:
Winkler, Hartmut: Similarity. Mosaic Pieces for a Media Semiotic 2.0
Online publication 2024
https://homepages.uni-paderborn.de/winkler/Winkler--Similarity.pdf
First published in German: W., H.: Ähnlichkeit. Berlin: Kadmos 2021
The German text is available online:
https://homepages.uni-paderborn.de/winkler/Winkler--Aehnlichkeit.pdf



4 First Sketch:

Similarity and Schema Formation in Different Media [1]

1. Intro

I would now like to turn to media and take up what I have already discussed in the previous chapter: There I argued, with regard to language and visual media, that similarity undermines all too clear notions of 'identity' as well as of 'difference.' I would now like to turn this into a positive thing.

Because parallel to this, one can show that similarity itself takes on a function and is of essential importance for different media, not although, but precisely *because* it does not operate with precision, but rather with certain tolerances and in a 'soft' manner.

The focus of what follows will be schema formation. How this relates to the question of similarity will become clear step by step; as in other chapters to follow, I will go through various media complexes; and I would like to begin with certain theses on perception which have already played a role in some of my earlier publications[2] and which I now question again with regard to similarity, as they form the basis of what follows.

2. Perception

Human perception already proceeds according to criteria of similarity. Around 1900, the Gestalt theory but also perception psychologists like Stern or Ebbinghaus emphasized that we would have great difficulties to perceive things at all, to free them from their background, if we would not *recognize* them. Thus, Ebbinghaus describes the situation of an infant lying in its carriage:

> "A very young child looks from a certain place into a certain room. It receives from it a little-structured, diffuse general impression; as often as it repeats its look, this impression repeats itself. Now the child is pushed by the mother in its carriage into a neighboring room; in the main, another overall impression takes the place of the first. Yet the mother and the carriage have remained the same. The optical stimuli emanating from them thus find their possible material and mental effect already somewhat predetermined [...]; the remaining, modified partial stimuli lack this dual foundation. [...] On the one hand, the impression stemming from the sight of the mother thus comes about more and more easily; on the other hand, it tears itself more and more loose from the various diffuse backgrounds in which it was originally absorbed: the view of the mother becomes an ever more independent member of the respective overall impression."[3]

---

[1] This fourth chapter is based on a lecture as well: I presented the central theses at the colloquium 'Similarity' at the University of Tübingen (Germany) in June 2019.

[2] First in: W., H.: Der filmische Raum und der Zuschauer. Apparatus – Semantik – Ideology. Heidelberg: Winter 1992, p. 131f.

[3] Ebbinghaus, Hermann: Grundzüge der Psychologie. Vol. 2, Leipzig: Veit 1913, p. 15 (transl. H. W.).

For the question of similarity, the following is important here: First, that objects of perception are not simply given, but are only slowly, step by step, exposed against their background and constituted as objects. And closely connected to this: That the impressions stabilize only in *repetition.*

Repetition, however, means similarity. Impressions that are similar to one another gain contour and – in tension with the changing contexts – pile up into fixed perceptual concepts. This subsequently allows for recognition; perception is no longer confronted exclusively with something new but can apply what it has already obtained from experience. This includes the fact that memory is always already involved in the process of perception; with the consequence that the respective present perception is by no means only a present perception.

"Does not this amount to saying", Bergson writes around the same time,

> "that distinct perception is brought about by *two opposite currents*, of which the one, centripetal, comes from the external object, and the other, centrifugal, has for its point of departure that which we term 'pure memory'? […] Together, these two currents make up, at their point of confluence, the perception that is distinct and recognized."[4]

What Bergson is explaining here is the fact that perception always has a projective component. The stream of perceptions coming from the outside is opposed by a second stream coming from within; the patterns that make it possible to identify what is perceived are fed by memory.

