

Article

## The Self-Referential Aspect of Consciousness

Cosmin Vişan\*

### Abstract

Following the phenomenology that is revealed by the emergent structure of consciousness, the path will lead to the acknowledgement of consciousness having a self-referential aspect. By following phenomenological clues, properties of self-reference will be revealed. The two most prominent properties of self-reference will be shown to be inclusion and transcendence that will be shown to be found everywhere in the phenomenology of consciousness. Also, self-reference will turn out to be unformalizable, this imposing limits on what a theory of consciousness can ever achieve. The unformalizability of self-reference would be shown to be because self-reference: is itself, includes itself, and transcends itself, all at the same time. Nevertheless, unformalizability will be shown to be an essential feature needed to bring essences into existence.

**Keywords:** Consciousness, self-reference, emergent structure, phenomenology, unformalizable.

### Introduction

In my previous paper, “The Emergent Structure of Consciousness”[1], consciousness was shown to be structured on an emergent hierarchy of levels. Though the analysis was presented in great details, there are certain aspects of the analysis that are begging for more explanations. The main problems that emergence is rising are why is consciousness structured on a hierarchy of levels and what more can there be said generally about the functioning of emergence. This paper will continue the emergent path and will try to reveal even more about the functioning of consciousness as revealed by its emergent structuring. Because of this, the reader will need to be familiar with the phenomenology presented in the previous paper. The retaking of the previous analysis here will take too much space and time and there will not be much energy left for bringing new elements to the analysis. Where necessary, certain aspects of the previous analysis will be briefly reminded, but to fully understand the new elements that this paper will bring, a reading of the previous paper is strongly recommended.

Another thing that needs to be mentioned right from the beginning, is that because, as it will be shown, self-reference is unformalizable, difficulties might arise regarding to what exactly the central entity of this paper is actually about. The unformalizability will make self-reference to always be outside of any conceptual attempt at capturing it, so this will create difficulties in grasping what exactly self-reference is and how it should be regarded. In order to overcome this fundamental shortcoming, lights will be shed on self-reference from many different angles, showing how it behaves in many different situations, thus helping the reader to form an idea as close as possible to what self-reference is. A coherent picture of self-reference will only surface

---

\* Correspondence: Cosmin Vişan, Independent Researcher. Email: visancosmin17@yahoo.com Note: This article was first published in JCER 8(11):

towards the end of the paper. Until then, there will only be various unconnected aspects of self-reference.

Also, until the last sections of the paper, all the other sections should only be regarded as tools designed to help grasping the nature of self-reference, not as actual correct information. They should only be regarded as maps that try to point towards the correct direction, but from those maps various information might be missing. Some might better highlight the path, some might better highlight the scenery, but none of them would actually be about self-reference. Self-reference will only be reached towards the end of the paper after the paths suggested by the maps have been traveled. Unfortunately, due to its unformalizable nature, I don't see a more straightforward way of illuminating self-reference. Hopefully, the reader will follow along the specified paths and in the end, he will finally encounter self-reference in its bizarre apparently contradictory nature.

## **The unformalizability of self-reference**

Because no formalization of self-reference is possible, is better to start with this aspect in order to avoid chasing impossible goals. Let's see what is self-reference in the vaguest sense and why it is unformalizable. A classic example where self-reference is mentioned is as follow: You first look at a unicorn. Then, you reflect on you looking at the unicorn. This is a trivial and easy to understand example. Fortunately, in order to show the unformalizability of self-reference we don't need any more complicated examples. Let's see thus why self-reference cannot be formalized. Let's say that you want to capture somehow the fact that you are seeing a unicorn. This seems simple. You just take a paper, and you write: "I see a unicorn.". This is what a formalization of consciousness is supposed to look like and it seems that you just did it. But on a careful second look, you notice that by writing down that sentence, you just took yourself outside of the very quale that you wanted to capture.

Now, instead of being in the state that you wanted to capture, you find yourself in another state, namely: "I write on the paper that I see a unicorn." So, all of a sudden, you realize that what you thought that was a successful attempt at formalizing consciousness, just slipped through your fingers. One thing that you can do at this moment, is to try again to capture consciousness. So, you take another piece of paper and you write the new state in which you find yourself: "I write on the paper that I see a unicorn.". But by doing this, you again exited the very state that you tried to capture, and you ended up in the new state: "I write on paper that I write on paper that I see a unicorn.". Thus, the next realization that you come up to is that no matter how many times you try to capture your current state, the very act of trying to capture that state will take you out of that state. So, you start to envision the possibility that maybe there is something about consciousness that will forever escape formalization.

Having seen this first example of how consciousness escapes formalization, we can try to pin down more exactly what is the actual reason for consciousness escaping formalization. This reason is the nature of self-reference. Let's see exactly how this works. First thing to mention, is

that, of course, formalization doesn't require writing on a piece of paper. It suffices if it is all done in one's own mind. So, if we want to formalize the act of seeing a unicorn, we need to reflect back on the state of seeing the unicorn, so we need to form the state: "I know that I see a unicorn". But this has the above mentioned effect of taking us out of the very state that we want to capture. So, let's write more precisely the two states that we are dealing with.

