Extended Concept for Meaning Based Inferences
Version 1
*

Gerd Doeben-Henisch
gerd@doeben-henisch

August 30, 2020

Abstract

In the posts 'CASE STUDY SIMULATION GAMES - PHASE 1: Observer-World-Framework' (16.July 2020)[1] and later in the post 'The Simulator as a Learning Artificial Actor [LAA]. Version 1' (23.August 2020)[2] it has been pointed out that the concept of *truth* can sufficiently only be explained if it is embedded in the far more wider concept of *meaning* as part of a working *everyday language* L_0. And this points back to the hearer-speaker of the language; until today only human persons are known to be capable to handle an everyday language sufficiently well. One consequence of this matter is the fact that questions of possible *inferences* from some given facts $f \subseteq F$ to some possible other fact $f' \subseteq F$ can not be disconnected from the facts themselves; the usage of expressions E of the used language – written as $E \subseteq L_0$ – can not be explained without the relation to the learned *meaning relation* μ which has to be assumed as part of the inner cognitive structure of a speaker-hearer. This fact leads to an extended view of the inference from a given state – a set of facts assumed to be true – and some condition Φ which shall be decided to be satisfied by the state S.

1 Basic Application Scenario

In the ongoing project it is assumed that human persons are talking with each other in a *shared everyday language* L_0 with the *goal*, to collect all their individual knowledge which is related to an *agreed question* Q and write this shared knowledge down in two documents: (i) a description of an *agreed actual static state* S representing a finite set of empirical facts F and (ii) a description of a

[1]See: `https://www.uffmm.org/2020/07/16/the-observer-world-framework-part-of-case-studies-phase-1/`

[2]See: `https://www.uffmm.org/2020/08/23/the-simulator-as-a-learning-artificial-actor-laa-version-1/`

set of *change rules* X, which shall be applied to the actual state S to describe which kinds of changes can happen with the assumed state S and thereby determine some *follow-up state* S'.

Thus the human actors A_{hs} are living in some *area* R of a real world RW and these human actors are either part of a static situation S at a certain point of time $t \in T$ or they *know* about such a state S or they can *imagine* such a state S in a way that the human actors are able to *speak about* this state S or to *write about* this state S in the format of a document D_S with expressions $E_{L.0}$ of the shared everyday language L_0.

While the expressions of the document as such have *no meaning*, there are *facts* F in the state S which shall be *encoded* by the expressions of the document $D_S \subset E_{L.0}$. This is only possible if there exists an explicit mapping between the (real/ known/ imagined) facts F of the (real/ known/ imagined) state S and the real expressions of the document D_S.

From human actors it is known that the brain of a human person not only collects continuously sens data from the environment *external* to the body as well as *internal* to the body. These collected data are all represented as *neural correlates* NN of external events called here collectively the *internal states* IS_{NN} of the system.[3]

It is known that these internal states in the brain are organized in a complex way allowing e.g. a distinction between those neural correlates $E_{NN} \subset IS_{NN}$ which represent external expressions E_{RW} and those neural correlates $F_{NN} \subset IS_{NN}$ which represent all kinds of facts F_{RW} from a situation, including even expressions.[4] It is further known that there exists complex mappings inside the brain between these both sets in the format $\mu : E_{NN} \longleftrightarrow F_{NN}$ thereby enabling an *encoding of meaning* by the meaning function μ either from expressions into facts or from facts into expressions. We call the transformation from external events from the real world RW into internal states IS_{NN} general *perception* $perc$ with $perc : RW \longmapsto IS_{NN}$, and vice versa the transformation of internal neural expressions E_{NN} into spoken/ written expressions E_{RW} as either talk $tlk : E_{NN} \longmapsto E_{RW.tlk}$ or write $wrt : E_{NN} \longmapsto E_{RW.wrt}$.

[3] For a first introduction into this perspective see e.g. Baars and Gage (2010) [BM10] and Gage an Baars (2018) [GB18]

[4] These neural representations of induced facts from outside the brain are *not 1-1*! The brain is transforming the incoming data in a manifold way into internal structures which serves primarily the goal to survive including many pre-fixed sub-goals. To understand the world *outside the brain* is a task on its own. For this the brain has to compute on its own structure to get hints how the world outside the drain probably is. There exists no simple and direct way to conclude from the inside of the brain to the outside!