On the basis of these early psychologies of perception, extended *schema theories* have been developed. These were elaborated mainly by Cognitive Psychology, and they are not without problems.[5] The concept of schema itself, however, is a great gain; first, because, as I will show, starting from perception it can be used in other fields as well; and second, because – similar to similarity – it always admits a certain vagueness. Schemata, as the term itself already implies, do not simply coincide with the schematized. They are more abstract than the schematized; and one would expect them to be stable, but at the same time always dependent on development, always under reconstruction.

Perception, then, *seeks* similarity. It observes similarity and extracts it from the amorphous material that the senses offer. And this begins outside the human sphere already in nature, because many animals have not only innate but also acquired schemata. The process of schema formation itself is inherent and completely automated; and it goes far beyond object recognition, insofar as temporal processes are also stored in process schemata; and finally, there is a transition to nearly all other types of regularities. Schema formation, one might say, is the basic mechanism to which *abstraction* and *form* can ultimately be traced. Let us now pass from perception to media, and first of all to language.

3. Medium Language

In language, its inner structure and its functioning, similarity plays a prominent role, and this on the most different of levels; e.g., when one distinguishes between contiguity and similarity, or when de Saussure contrasts the syntagmatic axis with the 'associative' chains which follow a logic of similarity;[6] or in the different dimensions of linguistic mimesis, which always imply similarity.

---

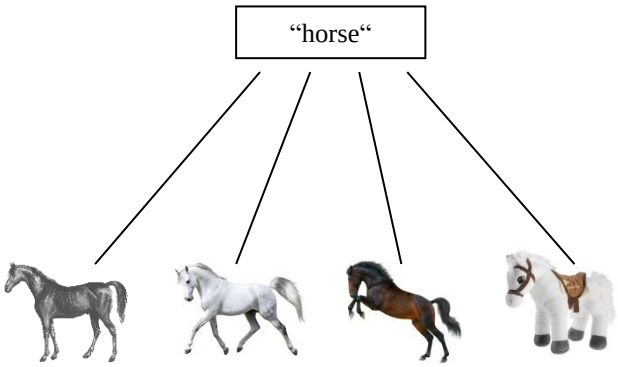
[4] Bergson, Henri: Matter and Memory. London: Allen & Unwin 1970, p. 163 [1896] (emph. H. W.).

[5] I will present some of these problems in my ninth chapter.

[6] De Saussure, Ferdinand: Course in General Linguistics. NY: The Philosophical Library 1959, pp. 128ff. [1916].

Out of all these possibilities, I would like to select just one and pursue it more closely, and this is the mechanism of subsumption, which is of absolutely fundamental importance for the functioning of language, semantics, and the formation of concepts. How concepts arise is a matter of some dispute. Linguistic theory would usually say that words/concepts come about through distinction and that vocabulary is organized according to the rules of contrast and difference.[7] I will show that similarity plays an equally important role; first, because distinction, difference, and contrast are not simply 'the other' of similarity, but remain related to similarity and even to 'identity.' And secondly, quite directly, insofar as concepts always and fundamentally subsume. The term 'horse,' to take an example, does not refer to a single specimen, but to a category or genus that includes the most diverse specimens; and this applies analogously, if we exclude names, to all terms and words.

Language, literature:



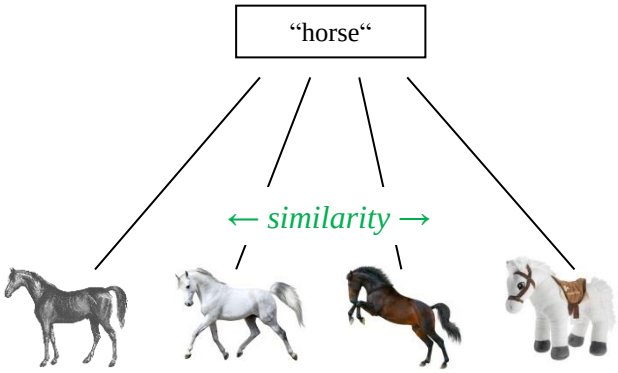
terms / language

perception / referents

Herein lies one of the crucial differences between language and the world, signs and the sphere of referents; both qualitatively as well as with regard to quantities: An infinitely large number of exemplars is confronted with a relatively limited set of terms; resulting in the fact that while the world might not fit into our small heads, language does so very well.