***The first state is:*** "I am in the state "I see a unicorn"."

***The second state is:*** "I am in the state that I know that I am in the state "I see a unicorn"."

So, what we see here is actually an "I" that by trying to capture its current state, it finds itself in another state. The way the "I" is trying to capture its current state, is by reflecting on that state. But since that state is part of itself, the "I" is reflecting on itself, so it is self-reflecting. And here we identify the true reason that consciousness escapes formalization: The act of self-reflection takes the "I"/Self out of that state, so it makes impossible to capture it in a formal manner. This act of self-reflection is what we take a first vague meaning of self-reference to be. Thus, we have our first clue that self-reference is unformalizable. For every "*I am X*" state in which the Self finds itself, there is always another "*I am "I am X"*" state that the Self is pushed into, if it tries to formalize the state "*I am X*". And the reason is simple: in order to formalize the current "*I am X*" state, the Self must think about it, but by thinking about it, it becomes that new thought, so it takes the new form "*I am "I am X"*", which replaces the very state that the Self wanted to formalize.

This is only just the first and easiest example where we spot the unformalizability of self-reference. Throughout the paper, by following along the emergent structure of consciousness, we will spot many other ways in which the unformalizability of self-reference will pop up. Those cases will become more difficult and an increased attention will be required in order to appreciate their true significance.

A note to make here is that even though it might appear that unformalizability represents a dead end in any attempt to gain more knowledge about consciousness, it will be shown that we can actually gain more knowledge about consciousness if we accept this fact. It will be shown that the entire emergent structuring of consciousness is possible precisely because self-reference is unformalizable.

## **General ways to regard self-reference**

Before looking in details at the more intricate manifestations of self-reference that are responsible for the emergent structure of consciousness, let's familiarize ourselves with some simple ways of looking at it. Some cases of self-reference that appear in literature are those of Russel's Paradox[2] in Mathematics, or the Liar Paradox[3] in Logics as in the sentence "this sentence is false". Why are they paradoxes? Or in other words: Why are paradoxes such as these apparently possible? Is there something wrong with the universe? Or is there something wrong with how these problems are being formulated? I will argue for the second option, the reason

being that these cases are trying to make use of self-reference in improper ways. They define some random entities, like the entity “this sentence is false” and they attribute to those entities the property of self-reference without ensuring that the property of self-reference can even be attributed to those entities. And because of this careless construction of those entities, paradoxes appear. So, let’s see why self-reference is used in improper ways in these cases. After all, “this sentence is false” sounds like a perfect self-referential entity: it is a sentence that makes a statement about itself, so what might be wrong with it? Why wouldn’t it be a self-referential entity?

The reason stems from the nature of Mathematics or Logics. Even though they are treated and researched as if being entities that exist independent of consciousness, in the end they are actually entities that exist within consciousness. I will not argue here for problems like “Is Mathematics discovered or invented?”. But even if Mathematics is discovered, it still is an entity that lies in consciousness. The only Mathematics that we discover and will ever discover, is a Mathematics that lies in consciousness, so it cannot be an entity that is fully independent of consciousness. Therefore, the Mathematics that we discover is bound to be a Mathematics that is in some way connected to consciousness. Therefore, to attempt to formulate a Mathematics that bears no connections to consciousness, will inevitably lead to paradoxes such as Russel’s Paradox. To invent entities such as “this sentence is false” as if they are entities that exist independent of consciousness, will inevitably lead at some point or another to inconsistencies. To invent a self-reference that exists independent of consciousness, will be an improper way to invent such entity.

So then, how are we to sort this situation and to give proper treatment to the entities that we want to use in our analysis? Drawing from the previous section, self-reference appears to be related to the action of the Self reflecting back on itself. And this time, this look at self-reference is indeed a proper one. The reason is that the Self is an actual real entity (as opposed to the entity “this sentence is false”). And as a real entity, the reflection on itself is also a real entity. So, we come up with the first proper look at self-reference: ***Self-reference is the action of the Self of looking back at itself.*** This is the only self-reference that can exist, for the reason that the Self is the only real entity that can refer back to itself. Entities like “this sentence is false” or “the set of all sets” are just mirages, “sentence” and “set” not being entities with independent existence, but being ideas in consciousness. Therefore, we arrive at a view of self-reference that is not the usual one that is found in literature. But given the fact that we showed that the notion found in literature is an improper one, and the view that we will employ in this paper is a proper one, we are assured that we are in the possession of a good view of self-reference. As it will be shown later on, this view is not complete. But for the beginning, it is enough to understand that the proper meaning of self-reference is the act of the Self of self-reflecting.

## Levels of self-reference

In our quest of unearthing self-reference, we will use in what follows some formalizations. But because, as we shown, self-reference is unformalizable, we will need to always keep in mind that the notations that we will employ will be merely pseudo-formalizations. The decision to employ them nevertheless is that they will prove useful in revealing properties of self-reference. But at all time, the reader must remember that they are not true formalizations. They might appear at first convincing formalizations, but later on will be shown why they are merely pseudo-formalizations.

