Homunculus’ Brain and Categorical Logic

Michael Heller
Copernicus Center for Interdisciplinary Studies
ul. Szczepańska 1/5, 31-011 Cracow, Poland

March 11, 2019

Abstract

The interaction between syntax (formal language) and its semantics (meanings of language) is well studied in categorical logic. Results of this study are employed to understand how the brain could create meanings. To emphasize the toy character of the proposed model, we prefer to speak on homunculus’ brain rather than just on the brain. Homunculus brain consists of neurons, each of which is modeled by a category, and axons between neurons, which are modeled by functors between the corresponding neuron-categories. Each neuron (category) has its own program enabling its working, i.e. a theory of this neuron. In analogy with what is known from categorical logic, we postulate the existence of the pair of adjoint functors, called Lang and Syn, from a category, now called BRAIN, of categories, to a category, now called MIND, of theories. Our homunculus is a kind of “mathematical robot”, the neuronal architecture of which is not important. Its only aim is to provide us with the opportunity to study how such a simple brain-like structure could “create meanings” out of its purely syntactic program. The pair of adjoint functors Lang and Syn models mutual dependencies between the syntactical structure of a given theory of MIND and the internal logic of its semantics given by a category of BRAIN. In this way, a formal language (syntax) and its meanings (semantics) are interwoven with each other in a manner corresponding to the adjointness of the functors Lang and Syn. Categories BRAIN and MIND interact with each other with their entire structures and, at the same time, these very structures are shaped by this interaction.
1 Introduction: On the Computer Screen

Together with my colleagues I was preparing a paper for publication. A phase portrait was nicely displayed on the computer screen. Network of trajectories represented a class of solutions to the equation we were interested in. At some points, called critical points, certain trajectories crossed each other. These points were important for our analysis. Some of the diagrams we worked with appeared later as figures in our publication [31]. The figures had to be explained, so we decided to attach appropriate labels to some of the critical points. We attached the label “stable saddle” to one of them. No problem. Then we proceeded to attach the label “unstable saddle” to another one. But the label jumped up. We tried to fix it up, but it jumped down. Then we started laughing. After all, it is an unstable point!

Let us try to understand the situation. We investigate an equation that (virtually) contains in itself its space of solutions (irrespectively of whether we explicitly know them or not). Through the suitable computer program and some “electronic circuits”, which are activated by the program, this space of solutions is mapped into the phase portrait displayed on the computer screen. The diagram we see on the screen is certainly something more than just a picture. It does not simply show stable and unstable critical points; it also does what the abstract equation orders its solutions to do (labels jump up and down at instabilities).

Let us go a step forward. In fact, the phase portrait on the screen is a substitute of the world. For suppose that our equation “describes” (or better – models) a mechanical system (e.g., a pendulum or oscillator). Then the unstable critical points of our equation correspond to physical situations in which the considered mechanical system behaves in an unstable way. We thus have, on the one hand, an equation (or a set of equations) or, more broadly, a mathematical theory and, on the other hand, a domain (or an aspect) of the physical world of which the considered mathematical theory is a model. Between the mathematical model and the domain (or aspect) of the physical world there is a mysterious correspondence – correspondence in the root-meaning of this word: both sides co-respond to each other. It is an active correspondence, and the activity goes both ways: it looks as if the domain of the world informed the theory about its own internal structure, and the theory answered by prescribing what the domain should do. And

\footnote{The equations we considered in our publication referred to a cosmological situation.}
the domain does it. The equations prescribe what the world should do, and the world executes this. The equations and the world are coupled with each other and act in unison.

And the screen on my computer? It is a part of the world. The program we have constructed reads the structure of the equations and executes what the equations tell it. And because of the coupling between the equations and the world, the computer does, in miniature, what does the world on its own scale. This is the reason why computers are so effective in our reading the structure of the world.