Finally, because the real world is constantly changing – including the human actors – , one has to assume that the system is capable to adapt to these changes. In this text this is called *learning* λ written as: $\lambda : IS_{NN} \longmapsto IS_{NN}$. Formally this can look like a tautological structure, but because the perceptions and other operations change the internal states constantly this self-referential structure is in itself no tautology.

The human actor model HA so far is simple because it makes only a small fraction of all the important aspects of a human actor a subject of discussion. This simple model can be understood as the minimal schema for a *semiotic agent* as discussed in the field of semiotics.[5] What is completely missing at this point are topics like *goal directed behavior*, *preferences*, *desires and emotions*, and much more.

2 Truth Theory

Making all these assumptions from above it is further assumed, that human actors are in an everyday situation capable to assert that the expressions E of a state description D_S are *true statements*. The concept of a *true statement* includes the following details:

Defining True Expressions: From the above described HA-Model it follows that we have at least the following entities: (i) The *real* expressions E_{RW} of the *real* document $D_{RW.S}$, (ii) the *neural correlates* of these expressions E_{NN} of $D_{NN.S}$, (iii) the *meaning function* μ working with neural correlates by being itself a neural process, (iv) the *corresponding meaning* of these expressions as neural correlates $\mu(D_{NN.S}) = F_{NN.D.S}$, (v) the actual *perception* of the real state S as $perc(S) = IS_{NN.S}$ with $F_{NN.S} \subseteq IS_{NN.S}$.

A possible *definition* of *true expressions* E with regard to some aspect F of the real world representing a *state* S could then be, that a human actor trained in the language L_0 can translate these real expressions E into a neural correlates E_{NN} which can with the aid of the learned meaning function μ be mapped into possible neural constructs representing facts $F_{NN.S}$ and *compare* these neural facts with those neural correlates $F_{NN.RW.S}$ gained through perception from the state S_{RW} in the real world, written as $cmp(F_{NN.S}, F_{NN.RW.S}) = x, x \in [0, 1]$. If there exists a *sufficient similar correspondence* $x > \theta$ between the both sets $F_{NN.S}, F_{NN.RW.S}$ then one can say that the given expression from the document $D_{RW.S}$ is *true* with regard to the perceived situation S_{RW}, otherwise

[5] See for a good introduction Noeth (1990)[N9̈0]. A more recent edition can be found as a German edition (2000) [N0̈0]. See also Doeben-Henisch (1998) [DH]

not.[6].[7]

Not Stand Alone, but Models: One sees easily that this definition has a lot of presuppositions which not automatically are given in any case. Besides the general problem that the *meaning functions* of the participating speaker-hearer can always be *different* one has to consider the fact that facts which are *completely given* in a real situation[8] represent only a subset of interesting facts. If a police is reaching a location with an accident then is is not sufficient to clarify, *what can be observed* at this location at the time of arrival, but one has additional to clarify, *how the actual situation came into existence*. To clarify this the police has to examine all important circumstances, has to interrogate all important participants and from this the police has to built some *coherent view* of the situation in the past and how it changed to reach the actual state. Such a 'coherent view' is nothing else than a *model* or even a *theory* constructed on the basis of individual, single facts which have to be *inserted* into *relations* which only can be *constructed* within the space of neural correlates as minimal *cognitive structures* which in this text are understood as *complex facts* F_M forming together a *model* M or a *theory* TH. The complex facts as such F_M are principally language independent, but to communicate such *inner models* of the real world one needs a *symbolic language* like the everyday language L_0 which can map these inner models into expressions E_{NN} of the language and then utter these expressions by speaking or writing as expressions of the real world E. Because every inner model represents more than simple facts the uttered expressions talk about relations given as neural correlates which are *assumed* to have been *valid/ true* in the *past*. In some cases they do further assume to be *valid in some future state*, which has to be judged then when the situation will occur.

State Descriptions From Memory: If the police tries to *reconstruct* some past events the police itself has no own *memories* which they can exploit. But in an everyday world the individual actors have all the time *direct experiences* based on their perceptions. In the model of human actors HA it is assumed

[6] The threshold θ used here for being 'similar enough' is completely speculative. To know more exact values one had to investigate these internal neural processes in more detail. With the actual available methods and tools in Neuroscience this seems to be still impossible.