So, let's say that we want to capture self-reference somehow. What would be the easiest way in which we can write down self-reference? Given that, as we saw, self-reference is the act of the Self of self-reflecting, this can be captured in the easiest form as follow:

**I am.**

The Self reflects back on itself and notices its own existence. Note here, as I also warned throughout my previous papers, that the Self must not be confused with the ego. The Self looking back at itself is not me, John, looking back at myself. The Self is the ontological entity that lies at the heart of consciousness and gives consciousness its subjective ontology. The Self is also involved in the ego, but the ego is just one entity among others that are present in consciousness.

We can represent the "I am" as in the Figure 1.

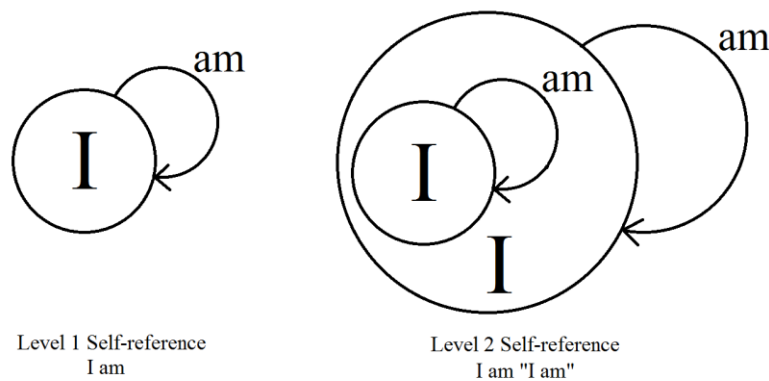


Figure 1. Levels of self-reference

We can imagine self-reference in a metaphorical way, as follow: The "I" is sending its arrow-like reference into the unknown. If the arrow doesn't hit anything, everything remains in the dark. But if the arrow returns to the "I", by the act of hitting the "I" the "I" acknowledges its own existence. Only through this acknowledgement that the "I" exists. Only through self-validation something can exist. So, "existence" is the act of the Self of reflecting back on itself. "I am" is an eternal logical structure in which the "I" is self-validating itself by sending the reference back to itself.

Now, once the Self validated itself, a new object appeared into existence. And that object is the “I am”. Now, the Self can identify with this object and by sending its reference back to itself, it finds itself as being:

**I am “I am”.**

This way, self-reference reflected back on itself and created a second level of itself. The process can continue indefinitely and self-reference can send further arrows to itself, reaching higher levels, like the third level:

**I am “I am “I am””.**

This seems rather an abstract analysis. We need to ground it in phenomenology in order to assess its validity. So, let’s finally look at some of the phenomenology revealed by the emergent structure of consciousness and see how it relates to this functioning of self-reference of increasing in levels.

## **Levels of phenomenology**

As disclosed in the previous paper[1], consciousness is structured on a hierarchy of levels. Since we see here that self-reference can reflect back on itself indefinitely and thus creating various levels of itself, is natural to try to make a connection between the levels of self-reference and the emergent levels of consciousness. For this purpose, we will employ the most basic levels of the emergent structure of consciousness, namely: *Self*, *Vividness*, *Diversity* and *Memory*. Let’s remind here the qualities of these levels:

*Self* – the quality of existence, and everything that comes with this: self-reference, ontological subjectivity, etc.

*Vividness* – the quality of “more of itself”

*Diversity* – the quality of diversity

*Memory* – the quality of memorizing

We will now try to make connections between these levels of consciousness and levels of self-reference.

The first connection that we make, is already made, since the quality of the Self is already identified with self-reference. So, for Self, we will have:

*Self*: I am.

Going to Vividness, and reflecting on its quality of “more of itself”, this quality suggests us a kind of self-reference that by reflecting back on itself, it become more than itself. But we already have this. This is the second level of self-reference. So, we have:

*Vividness*: I am “I am”.

Going to Diversity, thing starts to become a little subtler. Since Diversity has the quality of diversity, and since the third level of self-reference (I am “I am “I am””) doesn’t suggest such a quality, another form of self-reference is required for Diversity. So, let’s try something that suggests a diversity. Let this form be:

**Diversity:** I am “I am” & I am.

Going to Memory, things become even more complicated. Since Memory has the quality of memorizing, what combinations of “I am”s can we make in order to obtain something that suggests memorizing? Let’s reflect a little on what memorizing entails. Memorizing entails something that stores forever something that has been. For example, we see a chair and we store it in Memory. Given the fact that the chair is a quale, it comes with a whole emergent structure. So, to store a quale, Memory must be able to act upon the entire emergent structure of the quale of chair (shapes, colors, black-and-white, etc.). So, the memorizing process seems to be something that has to do with what a quale is in its entire emergent structure. The second part of trying to find the form of Memory is to note that Memory emerges from Diversity. And as a freshly emerged level, the only quale that Memory can store within itself, is itself. So, the memorizing process must act on the entire emergent structure of Memory. So, Memory, in order to memorize itself, must have a form that contains within itself all of its emergent structure, so all of its previous levels: Self, Vividness, Diversity. Thus, the form that we identify for Memory is:

**Memory:** [Diversity] & [Vividness] & [Self] ⇔ [I am “I am” & I am] & [I am “I am”] & [I am].

We will not go further with the next emergent level from Memory, namely the level of Time, because the analysis becomes quickly complicated. For the purposes of this paper, these first 4 levels of consciousness are enough to expose certain traits of self-reference.