There is another domain in which a formal structure reveals its effective power and produces real effects. Such processes occur in the brain. The formal structure in question consists of electric signals propagating along nerve fibres between neurons across synapses, and the world of meanings should be regarded as a product of this activity. The interaction seems to go both ways: the “language of neurons” (what happens in the brain) produces the meanings related to this language (in the mind), and the meanings somehow influence the architecture of neurons.

It seems that in both these cases (mathematical laws and their effects in the real world, and the brain–mind interactions) we meet two instances of the same working of logic where syntax (a formal structure), by effectively interacting with its semantics, produces real effects. This kind of interaction, although kept strictly on the level of logic (i.e. with no reference to processes in the real world), is well known in the categorical logic. In the present paper, we attempt to employ these achievements of categorical logic to try to understand the brain–mind interaction.

The traditional terminology of brain and mind (irrespectively of current trends in cognitive sciences to get rid of their conceptual load) seems especially well adapted to the present context in which general ideas are more important than structural details. Moreover, to avoid too hasty associations with the human brain and to emphasize the toy character of the proposed model, We prefer to speak on homunculus’ brain rather than just on brain.

The action of our argument develops along the following lines. Section 2 is a reminder on formal language, its syntax and semantics. Sections 3 and 4 briefly review those parts of categorical logic that refer to these concepts. Every category, call it $C$, has its internal logic, and if this logic is sufficiently rich, the category provides semantics for a certain formal theory $T$. Moreover, there exists a pair of adjoint functors, called Lang and Syn, from a category, called CATEGORIES, of categories belonging to a certain class (for instance,
coherent categories) to a category, called THEORIES, of theories and *vice versa*, which describe mutual dependencies between the syntactical structure of $T$ and the internal logic of its semantics given by $C$. This is described in section 3. In this way, syntax and semantics are interwoven with each other in a manner corresponding to the adjointness of the functors Lang and Syn. This is explored in section 4.

In section 5, the category CATEGORIES becomes the category BRAIN. It constitutes a simple model of a homunculus’ brain. Objects of this category are categories (belonging to a certain class); every such category models a neuron. Morphisms of this category model signals propagating along nerve fibres between neurons. The category THEORIES becomes the category MIND. Its objects are “theories of neurons”; more precisely, if $C \in$ BRAIN, then its “theory” is Lang($C$) in MIND. Morphisms of this category are functors between the corresponding syntactic theories; more precisely, if $T_1, T_2 \in$ MIND, then the morphism between them is Syn($T_1$) → Syn($T_2$). The pair of adjoint functors Lang and Syn model the interaction between the syntax of “theories” and their semantics, i.e. the network of neurons. The categories BRAIN and MIND are indeed somehow related to what their names refer to, at least as far as homunculus’ brain and mind are concerned.

After a seminal paper of W.S McCulloch and W. Pitts [18], published as early as in 1943, that proposed using classical logic to model neural processes in the brain, there were so many papers, developing and modifying (with various logical systems) this idea, that to quote even a sample of them would be immaterial (for a relatively recent state of art see a short review [15]). A. Ehresmann [6] claims that it was R. Rosen [23] who was the first to employ category theory to model biological systems. A series of works followed (a non-representative sample: [9, 12, 19, 20]) proposing to use various parts of category theory to model different aspects of the brain activity. In particular, adjoint functors were suggested to model “a range of universal-selectionist mechanisms” [7]. However, I have nowhere found something similar to modeling the interaction between brain’s language and its meaning.

## 2 Syntax and Semantics

In linguistics, syntax and semantics are regarded as parts of semiotics, the study of signs. Syntax studies relations between signs, and semantics rela-
tions between signs and what the signs refer to. Syntactic properties are attributed to linguistic expressions entirely with respect to their shape (or form). Semantics, on the other hand, endows them with meaning by referring signs to what they signify. Logic adapts these ideas to its own needs. Since it is a formal science, signs it considers should be elements of a formal language, and they cannot refer to anything external. As put by Halvorson, “But a formal language is really not a language at all, since nobody reads or writes in a formal language. Indeed, one of the primary features of these so-called formal languages is that the symbols don’t have any meaning” [10]. This is why the meaning should “artificially” be constructed for them. The idea of how this should be done can best be seen in Tarski’s prototype of this procedure [28]. If a sentence $s$, the truth of which we want to define, belongs to a language $L$ then the definition of $s$ should be formulated in a metalanguage $M$ with respect to the language $L$. And the metalanguage $M$ should contain a copy of $s$ so that anything one can say with the help of $s$ in $L$, can also be said in $M$. The definition of “True” should be of the form