The crucial point now is that the specimens must *resemble* each other in order to fall under a common concept.

Language, literature:



one concept 'horse'

many concrete horses

The formation of the concept 'horse' *owes itself to similarity*. A linguistic community observes that certain animals are particularly similar to each other; and it records this observation by grouping them together and then giving them a group label, the linguistic designation 'horse.'

[7] Lyons, John: Semantics. Vol. 1, NY: Cambridge University Press 1977, pp. 270-335.

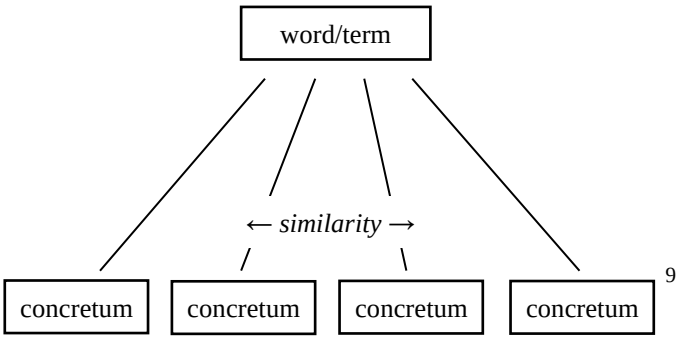
If one goes into detail, things become more complicated. In subsumption, only certain observations, certain similarities are taken into account; those characteristics which, according to the linguistic community, make a horse a horse. In other respects, the specimens can be markedly *dissimilar*, that is, they can differ greatly. In the case of 'horse,' for example, the color is irrelevant; the feature is suppressed for the conceptualization.

The observation of other similarities will accordingly lead to completely different groupings: Rabbits and stones, for example, are drastically different, i.e., quite predominantly *dissimilar*; and yet they may share the characteristic 'gray.' The designation 'gray' holds this one, very particular similarity.

A horse can be a horse and gray at the same time. This makes it clear that the categories overlap so that in semantics, one must assume a complex network of superjacent similarities and contrast axes. Further, one will have to note that similarities in semantics are not always those of perception; and finally, there are many abstract concepts whose world reference is indirect, complicated, or disputed. Not without reason did Walter Benjamin call language a system of *'nonsensuous similarities'*.[8]

Let us keep in mind, then, that that the individual terms subsume; and that this always implies similarity.

Language, literature:



[8] Benjamin, Walter: Lehre vom Ähnlichen. In: Ges. Schriften. vol. II/1, Frankfurt am Main: Suhrkamp 1980, pp. 204-210, here: p. 207 [1933].

[9] When I speak of concreta here, then rather neglectfully; as a technical term 'concretum' already means a concept, a concept, however, which – in contrast to abstracts – denotes things which are sensually experienceable. (Cf: Wikipedia (Germ.): Konkretum; http://de.wikipedia.org/wiki/Konkretum).

4. Visual Media: Photography and Film

As a second media complex, I would like to address the technical images, photography and film, precisely because at first glance, they function completely differently from language and writing. The most obvious dimension of similarity is iconicity, the fact that images, unlike words, 'resemble' what is depicted; but this is not what I am concerned with; rather, I would like to discuss here as well – isolated by way of example – only the problem of subsumption.

The point of photography and film is that they manage without terms and avoid the mechanism of subsumption. Where language uses the word 'table,' i.e., a general concept that initially evokes only a relatively abstract idea and could still designate a multitude of different tables, photography and film present the image of a *single* table. The images of photography and film provide *concretes*.[10]

Language and literature must resort to the means of description in order to concretize concepts; and that means to place further concepts at their sides. Images, on the other hand, always provide the description; the color and texture of the photographed table is always already determined; it is completely impossible for even one of these determinations to remain vague.