As we can see, if we take phenomenology to be expressible in terms of self-reference, the levels of self-reference required are not only the linear ones that we saw on the previous section, but it appears that each new object that self-reference creates is then available for the next self-reflection to incorporate it. For example, after the first two levels are created, *I am, I am “I am”*, self-reference can either self-reflect on only the second level, giving birth to a third level of the form *I am “I am “I am”*”, or it can self-reflect on both the available objects, giving birth to a third level of the form *I am “I am” & I am*. Therefore, let’s assume that all these self-reflections are possible and let’s make a table with all the possible combinations. By “*level*” I will name the number of times self-reference reflects on itself, and by “*type*” I will name the number of objects self-reference incorporates in itself for a particular level. The results are shown in Table1 for the first 4 levels and the first 3 types. We also specified in the table the 4 levels of phenomenology that we analyzed. Thus, we can see the following characteristics for the first 4 emergent levels of consciousness:

**Self** is a Level 1/Type I self-reference

**Vividness** is a Level 2/Type I self-reference

**Diversity** is a Level 3/Type II self-reference

**Memory** is a Level 4/Type III self-reference

A natural question arises here: What about all the other manifestations of self-reference? Do they contribute in any way to consciousness? My answer is that I don't know. They might be, but it might be a different consciousness than our human consciousness. It might be the case that all are equally valid and they manifest themselves at some point, but due to evolutionary pressures they might not provide survival advantages so they are filtered out. Or they might be suitable only for certain ecological niches and not for other. Maybe they offer survival advantage for elementary particles, governing thus Particle Physics, but when the complexity increases and one gets to more evolved biological organisms, they might not present survival advantage. Or they might be part of our human consciousness, but my analysis failed to reveal them. At this point I don't have an answer. At least in principle, they all seem valid manifestations of self-reference.

**Table 1.** Level/Type classification of self-reference

Level Type	1	2	3	4
I	I am ( <i>Self</i> )	I am "I am" ( <i>Vividness</i> )	I am "I am "I am""	I am "I am "I am "I am""
II			I am "I am" & I am ( <i>Diversity</i> )	I am "I am "I am"" & I am "I am"  I am "I am "I am"" & I am  [I am "I am" & I am] & I am "I am "I am""  [I am "I am" & I am] & I am "I am"  [I am "I am" & I am] & I am
III				[I am "I am" & I am] & I am "I am "I am"" & I am "I am"  [I am "I am" & I am] & I am "I am "I am"" & I am  [I am "I am" & I am] & I am "I am" & I am ( <i>Memory</i> )  I am "I am "I am"" & I am "I am" & I am



## First glimpses of inclusion and transcendence

Let's take now these 4 levels of phenomenology and see the first glimpses of the properties of inclusion and transcendence. We already know from the analysis of the emergent structure of consciousness[1] that each emergent level is more than the level from which it emerges and that it cannot be reduced to that level. But let's see now the same thing in the identified self-referential forms of these phenomenological levels.

*Self* → *Vividness*. We can see how the <I am "I am"> form contains the form <I am>.

*Vividness* → *Diversity*. The <I am "I am" & I am> form contains both the <I am "I am"> and the <I am> form.

*Diversity* → *Memory*. The < [I am "I am" & I am] & [I am "I am"] & [I am]> form contains all the <I am "I am" & I am>, <I am "I am"> and <I am> forms.

This is in perfect accordance with the general phenomenology of emergence as revealed in [1], where for example, a full visual scene doesn't include only the objects, but all that the objects bring with them: shapes, colors, black-and-white, temporal passage, Self, etc. We see thus that the self-referential approach to consciousness follows beautifully the phenomenology of consciousness. We are thus encouraged to continue our goal of understanding self-reference.

Of course, it is unlikely that the entire study of consciousness can be cast in term of "I am"s, for example red being unlikely to be expressed in terms of "I am"s. But this is not needed. As it will be shown, self-reference is actually not the "I am". "I am" is just one form of manifestation of self-reference. But self-reference can take many other forms (actually, all qualia are forms of manifestations of self-reference), so even though red might not be expressible in terms of "I am"s, it will still be a form of manifestation of self-reference. So far, we are only exploring maps that are pointing towards self-reference. We are not yet in the position to tackle self-reference directly.

Let's now reveal better what inclusion is. For this, let's take these first 4 phenomenological levels and do some analysis. The fact that <I am "I am"> contains the form <I am> is not what inclusion is. Inclusion is a little subtler. So, let's see what it actually is. Since self-reference is unformalizable, we shouldn't look for its properties in the pseudo-formalizations that we write down. We need to look behind them and understand more abstractly what is going on. Therefore, let's give a guiding definition for inclusion:

**Inclusion is when an entity X obtains the character of another entity Y.**

The best example here is when a color emerges from shades of gray. The way in which the color includes the shades of gray is by obtaining the character of darkness-lightness. We notice here that when a color includes the shades of gray, the shades of gray as such are nowhere to be seen. Their individuality disappears as such, but it appears undercover in the fact that the color displays a darkness-lightness variability.