$$\text{For all } x, \text{True}(x) \text{ if and only if } \varphi(x)$$

with the condition that “True” does not occur in $\varphi$. Here $x$ stands for the copy of the sentence $s$ in the metalanguage $L$, and $\varphi(x)$ describes, also in $M$, the state of affairs of which the sentence $s$ in $L$ reports (for more details see [13, 24]). Metalinguistic copy of $s$ could also be expressed as “$s$” (taken in quotes). Tarski’s own example:

“It snows” is true iff it snows.

For pedagogical reasons, this example is taken from the colloquial language, but strictly speaking Tarski’s definition refers to formal languages. The formal language $L$ has its own syntax (since it is a formal language), but is lacking its semantic reference. As we have seen, such a reference had to be constructed for it with the help of the metalanguage $M$.

Now, the idea is to improve the situation, by looking for such a conceptual context in which a semantics for a given theory would arise in a more natural (or even spontaneous) way.

---

2Sometimes one distinguishes also pragmatics which studies relations between signs and their users.
3 Categorical Semantics

To do so we must first define precisely what we mean by language. Since the definition must be precise, let us choose as an example the language of mathematics based on the standard first order logic (which is enough for most of the usual mathematics). Many other languages may be formalized in a similar way. In such a language we distinguish:

- constants: 0, 1, 2, ..., a, b, c, ..., and variables: x, y, z, ..., which can be combined by primitive operations to give

- terms, for example: x + y, x^3, ... which, in turn, can be combined, with the help of primitive relations, such as =, <, ≤, ..., to produce

- formulae, for example: x + y = z, x ≤ y, ... which, in turn can be combined, with the help of the usual logical connectives and quantifiers, into

- more complicated formulae.

To make the language more flexible and more adapted for concrete applications, we diversify its expressions into various types (called also sorts). In mathematics, we might use different letters for natural and real numbers, or different symbols for vectors and scalars. We say that, in both cases, we are using a two-typed language. There may be languages with as many types as is needed.

What we need is not so much a language, but rather a theory. In mathematical logic theory is almost the same as language; it is a formal language aimed at axiomatizing a certain class of sentences. The concept of theory, as it is functioning in modern physics can, in principle, be regarded as the special case of the logical concept of theory, although in scientific practice theories are rarely formulated with the full logical rigor.

Let then T be a theory expressed in a multi-type language. Such a theory is defined to consist of the following data:

1. A set of types \{X_1, X_2, ... X, Y, \ldots \}.

2. A set of variables \{x, y, z, \ldots , x_1, x_2, x_3, \ldots \} with a type assigned to each variable.
3. A set of function symbols with a type assigned to each domain and codomain of every function symbol; for instance, to the term \( x + y \), with the variable \( x_1 \) of type \( X_1 \) and the variable \( x_2 \) of type \( X_2 \), there corresponds the function symbol \( f : X_1 \times X_2 \rightarrow Y \), and the term \( f(x_1, y_1) = x_1 + x_2 \) is of type \( Y \).

4. A set of relation symbols with a type assigned to each argument of every relation symbol; for instance, to the formula \( x + y = z \), with the variable \( x \) of type \( X_1 \), the variable \( y \) of type \( X_2 \) and the variable \( z \) of type \( X_3 \), there corresponds the relation symbol \( R \subseteq: X_1 \times X_2 \times X_3 \), and \( R(x, y, z) \) is an atomic formula.

5. A set of logical symbols.

6. A set of axioms for a given theory build up from terms and relation symbols with the help of logical connectives and quantifiers, respecting types of all terms.

