Self-reference and self-consciousness

Michał Walicki

March, 1987

1 Introduction

The main object of this paper is the concept of self-consciousness. Since, considering this concept, I have found it very instructive to compare it with the seemingly similar notion of formal self-reference, I will give here an account of this comparison. The first part (section 2) is concerned with the self-reference in some formal systems. Example of the universal Turing machine (subsection 2.1) will illustrate the first approximate realization of the concept. Then the formula from the famous incompleteness theorem of Gödel (2.2) and the non standard universe of set theory (2.3) will be described to characterize the appropriate relation.

In these examples I am not going to concentrate on any particular theory of reference. "Reference" will be treated rather freely merely as a ground for distinguishing it from self-reference. What makes up the difference between both is the fact that the object which is referred to by another object shows itself to be identical with the referring one. The main problem in this context will be the meaning of the identity relation between both objects which can be eventually bestowed on them by a formal system.

The second part (section 3) is devoted to self-consciousness. For this purpose I have chosen the expositions given by Fichte in his "Introduction to the Theory of Science" [ITS] together with Sartre's "The Transcendence of Ego" [ToE] and "Being and Nothingness" [B&N]. It appears to me that there are many points shared by both philosophers. Section 3 will focus exactly on these common points. It should be emphasized that choosing the relation between consciousness and self-consciousness as the object of discussion, I have had to ignore the differences which exist between the two thinkers in other areas. The moral philosophy, the meaning of faith in God, the ontological status of the trancendent being and many other questions do not fall within the scope of the current presentation.

The main thesis of the paper is the incommensurability of the notions of self-reference and self-consciousness. The thesis will be developed gradually in section 3 by following the distinctions introduced in section 2. The structural similarity of the two notions is found in the existence of the aspect of duality (reference and consciousness, respectively) and unity (identity and self-consciousness). The important distinction concernes the relation of the two aspects and the character of the latter. Unlike the identity of the referring and the referred which is established on te basis of the dual aspect, "self-consciousness" will appear as a unified phenomenon which is inseparable from the duality of consciousness. The distinction between consciousness and self-consciousness is made only by the philosophical reflection. Here it will be retained for the sake of the clarity of the exposition.

The reconstruction of Fichte's concept of self-consciousness will proceed along with the gradual explanation of his philosophical deduction. The deduction begins with a pure self-consciousness and Fichte claims that from it one can derive empirical consciousness and the consciousness of a transcendent being. The different steps of the inference and their order seem to cover the essential features of the notion.

I also hope to show that the views of Fichte and Sartre on the present subject are far from divergent. Establishing a close connection between the two, we can thus use the elegant formulations of the French erudite to throw some light upon the thoughts hidden in the cryptic sentences of the German professor.

In the Summary (section 4) I give an abstract of the whole discussion. The particular emphasis put on the oppositions between the results of both preceding sections will clarify the relevance of considering self-reference in the inquiry about self-consciousness, as well as the inadequacy of mixing the two concepts together.

2 Self-reference in the formal systems

In this section we study the formal notion of self-reference on three examples: Turing machines (2.1), the incompleteness theorem of Gödel (2.2), and the non-well founded set theory (2.3). The last subsection 2.4 summarizes the main aspects of the notion.

2.1 Turing Machines

A Turing machine (TM) is a set of states and instructions working on an infinite input tape which is printed with the symbols from a given alphabet. An instruction tells what the machine is to do when in a given state it is reading a given symbol from the input tape. It can print a new (or the same) symbol, pass to another state (or remain in the old one) and move its reading head to the left or to the right along the tape. TM machine represents an algorithm which processes the data supplied to it from the input tape. It should not sound implausibly to call this process "reference" – the machine M refers to the actually processed input $x : M(x)$.

Now, there is obviously a fundamental difference between an algorithm and an input. Both are intimately connected (algorithm accepts only some kind of data and data must suit the algorithm) but it is machine which stands for an active agent – data are only a passive object to be modelled. The more complicated situations appear when we discover that algorithms can also be coded and provided as input to other algorithms. This trick makes it possible to approach construction of a self-referring system.