At the same time, as I have discussed in the previous chapter, images *weaken* the contrasts and boundaries that seemed guaranteed in language. Photography and film undermine identificatory thinking; clouds can look like horses, a horse can look like a shadow, a shadow can resemble a speck more than any other shadow...

This is one side of technical images. But they have another, unsuspected side, and here it finally becomes interesting for the concept of similarity. For what has been stated above does not mean that every technical image is actually unique, or that every photographed object stands only for itself. For, of course, we can only understand photos or films if we identify or recognize the table – despite all its concreteness – as a table.

---

[10] Fig.: Pixabay.

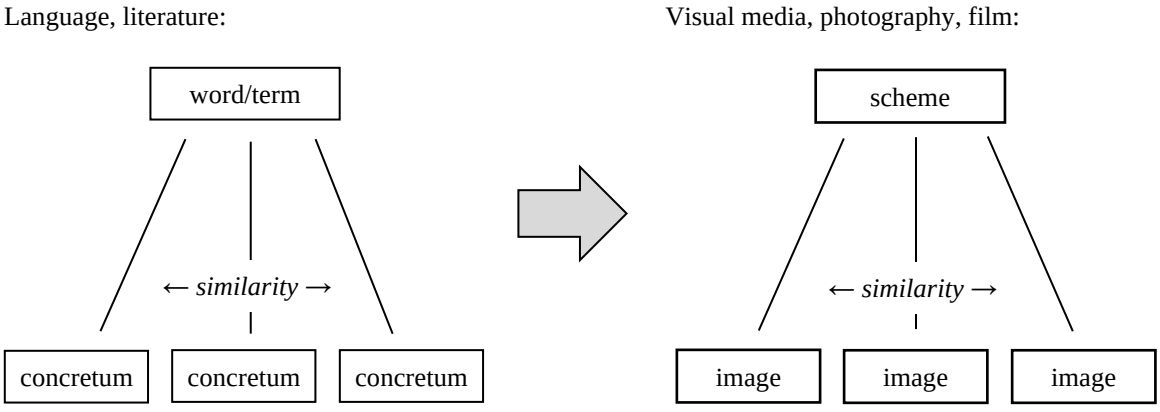
In consequence, this means that identificatory thinking plays a crucial role here as well; and accordingly, so does similarity. We can only identify the table as a table if we already have a concept, a schema, the stereotype of a table. Cognition theory tells us that we carry within us a whole set, a system of schemata and stereotypes that allows us to face the world and engage with the world in the first place. It is the treasure of a prior knowledge in which media competence and real experience are mixed.

Accepting the centrality of these schemata is not easy for film studies, for instance. Mostly, it is simply taken for granted that there are schemata, and often they are regarded pejoratively; and likewise, it is taken for granted that film images are ‘representational‘ images. In the visual arts, this is different. The advent of abstraction in art and, for that matter, also in experimental film has shown how great the confusion is when the images are no longer representational ones, when they purposefully subvert object recognition and our set of established schemata. To include the system of schemata in the analysis of visual media is of central importance, precisely because without it the medium does not function and it becomes impossible to explain how and why we understand images.

Behind the different surfaces of the images, therefore, the schemata and accordingly similarity prevail. The set of schemata and stereotypes forms a second level that lies – barely visible – behind the images. And these schemata, insofar as schemata are always abstract, stand in ~~a~~ systematic tension with the concreteness of the images.