This is, in fact, a purely syntactic definition of theory (for details see [3, pp. 344-348], [17, pp. 527-530]). Now, we want to create a semantics, i.e. a model, for a theory \( T \). This is done by constructing a category \( C_T \) which will serve us as such a model. The construction is almost obvious:

1. each type of \( T \) is an object of \( C_T \),

2. for each function symbol \( f \) with the types \( A \) and \( B \) of its domain and codomain in \( T \), correspondingly, \( f \) is a morphism from the object \( A \) to the object \( B \) in \( C_T \)

3. each variable is an identity morphism in \( C_T \),

4. for each relation symbol \( R \) in \( T \), its counterpart in \( C_T \) is a subobject in \( C_T \). Suppose \( \phi \) is a subobject of an object \( A \) in \( C_T \) then, by analogy with the usual theory of sets, \( \phi \) can be thought of as a collection of all things of type \( A \) that verify \( \phi \).

This definition must be supplemented with all of (first order) logic which is used to express axioms in \( T \) (for details [14]). Roughly speaking, since

\[ \text{Since } f \text{ is now in } C_T \text{ rather than in } T, \text{ it should formally be denoted by a different symbol.} \]
formulae correspond to subobjects, and all subobjects of a given object are partially ordered by inclusions (they form a poset), the axioms can be expressed in terms of the order relation on the subobject poset in the category \( C_T \). The category, defined in this way, is appropriately called the categorical semantics for a theory \( T \).

We have thus created (almost automatically!) a domain (the category \( C_T \)) the theory \( T \) refers to. Internal architecture of the category \( C_T \) exactly matches the logic involved in the theory \( T \).

Let us also mention that, vice versa, having a (sufficiently rich) category \( C' \), we can construct the formal theory \( T' \) the logic of which matches the internal architecture of the category \( C' \). This can be done by reading the above definition of the categorical semantics “backward”, i.e. we regard objects of \( C' \) as types of \( T' \), identity morphisms of \( C' \) as variables in \( T' \), etc. The theory \( T' \), reconstructed in this way from the category \( C' \), is called internal logic of \( C \). This entire process can be regarded as a functor, called \( \text{Lang} \), from a category of categories, call it \( \text{CATEGORIES} \), to a category of theories, call it \( \text{THEORIES} \).

\[
\text{Lang}: \text{CATEGORIES} \to \text{THEORIES}.
\]

For the time being this definition remains informal since neither \( \text{CATEGORIES} \) nor \( \text{THEORIES} \) have properly been defined, but it will be done below.

Let us start with a formal theory \( T \). We now want to organize it into a category \( \text{Syn}(T) \), called the syntactic category of \( T \). It is done in the following way.

Let \( \Gamma \) be a collection of type assertions, i.e. a collection of rules assigning a type to each term of a given theory, and \( \Phi \) a collection of all well defined formulae of \( T \). The pair \( (\Gamma, \Phi) \) is called a context. It is a formalization of what in ordinary language one means by this term.

If \( T \) is a type theory, its syntactic category, \( \text{Syn}(T) \), is defined as follows. Its objects are contexts \( (\Gamma, \Phi) \) and its morphisms \( (\Gamma, \Phi) \to (\Delta, \Psi) \) are interpretations (or substitutions) of variables. The latter means that for each type, prescribed by \( \Delta \), we must construct an expression of this type out of data contained in \( \Gamma \). In general, this is done by substituting terms from \( \Gamma \) for variables in \( \Delta \). We also must, for each assumption required by \( \Delta \) (if there are any), present a proof of this assumption out of assumptions contained in \( \Gamma \) (for details see \[8, 27\]).
The category $\text{Syn}(T)$, constructed in this way, is also called a category of contexts (for details see [8, 26]).

Since from a theory $T$ we have constructed the category $\text{Syn}(T)$, we can have a functor,

$$\text{Syn}: \text{THEORIES} \to \text{CATEGORIES}$$

provided we define the categories THEORIES and CATEGORIES. We do this in the next section.