When an algorithm X is coded as a piece of data and fed into another algorithm M, the question arises as to what kind of information X contains. We would like it to cover its original meaning, that is, that it is an algorithm – description of a particular procedure, a particular TM. If $M(X)$ satisfies this condition we say that M simulates X. It means: for any possible input i for X, the result of the computation $X(i)$ is the same as that of $M(X(i))$.

Providing X as the only input does not lead anywhere because X is only a description of a procedure (a codified TM). The input tape must be completed with some input i (acceptable for X). The expression $M(X)$ means something like: machine M looking for the rest of the input or machine X waiting for some actual input.

The next step is construction of a machine which simulates not only one particular algorithm but all of them. Such a machine is called the universal TM and is a kind of a theoretical pre-computer. An universal TM M represents, like any TM, an algorithm. Consequently, it can be coded and fed into some machine which can simulate it. But M is an universal TM, so it can simulate itself. The question about the result of such a computation has an immediate answer: M simulating M will do the same what M does $M(M) = M$. Consequently, also $M(M(M)) = M(M) = M$ and so on. Reference made by M to itself amounts to the same as referring to nothing.

But, the doubt can arise, the input containing only M is not complete. As an input i is necessary to give $M(X)$ some actual meaning, so here too, some additional input should be supplied. It, however, would not change the situation very much. The above equation $M(M) = M$ means in practice an equivalence – for any input $i : M(M(i)) = M(i)$. M simply ignores itself, its reference to itself does not change the produced output at all.

There is also another difficulty. We may construct another universal TM – N. M and N are different having different definitions and also quite different ways of working. Then, for any acceptable input i, the results of computations $M(M(i))$ and $M(N(i))$ are identical, that is $M(M) = M(N)$. So is $N(N) = N(M)$. Since the machine is the active agent in the sense that it can "recognize" two different inputs by producing two different outputs, it is reasonable to ask it to recognize the difference between itself and another machine (when both are given to it as two inputs). At least, it would be a necessary condition for calling a machine self-referential. Failing to recognize this difference, M fails also to recognize its identity with itself.

This is rather an unessential lack from the point of view of the theory which is interested exclusively in the data processing and in the results of computations. The universal machine is important as such, as *an* universal machine and not as *this particular* universal machine (at least as long as the questions about effectiveness are not asked). Accordingly, the equivalences between machines may play much more important role than identities. Nevertheless, nobody will deny that M and N are two different machines.

It is a question of definition, but because of the above failure (to recognize itself), I would consider the case just described only as an approximation to the self-reference. Except that the referring object must be capable of being referred to (as M does when it is coded on the input tape), the identity between both must be granted in one way or another (M being indistinguishable from N can not be, properly speaking, said to be self-identical).

The present case is not included to show that the formal systems are incapable of developing more sophisticated concepts peculiar only to us, humans. It is only to indicate the main points, the rotation axis of the whole topic, and to illustrate the first step in achieving self-reference. This step consists in

1. coding the referring (algorithms) in the form characteristic for the referred (input). We have also observed that

2. there is no effect of feeding an universal TM with its own description – it follows from

3. the lack of an internal identity relation which could be recognized by the system

In the next sections we will have the opportunity to see how the formal systems can approach the problem of conceding the identity of the referring and the referred, as if, from outside. Here there is simply an insufficient formal framework for "recognition" of such an identity. We, on the other hand, who stand outside, can see the different definitions of what is inseparable for the algorithm. The machine, although it acts on the definitions (an algorithm in an input form is a codified definition of a TM), "recognizes" only by means of the output. It never refers to itself – it refers to something else, not even to another machine but simply to an input. The system seems to be to weak to develop fully a self-referential structure, even if we may feel that there is an analogy between $M(M)$ and a structure.

2.2 A formula referring to itself

The incompleteness theorem of Gödel (IT) can be considered as the formal version of the liar paradox. It shows existence of a special formula in a wide class of formal theories. The formula has the property analogous to the statement "This sentence is false." – it is provably true if and only if it is false. *We* know that it is true but it is impossible to prove this fact within the system. The system is therefore incomplete (can not prove all the true statements) and hence the name of the incompleteness theorem. The formula has also another interesting property – it looks somehow like $G(G(x))$, as if it referred to itself. Impression that the two strange features are necessarily connected is not correct. As in other paradoxes of this kind, there exists a non-self-referential version which causes the same problems. To take the example of "This sentence is false.". The two sentences A: "B is false." and B: "A is true." will do the same job. The paradoxical character lies not so much in self-reference as in the concept of truth and its application – which is not our subject matter. Consequently, I will not look at what is probably the most interesting in the IT – its seemingly paradoxical result. I will concentrate on some points of its proof and will attempt to trace the way and the meaning in which self-reference appears there.