Therefore, taking the **Self** → **Vividness** case, we should not look for the <I am> form of the Self in the <I am “I am”> form of Vividness. The form of the Self must disappear as such when included in the form of Vividness. Therefore, we need to be subtler in exposing exactly how inclusion works for this case. We need to look behind the form and see the essence. For example, the essence of the shades of gray is the darkness-lightness spectrum, not the visual shades of gray as such. So, what is inherited/included in the emergent level of colors are not the visual shades of gray as such, but their essence of darkness-lightness. So, what about the essence of the Self behind the form <I am>? What possible essence can we spot behind this form? What about the essence of “unity”? Let’s see if this essence is inherited/included in the form of Vividness. If we look at the form of Vividness <I am “I am”>, we see that indeed it is a unity. So, we are on the right track. Also, we know from the phenomenology of the emergent structure of consciousness, that a certain quality is inherited all the way up in all the levels that emerge along a particular branch. So, when searching for inclusion, we must make sure that once the essence of a particular form has been identified, it then needs to be found in all the forms that emerge from it along a particular branch. So, if the essence of the first self-referential level, <I am>, is the unity, we must identify the unity everywhere in consciousness. And indeed, we do this, all qualia being unities, consciousness being unified.

If we take the next case, **Vividness** → **Diversity**, the analysis now moves faster. The essence of Vividness behind the <I am “I am”> form is that of “more of itself”, essence that is then spotted both in the forms of Diversity and Memory, given the fact that they are composed of multiple objects.

For the **Diversity** → **Memory** case, we identify the essence of Diversity as being diversity, because it is formed out of 2 different objects, and then we also identify this essence in the form of Memory, Memory being formed out of 3 different objects.

We thus identify a pattern here. While Self, Vividness, Diversity, Memory are *forms* of consciousness, they have behind the scenes *essences* that are the ones that are propagated higher in levels and manifest themselves in the higher forms of consciousness. We also begin to suspect that the mechanism through which this propagation higher in levels happens is by self-reference looking back at itself, with each self-reflection self-reference passing the essences higher in the hierarchy, thus including them in the higher order qualia. Also, each self-reflection brings into existence new essences that didn’t exist in the previous self-reflections. For example, while the form Vividness contains the “unity” essence of the form Self, it also brings into existence its own essence of “more than itself”. Then, while the form Diversity contains both the “unity” essence of the form Self and the “more than itself” essence of the form Vividness, it also brings into existence its own “diversity” essence. The same then goes when all these essences are passed on further to the form Memory, Memory bringing into existence its own essence of “memorizing”. We thus have the first glimpses at transcendence.

## Essence and Existence

Let's see some more examples from the higher levels of the emergent structure of consciousness and learn better to recognize the essences in a form. Let's take my favorite hierarchy of the written language. And let's take the levels of the forms *Shapes, Letters, Words, Sentences*.

**Shapes:** essence of “visual objects”: entities with spatially defined boundaries.

**Letters:** includes the essence of the Shapes, thus becoming themselves visual objects, and transcends it, bringing into existence its own essence of “unities of language”.

**Words:** includes the essence of the Shapes, being themselves visual objects, includes the essence of the Letters, being themselves unities of language (just more complex than letters), and transcends them all, bringing into existence its own essence of “carriers of linguistic meaning”.

**Sentences:** includes the essence of the Shapes, being themselves visual objects, includes the essence of the Letters, being themselves unities of languages (just more complex than both letters and words), includes the essence of the Words, being themselves carriers of linguistic meaning, and transcends them all, bringing into existence its own essence of “carriers of ideas”.

Of course, the precise essences and the precise forms into which essences are brought into existence might not be exactly the ones presented here. But what is important here to notice is that there are some entities, called essences, and some other entities called forms, the relation being that a form can include more essences. Traditionally in Philosophy, this is the problem of Essence and Existence[4].

**An essence is what makes a specific entity to be what it is.**

**An existence is an actual instantiation of an essence.**

For example, we can tell to a blind person about the essence of redness, but until he brings it into existence, he cannot know what redness is.

Let's reflect now on the dynamics of bringing essences into existence. If we look carefully, we see a drastic difference between essence and existence. Let's take for example the essence of “chairness”, the essence that makes a chair to be a chair. Let's say that by some means, we want to bring this essence into existence and experience the quale of a chair. But this is not possible directly. Even though essences are pure: chairness is chairness, existences/forms are not. You cannot just have the “pure” form of chair. The form of chair, in order to be/to exist, it needs to be made out of shapes, colors, shades of gray, spatial extension, vividness, first person ontology/Self, etc. So, we see that whatever mechanism might be responsible for bringing essences into existence, it is not a straightforward mechanism, but is one that needs to bring an entire emergent structure in order for that particular essence to be cast into a form.