[7] While this description is dealing with the functions of the brain in general one has to keep in mind, that the individual speaker-hearer does normally not have a clear consciousness about these processes, which are widely (more than 98%) *unconscious*! In this text it is assumed that the *subjective-introspective* counterpart of such a neural mapping process – often called the *phenomenological* perspective – is the *subjective feeling* of an *intuitive evidence* that the learned meaning is matching an actual perception. This then is enables the feeling of being *convinced* that some intended fact is really what it should be.

[8] e.g. the fact 'There is a white cup on the table', 'It rains', 'Peter is eating', ...

that one has to distinguish between (i) *actual perceptions* P_{NN}, (ii) *abstracted actual perceptions*$P_{NN.cat}$ and (iii) *stored abstracted perceptions* $P_{NN.cat.mem}$. In this text it is assumed that all actual perceptions are *individual, concrete neural correlates* which can be seen as *possible instances* of abstracted actual perceptions $P_{NN.cat}$ which can be stored as $P_{NN.cat.mem}$. The difference between the abstracted perceptions $P_{NN.cat}$ and the stored abstracted perceptions $P_{NN.cat.mem}$ is that the stored abstracted perceptions can associate multiple additional abstracted perceptions $P_{NN.cat}$ in a network-like structured thereby enhancing the category with many properties. The abstracted perception $P_{NN.cat}$ functions in this context like an *interface* between different concrete perceptions P_{NN} and a large set of stored abstracted perceptions $P_{NN.cat.mem}$.

From everyday experience we know that such stored abstracted concepts can become very complex. Look to the *perception of a smartphone*. If a young child perceives such an object the first times in its environment it looks like many other objects having some shape, some color, some weight, some tactile properties, making sometimes some noise etc. If it becomes older than perhaps it perceives other properties correlated with certain behaviors like making phone calls, receiving messages, looking to videos etc. All this will continuously be assembled in its space of neural correlates to a larger and larger network of constructs associated with each other in different ways. These concepts all together are representing the concept smartphone. And if the growing teenager will study some hardware or software skills connected with the smartphone the concept will be enlarged even more. There is no fixed limit for a concept.[9] As mentioned before all these *organized* neural correlates are part of the cognitive structure of a human actor forming many complex facts F_M and these complex facts can never be validated by only one actual situation. Thus the internally accumulated complex view of the real world is the product of many complex formation processes called *learning* – mostly unconscious! – which have their *evidence* in the past distributed in many different situations. Thus if a person in an actual situation would say that your smartphone has a software failure then the ability to decide whether this statement is *sound/ valid/ true* in this situation will depend from your understanding of at least the concepts 'smartphone' and 'software failure' and then from the possibility to check those properties which are *known by memory*, that these must be given to infer 'correctly' that the perceivable behavior of the smartphone points to a software failure.

State Descriptions by Imagination: Finally there is the case of an *imagined state* $S_{NN.img}$. This relates to the experience that human persons can change

[9]In Cognitive Psychology this topic is labeled *chunking* and has a long tradition. Newell & Rosenbloom have published some papers and books about it. See Newell (1990)[New90]. For a more modern background to chunking and memory see again the the books Baars & Gage(2010)[BM10] and Gage & Baars (2018)[GB18] .

a concept of a situation by *thinking*. What this means in detail is still not clarified completely by psychology. In this text we assume that an *imagined state* $S_{NN.img}$ is the result of a generative process where human actors take an *initial state* $S_{NN.img.0}$ as starting point and then they apply *conceivable* actions to the given state thus that the initial state does *change* in a way which finally leads to the imagined state $S_{NN.final}$. The initial state $S_{NN.0}$ has to be some *known* state or an *actual perceived* state and the applied changes X shall be changes which are known to be possible or they are candidates for to be shown to be possible. Such an *imagined state* $S_{NN.final}$ can then be classified as being *true* – or not – in the same way like the actual perceived state or the remembered state. But because the meaning of the imagined state can become quite complex it can last some time until the *verification* of the truth of such a description can be finalized. Imagined states $S_{NN.img}$ have also to be stored in the memory but they are not caused by direct perceptions but by *inner activities* of the system itself, in this case in the format of a *generative process*.[10].

3 Generating Imagined States of the Future

How the brain is working while generating new imagined states which suffice all the above mentioned requirements is still not really known. What one can do and what will be done in this project is to define an application scenario by *defining the allowed observable behavior* and then asking back which kind of processes are necessary to enable the generation of imagined future states.