## 4 Syntax – Semantics Interaction

Let us start with objects for both of these categories. It is obvious that they will be categories and theories, respectively. To have workable categories, one must restrict the class of theories as candidates of being objects in THEORIES (and analogously for CATEGORIES). The criterion one follows is the kind of logic that underlines a given theory. It could be what logicians call: finite product logic, regular logic, coherent logic, geometric logic, etc. (as it could be expected, the internal logic of the corresponding semantic category will be of the corresponding kind, i.e. finite product logic, regular logic, etc.) [14]. For our further analysis it is irrelevant which one will be chosen. However, for the concreteness sake we may think about the coherent logic. Roughly speaking, this is a fragment of the first order logic using only the connectives $\land$ and $\lor$, and the existential quantifier. Large parts of mathematics can be formalised with the help of this logic. To this logic there correspond coherent theories and coherent categories. They will constitute objects of THEORIES and CATEGORIES, respectively. Morphisms for CATEGORIES are obviously functors between corresponding categories; for instance coherent functors for coherent categories [1]. Let now $T_1$ and $T_2$ be objects in THEORIES. Morphism between $T_1$ and $T_2$, $T_1 \to T_2$, is a functor between their corresponding syntactic theories $\text{Syn}(T_1) \to \text{Syn}(T_2)$. Roughly speaking, this means that it is possible to express (to interpret) $T_1$ in terms of $T_2$ (for details and discussion see [11]).

As a side remark let us notice that by studying the category THEORIES, we could learn “how individual theories sit within it, and how theories are

---

4Strictly speaking, CATEGORIES is a 2-category (since its objects are categories and morphisms are functors), and THEORIES is a 2-category, in this case, called also a doctrine [3].
related to each other” [11, p. 413]. This is nicely consonant with a newer trend in the philosophy of science to investigate the so-called inter-theory relations [2, 22].

A truly remarkable fact is that the functors Lang and Syn constitute the pair of adjoint functors. Let us explain what does it mean.

Let us consider any pair of objects: $C$ of CATEGORIES and $T$ of THEORIES. Adjoint functors serve to compare them. However, they cannot be compared directly since they live in different categories. Adjoint functors serve to move each of them to the correct category so as to enable the comparison. Let us follow this process step by step [25, pp. 148-153].

Let us first consider the object $T$ which lives in THEORIES. We want to compare it with the object $C$ which lives in CATEGORIES. We thus move $C$ to THEORIES with the help of the functor Lang to obtain the object $\text{Lang}(C)$. We now make the comparison with the help of a suitable morphism,

$$f : \text{Lang}(C) \rightarrow T$$

in THEORIES. We do the same starting with $C$ in CATEGORIES and $T$ in THEORIES, and compare $C$ with $\text{Syn}(T)$,

$$g : C \rightarrow \text{Syn}(T)$$

in CATEGORIES. To complete the definition of adjunction we demand that morphisms $f$ and $g$ should constitute a pair of bijections which is natural both in $C$ and $T$ (see below).

The above definition can be put into the concise form

$$\text{THEORIES} (\text{Lang}(C), T) \cong \text{CATEGORIES} (C, \text{Syn}(T)), \quad (1)$$

expressing an isomorphism between the right and left hand sides of this formula that is natural in $C$ and $T$. The latter condition says that when $C$ varies in CATEGORIES and $T$ varies in THEORIES, the isomorphism between morphisms $\text{Lang}(C) \rightarrow T$ in THEORIES and $C \rightarrow \text{Syn}(T)$ in CATEGORIES vary in a way that is compatible with the composition of morphisms in CATEGORIES and THEORIES, correspondingly, and with the actions of Lang and Syn on both these categories (see [16, pp. 50-51]).