## Meeting self-reference

Equipped with all these new tools, we are finally ready to meet self-reference. I will present first an abstract theory of self-reference and then I will exemplify its manifestation as it is seen in the emergent structure of consciousness.

**Let there be an entity, called self-reference, with the property of looking back at itself.**

What can we say about such an entity? The first thing is that because self-reference has the property of looking back at itself, by employing this property, it finds itself when it looks back at itself. Second, the object that it finds by looking back at itself, namely itself, must lie somewhere. That object lies in the looking-back-at-itself of self-reference, so it lies in self-reference. We thus see, that by looking back at itself, self-reference does two things: it includes itself within itself and, by including itself within itself it transcends itself. We are thus dealing with an entity that both includes and transcends itself. We thus see the true reason why self-reference is unformalizable. We are dealing with an entity that is at the same time: itself, less than itself, and more than itself. If this were to be formalized, it would look something like:

$$x=x$$

$$x<x$$

$$x>x$$

which clearly would be simply absurd. At this point, the reader might try to conclude that we actually reached contradictory results, so self-reference must be rejected based on its inherent contradictions. But this conclusion doesn't stand. The reason is that the non-contradiction principle is a principle that has to do with formal entities, with actual existing entities. But it doesn't say anything about entities that are outside of the realm of existence. Indeed, if we would try to formalize self-reference, we would throw ourselves directly in the midst of contradictions. But something surprisingly happens if we let self-reference unformalizable: it can maintain all these properties without contradicting anything, not even itself.

Are these arguments sound? Or am I just trying to trick the reader into accepting meaningless words play? Because we are used with the forms which consciousness shows to us and the forms into which it puts us (as for example our bodies), we developed a Logic and a Mathematics based on these forms. And since forms are definite entities: red is red, sweet is sweet, etc., we developed the non-contradiction principle. But this principle is solely a principle of how forms are. If there are other entities that are not forms, this principle can't say anything about them. So then, is self-reference such an entity? Yes. I will claim that self-reference is an unformalizable non-existent entity. But then, if it doesn't exist, what good does it make? We will see that it does all the good in the world: it brings consciousness into existence. We will see that without self-reference, all that would have ever been would have been essences. But those essences would have stayed forever in a non-existence state. We will show how self-reference is the bridge between essence and existence, being the one that by taking essences and putting them through the mechanism of looking-back-at-itself brings them into existence, thus giving birth to the only entity that exists: consciousness.

Of course, as the reader notices, I employ words such as: self-reference “is”, chairness “is” chairness. This is just an artefact of ourselves being entities that are forms, and we cannot escape thinking in terms of forms. So even though there “are” (again the same inability to think outside forms) entities that are outside the realm of forms/existence, we cannot escape thinking and speaking about them also as if they are forms. So, if we are trapped in the realm of forms, how are we to even think about formless entities? The way in which we can think about them is to notice their effect upon the forms. The major effect that formless entities such as essences and self-reference have upon the realm of forms is the emergent structure of consciousness. We will present a mechanism for emergence based on self-reference acting on essences and making them look-back-at-themselves. We will show that even though formless entities are not subjected to the rules of forms, they are nevertheless inseparable from forms, forms only existing because of the mechanisms of formless entities, and formless entities always manifesting themselves in the forms, thus formless entities being within our grasps of researching them.

Let’s reflect a little more on the nature of self-reference in order to make it as clear as possible. One way to do this is to contrast it to a fractal. The reader might be tempted to view self-reference as a sort of fractal. And it is important to emphasize that they are radically different entities. A fractal is not self-referential, but it is recursive. A fractal takes its output and use it as an input in an iterative manner. But the looking-back-at-itself of self-reference is not iterative, but it is its very nature. Also, the self-similarity of a fractal is only obtained for certain zoom levels: you need to zoom in a certain amount in order for the fractal to become similar to itself. Self-reference on the other hand, is always itself, it doesn’t need to do anything. Is in its very nature to be always itself, to always find itself when it looks back at itself, and it eternally looks back at itself, so it is always itself.

Another comparison would be with a mirror. But the difference between a mirror and self-reference is that in order for the image in the mirror to form, a lot of physical processes need to happen, like for example photons traveling from to object to the mirror and back to the object, while self-reference is timeless. Self-reference looks back at itself in a timeless manner. The looking-back-at-itself is not an action that self-reference does, but it’s a property, is what self-reference is. Time is only a form that self-reference gives birth later on.