5. Media Comparison

If we now go back to what I have said about language, a startling parallel emerges: namely, the schemata/stereotypes occupy a similar position as the concepts do in the case of language:

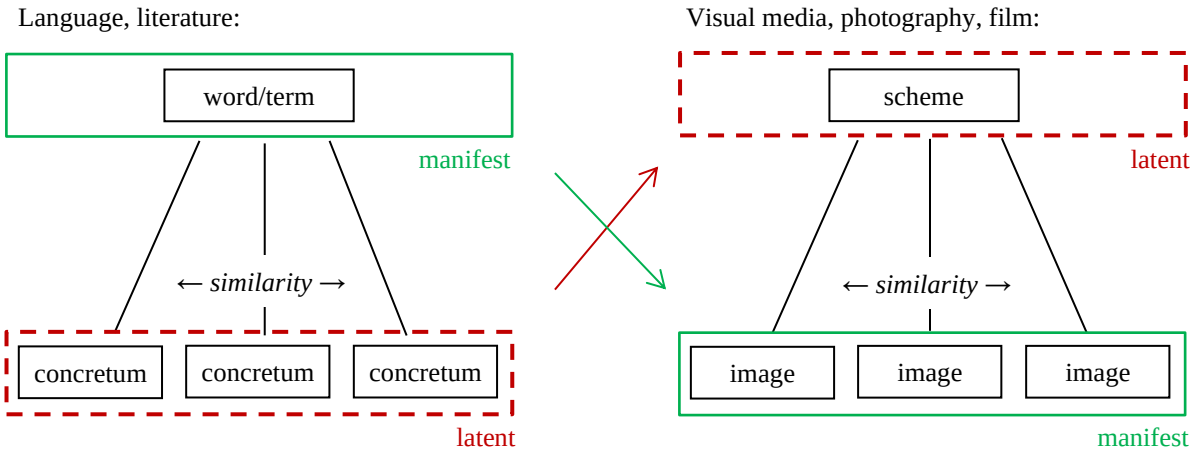


And as in the case of words/terms, it is again similarity that allows the leap from the respective individual to the schema. Immediately, however, decisive differences become apparent as well; especially with regard to the question of what separates the two levels and what significance they have in each case.

In the case of language and literature, it is texts that lie materially on the table and to which attention is directed. And since texts are made of words/concepts, it is the upper level of my sketch that is manifested and observable. Filling the concepts with concrete ideas, on the other hand, is left to the readers; here, their imagination is called upon; if a horse appears in the text, they must imagine what kind of horse it is and what color it is. This concretization remains in the mind and thus latent.

In the case of photography and film, it is exactly the other way around: since the materially manifested images provide the *concreta,* they are to be found on the lower level of the sketch.

And accordingly, it is precisely the schemata that remain latent here, because the recipients have to provide them – as media competence or knowledge of the world.



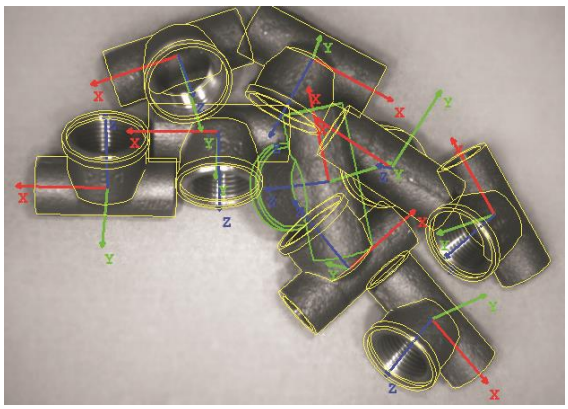
But how on earth can such a difference come about? Especially for similarity: Why can similarity once concern the level of the latent, that is, the imagined concreta, and once the material, manifest images?

My answer would be that exactly such differences are the reason why there are different media at all. Possibly more important than the usual way of distinguishing language from images, conceptual from visual, and iconic from arbitrary signs, would be how media deal with similarity, organize their schema formation, their type of similarity.

6. Computer, Machine Object Recognition

To ensure that my comparison of media does not appear too abstract, as a glass bead game that might just as well be left alone, I would like to address a third medium, because this one sums up nearly everything that has been said so far and shows that we are dealing with a very concrete and practically relevant question. And this medium is the computer, or – to stay in the field of images – one of its applications, machine object recognition.