We should notice that in the above definition, we compare, in fact, not objects of two different categories, but rather categories themselves (objects

---

5For a full definition of adjoint functors see any textbook on category theory.
$C$ and $T$ are any pair of objects). Moreover, comparing two categories we are not so much interested in their objects, but rather in morphisms between objects. This is clear from the fact that at the end, we have identified those morphisms of two categories that are pairwise naturally isomorphic among themselves.

As we can see, categorical logic does not simply creates a semantics for a given language, but shows that dependencies between them go both ways: in a sense, syntax and semantics create each other. More precisely, they condition each other through the adjointness relation.

5 Categories BRAIN and MIND

So far everything that has been said was just a reminder of standard and well known things. From now on, everything will be hypothetical and highly simplified. The bold and maximally simplified hypothesis is that neurons in the brain can be modeled as categories, the internal logic of which is sufficiently complex (yet manageable). Of course, our inspiring motive is the human brain and in constructing our model we shall try to imitate what is going on it; however, being conscious of our simplified and highly idealized assumptions, we prefer to speak about a homunculus brain. Our homunculus is a kind of “mathematical robot” the aim of which is to provide us with the opportunity to study how such a simple brain-like structure could “create meanings” out of its purely syntactic program. Our other drastically simplifying assumption consists in systematically ignoring all brain’s functions and processes that are not directly related to the proposed syntax–semantics relationship.

As it is well known, neurons communicate through signals transmitted via: presynaptic (source) neuron – axon – synapse – dendrite – postsynaptic (target) neuron, and this via is unidirectional. In our homunculus model, these transmission processes will be regarded as functors between categories (neurons).

Let us consider the category CATEGORIES, which we now aptly call BRAIN. Its objects are categories modeling neurons, and morphisms are functors between these categories.

We thus assume that each neuron in the homunculus brain is represented by a category (belonging to a certain class of categories; in the following we shall simply say that a neuron is a category). At the moment, we are not
interested which biological mechanisms implement this assumption. Everything that counts in this model is the assumption that neurons consist of collections of objects and morphisms satisfying conditions from the category definition. We should have in mind that these simple conditions could lead to highly complicated structures.

Morphisms (arrows) in the category CATEGORIES are functors between object-categories that is to say axons through which neurons communicate with each other. The crucial thing is that they must satisfy the usual conditions for morphisms: composition of morphisms, its associativity, the existence of identity morphisms. With the latter there is no problem: no output from a neuron counts as its identity morphism. To verify whether two other conditions are verified in the human brain would require going deeper into the neural structure of our brain. In the case of the homunculus brain this is not necessary. Since the homunculus is of our construction, we simply assume that synapses in its brain well compose and do that in the associative way.

The next step seems obvious. Each neuron (modeled as a category $C \in$ BRAIN) has its own program enabling its working, i.e. an internal logic underlying this program. We thus can define a counterpart of Lang$(C)$ which h is a “theory” of this neuron. It is reasonably to claim that it is an object of the category THEORIES which we now call MIND, and the functor Lang: BRAIN $\to$ MIND is defined in analogy to that between CATEGORIES and THEORIES.

What about the morphisms between such objects? We proceed in strict analogy with what has been done in THEORIES. Let now $T_1$ and $T_2$ be objects in MIND, a morphism between them, $T_1, \to T_2$, is a functor between their corresponding syntactic theories, i.e. Syn($T_1$) $\to$ Syn($T_2$), where the functor Syn: MIND $\to$ BRAIN is defined in analogy to that between THEORIES and CATEGORIES.

The analogy is only apparently straightforward. In fact, it is based on a huge extrapolation, and as such highly hypothetical, but it is worthwhile to be explored since the problem at stake deserves even a high risk. By pursuing this analogy we could claim that also in this case the functors Lang and Syn are adjoint functors. If so, we have a very interesting conjunction between brain and mind; it is interesting even if brain and mind are modeled by such a naive construction.

Neurons, their interactions and programs underlying their working are, in contrast with abstract categories like CATEGORIES and THEORIES, real
things, at least in the homunculus world, and we are entitled to suppose that
the functors Lang and Syn between Brain and Mind really do what they
formally signify (like our phase portrait on the computer screen really did
what the program told it to do).