Having made these comparisons, we now understand better that the looking-back-at-itself is not a process, but a nature. And the finding of itself inside itself is also not a dynamical process, but is also a nature. Also, the fact that by finding itself inside itself leads self-reference to include and transcend itself are also not dynamical processes that happen over time. They are all the nature of self-reference. So, all these properties that self-reference has, are all parts of self-reference at the same time. And as mentioned before, this is alright, because self-reference is an unformalizable non-existent entity. Let’s summarize the properties of self-reference that we have so far:

**Identity:** self-reference is itself

**Inclusion:** self-reference includes itself

**Transcendence:** self-reference is more than itself

One important thing to clarify further before going to apply self-reference to the emergent phenomenology of consciousness, is the property of identity. The property of identity is important not only in itself, but also for the two other properties. When self-reference looks back at itself, the only way in which it can include itself is if it identifies with itself when it finds itself. If there would be no identification, self-reference even though it might see itself inside itself, it wouldn't include itself. In order for the inclusion to happen, self-reference must first identify with the itself that it finds inside itself, and only because of this identification that it can then be included in itself and then transcends itself. The true importance of this might not be clear at this moment, but it will become clearer when we will go to phenomenology and we will see that emergence can only happen if self-reference identifies with itself that it finds inside itself. And also, it can transcend itself only if it identifies with its transcendent form.

## **Bringing consciousness into existence**

Having presented an abstract analysis of self-reference, it is time to see how self-reference is responsible for the emergent structure of consciousness. As we saw in the previous paper[1], the most basic level of consciousness was the Self. Having now the functioning of self-reference presented, we can see how this is done. First, when self-reference looks back at itself, it finds itself. Finding itself, it identifies with itself. So, it goes from being a formless entity, to being an entity that has form. By identifying with itself, self-reference acquires the form "I am [myself]." Acquiring this form, the Self/I is created. So, self-reference is at the same time both a formless entity and a form entity. Its nature remains forever formless, and thus all its apparently contradictory properties (identity, inclusion, transcendence) remain, but precisely because of these properties, self-reference can acquire form. By transcendence, self-reference transcends its formless nature and acquires form. By identity, it identifies that form as being the Self. And by inclusion it maintains its formless nature included in the newly acquired form. In a way, the Self is self-reference itself, but is only that part of self-reference that identifies itself with form. Form comes about because of the formless dynamics of self-reference of looking-back-at-itself. Form and formless are inseparable. Form cannot exist without the dynamics of the formless, and formless cannot manifest its dynamics/be itself without bringing form into existence.

We thus have our first explanation of how self-reference brings forms into existence. Of course, everything about this explanation looks suspicious. It seems like an ad hoc explanation. Why would the three properties of self-reference work like that? In order to better see why this is the mechanism of self-reference that brings forms into existence, let's have a look at a clearer example. Let's take the retentional passage of time. We know from the analysis of the emergent structure of consciousness[1], that Time emerges out of Memory. This is done by the present moment being itself a memory and then taken as a memory and fed-back into the present moment. This way, the present moment is both itself as present and itself as past, and thus a feeling of a passage is obtained. Let's see how this is obtained by the mechanisms of self-reference. I will use capital letters for forms and small letters for essences.

**Identity:** time is itself.

**Inclusion:** time, by looking back at itself, it finds memory. This is inclusion because: Self-reference (time) includes itself (memory). This does makes sense, because both time and memory are self-referential entities. Self-reference doesn't discriminate between its various manifestations. All it cares is to be itself.

**Transcendence:** Time, by including both itself (time) and memory, becomes more than itself as an essence and becomes a form.

The final result is Time having a feeling of a passage.

Let's analyze carefully and understand what it happening. First of all, at this moment I must state that I don't know how self-reference actually picks up essences. But given phenomenology, we must assume that somehow it actually interacts with essences. So, for the present moment, let's just take for granted that self-reference interacts with essences. Let's see how this interaction takes place.

First, self-reference meets the essence of time and it identifies with itself, thus the essence of time acquiring itself a self-referential character, and thus time becoming itself. But it is not yet a form. As we saw in the Essence and Existence section, in order for an essence to come into existence, an entire emergent hierarchy must be brought up. This is done in the next steps of self-reference looking back at itself. Once time has become a self-referential entity, it looks back at itself and finds memory; memory is already a self-referential essence, having acquired the self-referential character in a previous step of self-reference looking back at itself. Now, as we stressed in the previous section, inclusion can only take place if self-reference identify itself with the object that it finds inside itself. Since the object that it now finds inside itself, namely memory, is also self-referential, self-reference can identify with it, so memory is included in time. Notice that memory already has its entire emergent hierarchy within it, so by including memory, time includes all the necessary emergent hierarchy. But it is still not a form yet. For the moment it is only a self-referential essence (from Identity) and includes the emergent hierarchy of memory (from Inclusion). But in order to become a form, it must transcend itself. So, for the third step, time includes both itself as a self-referential essence and memory as a self-referential essence, and by doing this, it becomes more than itself as a self-referential essence, thus becoming a form, so coming into existence.

Notice that once time comes into existence, thus becoming Time, self-reference continues to identify with itself, so self-reference will also be Time. Self-reference is a bridge between essence and existence, thus self-reference is both of them. So, Time is a form of manifestation of self-reference. Since this mechanism applies to all the emergent levels of consciousness, all qualia are forms of manifestation of self-reference. We fail to notice this in day-by-day life. In day-by-day life, we just see objects and live our lives. We see a chair in front of us, and we perceive it as an object independent of consciousness, and clearly not an object that has anything to do with self-reference. But as we see now, all qualia are brought into existence through the mechanism of self-reference that: first identifies itself with the essence of the yet-to-be quale,

then by looking back at itself it includes within the newly identified essence an entire emergent hierarchy of essences for the identified essence, and then, by transcendence, it finally brings the entire emergent hierarchy into existence, thus giving birth to qualia. And finally, it identifies with qualia themselves. So, all qualia are in the end of the form “I am X.”, thus all qualia being self-referential.