As described above in this text it is assumed that a *follow up state* S' of a *given state* S will be generated by either (i) *deleting* some expressions E^- of the given state S if converted to S' or (ii) by extending the state S by some new expressions E^+ for the construction of S' by*creating* new expressions, written as $S' = S - E^- \cup E^+$.

For this to happen it is assumed in this text that a *change rule* $\xi \in X$ has the following format: *IF*-part and *THEN*-part, written as $X \subseteq X_{if} \times X_{then}$. While an element of the then-part has always the format $X_{then} \subseteq E^- \times E^+$ an element of the if-part is a set of expressions which have to be *valid*. One can generalize this case in the following way: If we assume that the expressions of the state S are all *true* with regard to the assumed part of the world then it has to be clarified for each expression ϕ in the if-part of a change rule whether this expression is *true* in the assumed state S or not. The following cases will

[10]On the *phenomenological* level this can be understood as *thinking ahead*. Thinking ahead can help to identify *possible states in a possible future* which perhaps can be more advantageous than others which can happen if we as human actors do not prevent the occurrence of the less advantageous states

be examined:

1. **Expression matches expression directly:** The IF-Part of a change rule $\xi \in X$ includes an expression Φ which matches directly at least one expression in the assumed state S. This can be a 1-1-match because there is an expression in S which is formally identical with Φ or the expression Φ has some *variables [V]* which can be substituted by some *numerical constants* as *[V/c]*. Eventually there exist more cases which allow such a direct replacement.

2. **Facts need a Model:** The IF-Part of a change rule $\xi \in X$ includes expressions Φ_i which ask for the existence of some *actor*, a human actor or some other actor. Actors are characterized as input-output systems with a *behavior function* β enabling an actor to compute a *response [O]* triggered by an *input [I]* by using its *internal states* IS. Every actor is assumed to have some internal *models* M which have been *learned* and which enable the actor to identify parts of a given input as parts of some more complex cognitive structure which in turn enables specific outputs.

While in the case of the direct matching of expressions only simple computations are needed on the site of the simulator σ, the existence of actors calls for a more complex answer which implies more complex computations.

In case of *real human actors [HA]* one can include human actors in the simulation thereby realizing an *interactive simulation [IS]*. In case of *artificial actors [AA]* one has to provide an *artificial behavior function* β as a realized *algorithm*. To a certain degree it can be helpful to define some artificial actors *mimicking* human behavior in a simplified way.[11] This allows simulations which can be clearly more realistic than without such modeling. In this case the simulator σ has to do the job of computing the different behavior functions.

References

[BM10] Bernard J. Baars and Gage Nicole M. *COGNITION, BRAIN, AND CONSCIOUSNESS. Introduction to Cognitive Neuroscience*. Elsevier - Academic Press, Amsterdam - Boston - Heilberg et al., 2 edition, 2010.

[CMN83] Stuart K. Card, Thomas P. Moran, and Allen Newell. *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, Inc., Mahwah (NJ), 1 edition, 1983.

[11] This method of using artificial actors to mimicking simplified human behavior has been introduced and succesfully being used by Card, Moran & Newell (1983)[CMN83]. In their book they introduced the *Model Human Processor* and the concept of *GOMS*-models.

[DH] Gerd Doeben-Henisch. Semiotic machines. In E.W.B. Hess-Lüttich et al., editors, *Signs & Space – Raum & Zeichen. An International Conference on the Semiotics of Space and Culture*. Gunter Narr Verlag. pp.313–327.

[GB18] Nicole M. Gage and Bernard J. Baars. *Fundamentals of Cognitive Neuroscience: A Beginner's Guid*. Academic Press - Elsevier, London - Oxford - San Diego - Cambridge (MA), 2 edition, 2018.

[N9̈0] Winfried Nöth. *Handbook of Semiotics*. Indiana University Press, Bolomington - Indianapolis, 1 edition, 1990. Enlarged and completely rewritten edition of the 'Handbuch der Semiotik' (1985).

[N0̈0] Winfried Nöth. *Handbuch der Semiotik*. J.B.Metzler, Stuttgart - Weimar, 2nd edition, 2000. Completely rewritten 2nd edition of the 'Handbook of Semiotics' (1990).

[New90] Allen Newell. *Unified Theories of Cognition*. Havard University Press, Cambridge (MA, USA) and London (England), 3 edition, 1990.