Object recognition is used in industrial production, for example, when a robot arm has to reach for a machine part. To control it, you use video cameras; and then you need a program that helps find the part, no matter where it is located in the pile.[11]

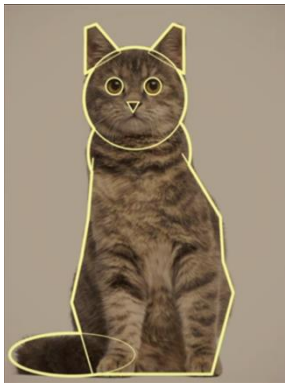


It becomes more difficult when *different* things are involved; for example, when you want to check whether a picture shows a dog or a cat. This task, playfully easy for a three-year-old

[11] Fig.: © MVTec Software; reprod. approved.

child, is a serious problem for a computer. First of all, ‘objects’ have to be isolated in the continuous image, which in itself does not know any objects; and then it is a matter of identifying the objects. For this, the computer must be able to distinguish; it must have categories to which it can assign the objects in each case.

So how does one go about such a task? In a first step, an attempt was made to compile lists of typical features for all objects. If something has a round head and pointed ears, it might be a cat.[12]

This approach, however, quickly failed because the concrete illustrations differed greatly and hardly any image fulfilled all the characteristics that one would consider indispensable.[13]




For this reason, a different approach has been taken and today, so-called ‘deep learning‘ and ‘convolutional neural networks‘ are being used.[14] The special feature of these programs is that they completely dispense with the definition of explicit features; instead, the programs are ‘trained’ by being fed large quantities of preexisting images. This requires a great number of learning cycles and the process is roughly divided into three phases: To begin with, people have to tell the network what each object is, which is called ‘labeling‘; in a second phase, corrective action is only taken if the network delivers incorrect results; so that in the third phase, the program finally runs independently and – without further human intervention – recognizes objects in images; or only makes as many errors as the creators consider acceptable.

The programs rely on an extremely large database. Since 2009, the ambitious project ‘Imagenet’ has built up a database that contains millions of images and organizes them according to 22,000 categories. This database was created with the help of 50,000 crowd-sourced contributors – in

---

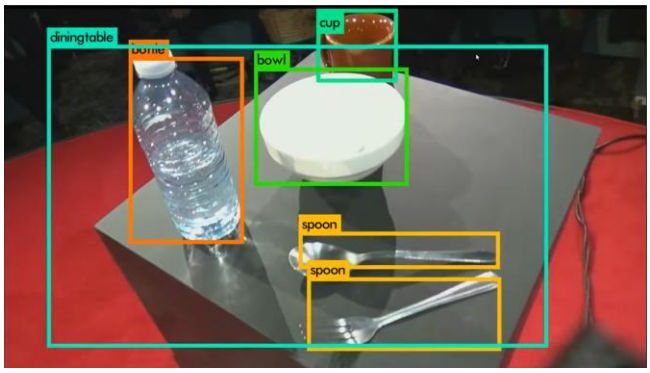
[12] Fig.: Video: Fei Fei Li: How we teach computers to understand pictures; https://www.youtube.com/watch?v=40riCqvRoMs, Min: 05:00ff.

[13] Ibid.

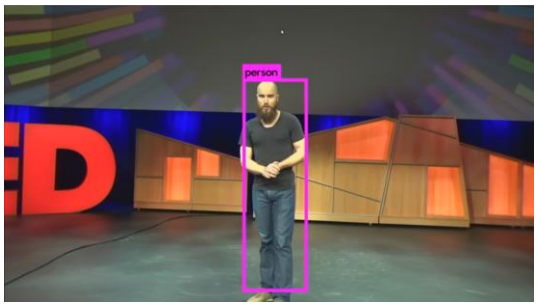
[14] Wikipedia (Germ.): Convolutional Neural Network; https://de.wikipedia.org/wiki/Convolutional_Neural_Network.