## The story so far

Let’s see where we stand now in the process of understanding consciousness. In the paper “I Exist”[5], I presented a first glimpse of self-reference. At that moment, I only showed the eternal existence of the Self. Retrospectively, having now the new insights, that paper only dealt with the Identity property of self-reference. Also, “I Exist” left open the problem of time, not seeing how from an eternal unchanging Self, time was possible.

Then, in “The Emergent Structure of Consciousness”[1], by analyzing the phenomenology of consciousness, time was shown to be an emergent level that acquired its quality of passage from the self-referential quality of the Self. Thus, time was shown to be in agreement with the eternal unchanging nature of the Self, the apparent dynamic character of time being just a quality, like the quality of red, so time itself being as eternal and unchanging as the Self. The new problem that that paper left open is how inheritance is being done from level to level in a hierarchical manner.

Now, the present paper solved the problem of inheritance, showing that it is being realized by an entity called self-reference that has the property of looking-back-at-itself. Thus, with each look back at itself, self-reference enriches consciousness bringing new essences into existence in a hierarchical manner. However, this opens up new problems, the main one being the way in which self-reference picks up essences. And this is actually a very tough problem, that is general for any theory of consciousness. In its most general form, it is the problem of why we have precisely the qualia that we have. Why do we see exactly the colors that we see and not other colors? Why do we have the qualia domains that we have (sight, hear, touch, etc.) and not other qualia domains? This somehow seems to me an insurmountable problem, one reason being that somehow you need to know in advance what qualia to select. And evolutionary selective principles don’t necessary help.

Because if the number of possible qualia and qualia domains are infinite, then how did evolution select what it needed from that infinite number? For example, let’s take the first being that acquired sight. Let’s say that what that being first saw was black-and-white. And let’s employ the standard evolutionary principles of variation and selection. For example, you had 1 million beings that experienced 1 million different qualia, and only that being that experienced black-and-white had a survival advantage. But this type of argument doesn’t really stand, because if the number of possible qualia are infinite, then you would need either an infinite number of beings in order to select that one that precisely had black-and-white needed for survival in his environment, or you either needed a huge amount of luck that from the infinite qualia, those 1



million beings happened to experience precisely black-and-white. And what is even more suspicious, is that this “luck” didn’t happen only for sight, but so many other qualia domain were brought into existence for the different ecological niches, and they are continuously brought into existence every moment. Each time a being is faced with a problem, somehow the solution is brought into existence. So, there must be some other mechanisms with which self-reference interacts in order to select precisely the essences that are needed to be brought into form in order to help a being solve a problem.

This problem is romantically put in words by Kant when talking about schematism:

*“The concept of dog means a rule according to which my imagination can always draw a general outline of the figure of a four-footed animal, without being restricted to any particular figure supplied by experience or to any possible imagine which I may draw in the concrete. This schematism of our understanding applied to phenomena and their mere form is an art hidden in the depth of the human soul, the true secrets of which we shall hardly ever be able to guess and reveal.”*[6]

Another related problem is the problem of the stability of our consciousness. It appears that, in a healthy consciousness, once a quale or qualia domain have been brought into existence, they are kept throughout the entire life of the being. Why should we see the colors that we see throughout our entire life? Why should the colors not fluctuate, always new ones coming into existence and old ones disappearing? “Because of brain” is not a solution, because “brain” is itself just a quale in consciousness, so the organizing principles that keep our consciousness stable must be found directly in the workings of consciousness, not in the “brain”.

Self-reference might be a step forward towards solving these problems, or the solution might come from totally different directions. For the moment, we have placed one more stepping stone towards the understanding of consciousness.

## References

- [1] Cosmin Vişan (2017), The Emergent Structure of Consciousness (Part I), *Journal of Consciousness Exploration & Research* 8(8): pp. 604-627.
- [2] Russell, Bertrand, 1903. “Appendix B: The Doctrine of Types,” in Bertrand Russell, *The Principles of Mathematics*, Cambridge: Cambridge University Press, 1903, 523–528.
- [3] Beall, Jc, Glanzberg, Michael and Ripley, David, "Liar Paradox", *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.),  
URL = <<https://plato.stanford.edu/archives/fall2017/entries/liar-paradox/>>.
- [4] Aquinas, T., *On Being and Essence*, trans. Armand Maurer, 1968, Toronto: Pontifical Institute Medieval Studies.
- [5] Cosmin Vişan (2015), I Exist, *Journal of Consciousness Exploration & Research* 6(3): pp. 185-193.
- [6] Friedrich Max Müller (1922). *Critique of Pure Reason*, by Immanuel Kant, translated by Friedrich Max Müller, New York, The Macmillan Company.