Roughly speaking the functor Lang provides a collection of theories
(mind) for a collection of neurons (brain), and the functor Syn transfer a
syntax of these theories to the network of neurons. The action of these two
functors is adjoint; consequently it determines a strict interaction between
BRAIN and MIND. Let $C$ be any object (a neuron) in BRAIN and $T$ any
object (the theory of this neuron) in MIND, then equation (1) assumes the
form

$$MIND(Lang(C), T) \cong BRAIN(C, Syn(T)).$$

The natural isomorphism $\cong$ appearing in this equation is crucial. It states
that when we go from neuron to neuron as objects in BRAIN, and their

$$\text{corresponding theories vary in THEORIES, then the isomorphism between}\n$$

morpisms $\text{Lang}(C) \to T$ in MIND and $C \to \text{Syn}(T)$ in BRAIN varies in a

$$\text{way that is compatible with the composition of morphisms in BRAIN and}\n$$

MIND, correspondingly, and with the actions of the functors Lang and Syn

$\text{(see \[16\].)}$ We could summarise the situation by saying that the categories

$\text{BRAIN and MIND interact with each other with their entire structures and,}\n$

$\text{at the same time, these very structures are shaped by this interaction.}\n$

6  A Comment

Interactions syntax – semantics are omnipresent both in our everyday con-

versations and in various forms of practicing science. The world around us

is full of meanings and of our attempts to decipher them. Science could be

regarded as a machine to produce signs, through experimentation and criti-

cal reasoning, and extract from combinations of them an information about

the structure of the world. Logicians put a lot of effort to make the syntax

– semantics interaction precise. As we have seen in section 2, in spite of the

fact that formal languages are lacking any external references, it was possible

to create for them semantical references by cleverly exploiting the relation

tween language and its metalanguage. In categorical logic the situation has

\[\text{For a full discussion of the role of the naturality condition in the definition of adjoint}\n\text{functors see any textbook on category theory.}\]
Any formal theory $T$, via the functor Syn generates the category $\text{Syn}(C) = C_T$ of which it is a theory, i.e. $C_T$ provides a “natural” semantics for $T$. And vice versa, any (sufficiently rich) category $C'$, via the functor Lang, generates its own theory $\text{Lang}(C') = T'_{C'}$ which constitutes the internal logic of $C'$. It is interesting to notice that $T_{C_T}$ does not coincide with $T$, they are only Morita equivalent. Here, we shall not go into technical details; it is enough to say that two Morita equivalent theories could be regarded as two interpretations of the same theory [10].

The fact that $T_{C_T}$ does not coincide with $T$ is the consequence of the fact that the functors Lan and Syn are not mutually inverse functors but constitute the pair of adjoint functors. This in turn implies that in categorical logic the interaction syntax – semantics is skillfully complex with creative influences coming both ways.

All the above discussed (in sections 3 and 4), properties of the syntax – semantics interaction are obviously preserved if applied to the categories BRAIN and MIND. There is only one big difference: now “neurons and their theories” are real things (although in a highly idealised version of the homunculus world). Nevertheless, the situation is not much different from that one meets in many empirical sciences in which some abstract mathematical structures model some real processes (always more or less idealised). We should not be surprised that the method of mathematical modeling works when applied to our cognitive processes, but rather that mathematical structures not only describe the real world (whether it is our brain or the world of physics), but that they are also effectively acting in it (like in the little arrow on the computer screen).

References

[1] Barrett, T., Halvorson, H., Morita Equivalence, arXiv:1506.04675 [math.LO]

[2] Batterman, R. Intertheory Relations in Physics, The Stanford Encyclopedia of Philosophy, https://plato.stanford.edu/archives/fall2016/entries/physics-interrelate

[3] Borceux, F., Handbook of Categorical Algebra, vol. 3: Categories of Sheaves, Cambridge University Press, Cambridge, 1994.
[4] Coherent Functors, nLab, 2011, https://ncatlab.org/nlab/show/coherent+functor

[5] Doctrine, in nLab, https://ncatlab.org/nlab/show/doctrine

[6] Ehresmann, A., Applications of Categories to Biology and Cognition, in: Categories for the Working Philosopher, ed. by E. Landry, Oxford University Press, Oxford, 2017, pp.358-380.