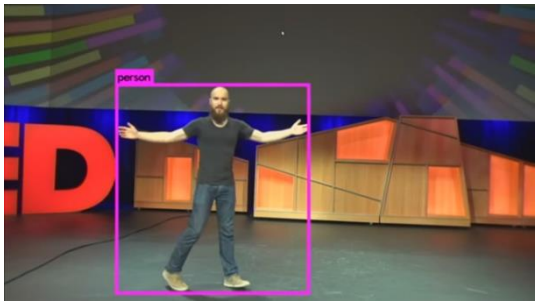
other words, manually, by people. And because it is publicly available, many of the automatic image recognition projects draw on this data to train their programs.

Image databases and neural networks have advanced image recognition to a degree that hardly anyone would have thought possible just a few years ago. Thus, when presented with photos, the programs can identify different objects.[15]

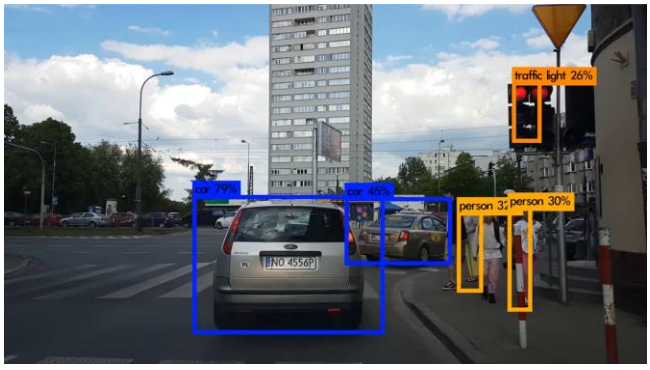


They also detect such objects that are in motion or change their size.[16]





And the third case is even more complicated: the control of autonomous vehicles through the complex traffic of a city. Here, the computer has to identify a large number of objects, even if they are in motion, distorted in perspective, or partially covered; and since driving is a time-critical affair and small children may run into the street rather suddenly, the whole thing has to happen in real time.[17]



[15] Fig.: Video: Joseph Redmon: How computers learn to recognize objects instantly; https://www.youtube.com/watch?v=Cgxsv1riJhI, Min: 03:10ff.

[16] Ibid.

[17] Fig.: Video: Yolo 9000 Object Detection #7; https://www.youtube.com/watch?v=_kxX09i4fds.

All of this is quite impressive, even if many problems remain unsolved and it will still be years before we see fully automated vehicles on the roads.[18] But how does machine object recognition relate to the issue I am pursuing here?

On the one hand, it is obvious that here, too, it is about similarity. Machines are to be enabled to recognize similarity, a skill that until now has been attributed exclusively to humans. And if it is possible to actually formulate this ability in algorithms, one must conclude that similarity is perhaps capable of theory after all.

Second, the programs imitate human perception, primarily insofar as they tie perception to recognition. Here, too, we can find the two levels I have shown above: There are the respective current perceptions and 'above' them, there are the schemata (even if these are called categories here). And as described above for the technical images, the aim is to leave the concretion of the images behind and find schemata in heterogeneous material.

An essential difference, however, is that the categories are fed to the artificial neural networks, whereas in the case of perception – that was Ebbinghaus' point – they emerge from the material in the process of repeated perception.[19] Perception forms its categories in the process of perception itself; this is something artificial neural networks are as yet not able to do.

7. Conclusions

This brings me to my final consideration. For how – we should now ask – do the various building blocks I have discussed fit together? Is there a transition between the mechanisms of perception and the functioning of media? Between the role of similarity in Gestalt-recognition and the conceptualization of language, or the schemata in image discourse? I think that such a transition can indeed be shown. And I think that my computer example has already illustrated this to some extent.

Let's start again with perception. Perception, as I have said, looks specifically for similarity. If all perception is recognition, bound to the ascertainment of similarity, to repetition and memory, and if our perceptual apparatus already possesses schema formation, then it suggests itself here to build a bridge to the schemata of image discourse.