[7] Ellerman, D., On Adjoint and Brain Functors, arXiv:1508.04036v1 [math.CT]

[8] Fu, Yuchen, Category Theory, Topos and Logic: A Quick Glance, http://charlesfu.me/repository/topos.pdf

[9] Gomez, J., Sanz, R., Modeling Cognitive Systems with Category Theory, Universidad Politecnica de Madrid, 2009.

[10] Halvorson, H., Scientific Theories, in: Oxford Handbook of the Philosophy of Science, ed. by P. Humphreys, Oxford University Press, Oxford, 2016, pp. 585-608.

[11] Halvorson, H., Tsementzis, D., Categories of Scientific Theories, in: Categories for the Working Philosopher, ed. by E. Landry, Oxford University Press, Oxford, 2017, pp. 402-429.

[12] Healy, M.J., Caudell, T.P., Ontologies and Worlds in Category Theory: Implications for Neural Systems, Axiomathes 16, 2006, 165-214.

[13] Hodges, W., Tarski’s Truth Definitions, The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), https://plato.stanford.edu/archives/fall2018/entries/tarski-truth/.

[14] Internal Logic, nLab, 2017, https://ncatlab.org/nlab/show/internal+logic

[15] Koch, Ch., Computation and the Single Neuron, Nature 385, 1997, 207-210.

[16] Leinster, T., Basic Category Theory, Cambridge University Press, Cambridge, 2014.

[17] Mac Lane, S., Moerdijk, I., Sheaves in Geometry and Logic, Springer, New York – Berlin – Heidelberg, 1992.
[18] McCulloch, W.S. and Pittts, W., A Logical Calculus of the Ideas Immanent in Nervous Activity, *Bulletin of Mathematical Biophysics* 5, 1943, 115-133.

[19] Mizraji, E., Lin, J., Logic in a Dynamic Brain, *Bulletin of Mathematical Biology* 73, 2011, 373-97, doi: 10.1007/s11538-010-9561-0

[20] Naotsugu Tsuchiy, Shigeru Taguchi, Hayato Saigo, Perspective Using Category Theory to Assess the Relationship between Consciousness and Integrated Information Theory, *Neuroscience Research* 107, 2016, 17

[21] Rogers, R., *Mathematical Logic and Formalised Theories*, North Holland, Amsterdam–London, 1971.

[22] Rosaler, J., Inter-Theory Relations in Physics: Case Studies from Quantum Mechanics and Quantum Field Theory, arXiv:1802.09350 [quant-ph]

[23] Rosen, R., Organisms as Causal Systems which are not Mechanisms: an Essay into the Nature of Complexity, *Theoretical Biology and Complexity* 1985, 165-203.

[24] Sher, G., What is Tarski’s Theory of Truth? *Topoi* 18, 1999, 149166.

[25] Simmons, H., *An Introduction to Category Theory*, Cambridge University Press, Cambridge, 2011.

[26] Substitution, in nLab, https://ncatlab.org/nlab/show/substitution

[27] Syntactic Category, in nLab, https://ncatlab.org/nlab/show/syntactic+category

[28] Tarski, A., *Pojęcie prawdy w językach nauk dedukcyjnych*, Towarzystwo Naukowe Warszawskie, Warszawa, 1933 (in Polish).

[29] Theory, nLab, https://ncatlab.org/nlab/show/theory

[30] Tsementzis, D., A Synthetic Characterization of Morita Equivalence, arXiv: 1507.02302

[31] Woszczyna, A., Heller M., Is a Horizon-free Cosmology Possible?, *General Relativity and Gravitation* 22, no 12, 1990, 1367-1386.