Quite obviously, it is similar mechanisms that Mother Nature first inscribed into our perceptual apparatus and which human history then installs as 'media' in the space of culture. Or, more plausibly: media seem to *imitate* and prolong what is preformed in the mechanisms of our perception.

If this is the case, however, it is shifted in a significant way: Media operate in intersubjective space, in the space of the collective, and this means: *between* people, which is why they are often hastily reduced to 'communication.' And they operate in the external space, manifest/material, insofar as they always make use of technology. If one wants to avoid short-circuits, it will be necessary to spell out the connection a little.

If perception seeks similarity, it is because of economic reasons. It is a matter of reducing the complexity of the world to such an extent that it loses its overwhelming character and becomes at least somewhat manageable; and if the perceptual apparatus forms schemata which allow it

---

[18] The problems can be clearly observed in the cited video itself: It is shocking how many objects the algorithm does *not* recognize.

[19] This difference is very far-reaching. In essence, the material organizes itself almost automatically in the process of repeated perception; the schemata and categories form themselves in repetition. In the world of computers, for example, clustering algorithms are used for this...

to recognize objects, it prevents itself from being swamped, as it would undoubtedly be swamped if it were confronted with material which is – always – new.

Exactly the same function – this is my assertion – *is performed by the media.* My consideration of subsumption has shown that here, too, reduction of complexity is the goal. It is possible to bring a very large number of individual items under a single term, a single scheme; with the consequence that subsequently, the term or the scheme takes their place, so that – incomparably more economically – only term or scheme has to be worked with. Here, too, it is a matter of building up prior knowledge that enables us to deal with what is new in each case.

Similarity is the condition and the rule that allows this reduction. Its ascertainment is always perspectival, reductionist, selective, and precarious, which is why concepts and schemata fundamentally 'lie'; but media that operate without this reduction do not exist. The functioning of media is bound to schema formation.

It seems remarkable to me that this, in my opinion, very central function does not play any role in common media definitions. One reason for this could be that it is taken for granted that media reduce complexity, and the *work* that the media do on this front is therefore overlooked. And perhaps the very fact that perception operates in a parallel manner makes it difficult to recognize the importance of the mechanism of schema formation. But much is lost from view in this way.

In any case, I think one should insist on inspecting the wheels of the machinery. The first consequence would be to shift the focus of what is considered to be the characteristics of media. Possibly, it is not – or not primarily – about 'communication,' not about technology or networks/ networking, not about environments, but about the function of processing the world with the help of technical-symbolic systems, of structuring it, of making it understandable and accessible. Once again, the *semiotic* aspect moves to the center of attention;[20] and the precarious and still unresolved question of what a semiotics should look like that actually makes this side of the media comprehensible.

This question, as I said, is open, and will certainly remain open for the time being. As a conclusion, however, it can be stated that the concept of the schema is an indispensable key when it comes to comparing and relating different media – despite their different characteristics and modes of operation. And it seems equally indisputable to me that similarity is at the center of schema formation.

Schema formation as well as similarity are bound to the criterion of 'fuzziness.' For both, it is true that haziness is not a defect, but the condition of their functioning; and even more: they make a very intelligent use of vagueness. To include different things under one concept *means* to overlook some of their particular differences; and the central point of a visual scheme is that no single perception fulfills the scheme completely.

All higher functions, the capacity of abstraction and generalization, of typification and form building, are based on this vagueness (as on similarity). To render this plausible, or more plausible, is the project that the following chapters pursue.

[20] Cf: Winkler, Hartmut: Zeichenmaschinen. Oder warum die semiotische Dimension für eine Definition der Medien unerlässlich ist. In: Münker, Stefan; Rösler, Alexander (eds.): Was ist ein Medium? Frankfurt am Main: Suhrkamp 2008. pp. 211-221.

[…]