First of all, let us notice the difference between the terms and the formulae in a formal language. Terms are interpreted as elements of some domain ("the real world"). Terms can be said to refer to their corresponding elements like a proper name refers to a person. But in our context it is more relevant to consider the formulae only, and the reference made by them to the terms. I am not going to differentiate between terms and objects or, in general, between the syntactic and the semantic side. Thus, saying that the formula "A is red" refers to A, we do not bother that A is only a name (that is "A") which denotes some individual. For us the referred objects are terms and the formulae are the referring agents : the terms are the ultimate objects of a formal theory, and the formulae (language) are the means for describing and referring to these objects.

The objects of "the real world" of the IT are natural numbers. The famous trick of Gödel is to code the formulae with numbers, so that to each formula there is assigned an unique number (the so called Gödel number – GN). In this way the formulae can be identified with numbers just like in the previous example the algorithms could have been given an input form. Consequently, the formulae (their GNs) can be referred to by other formulae and can be treated like usual objects.

The next step of the proof is not essential in our context. It consists in expressing some predicates (particularly the predicate "provable") within the language of the theory. The consequence of this step is that formulae can be described within the theory as, for instance, "(not)provable" – the theory can say "$G(x)$ is not provable". It can not be proved, in principle, whether some statement can be proved or not, but in order to get the final result it is sufficient to express provableness as a predicate.

What is more interesting for us is the last step. At this point the following formula has already been constructed:

> $G(x)$: "x has one free variable and x with its own GN substituted for this variable is not provable"

"x" figures here as the term to which the whole $G(x)$ refers. But it can be a usual number as well as a GN of some formula. In the latter case x can in turn have a term to which it refers. $G(x)$ says that it does have such a term. It is what the "free variable" stands for, except that the "free variable" is only a kind of empty place where some concrete term can appear. $G(x)$ says, furthermore, that in this empty place of the free variable we should put in the GN of the x itself. The result of this operation will be something like $x(x)$. And $G(x)$ says that that "$x(x)$ is not provable".

We have not reached the end yet. $x(x)$ is neither self-referring nor is it the final formula of the IT. We take now the GN of $G(x)$ and put it into the place of x (which itself is a free variable in $G(x)$). The result is roughly speaking $G\{G(x)\}$ or, using the formulation of Findlay:

> "The formula obtained by substituting the free variable in the following formula: {"the formula obtained by substituting the free variable in the following formula: x with its GN is not provable."} with its GN is not provable."

If $G(G(x))$ is false it follows that it is provable which would be a contradiction because one of the premises of the theorem is that we are dealing with a consistent theory. Consequently, $G(G(x))$ is true and from this fact there follows another one, namely, that it is not provable. This, however, does not enable us to call $G(G(x))$ self-referential any more than the implication $y = 5 \Rightarrow 4 < y < 6$ enables us to do so with respect to y.

The object referred by $G(G(x))$ is $G(x)$. And obviously $G(G(x)) \neq G(x)$ – they have different GNs. We can not think about the situation as if we had $G(x)$, took $G(x)$ and substituted it for x and got the same $G(x)$ referring to itself. After the substitution, it is no longer $G(x)$ which refers to $G(x)$ but $G(G(x))$ which refers to the inner $G(x)$. Like x is the object referred to by $G(x)$, so is $G(x)$ the object referred to by $G(G(x))$. Does it mean that the formula is not referring to itself?

We have the analogous situation with the sentence "Yields false when preceded by its own quotation." It is supposed to yield false when preceded by its own quotation. So we can try: " "Yields false when ... quotation." yields false when preceded by its own quotation." But now the expression "its own", according to its very nature, has changed the reference – it is referring to the whole (doubled) sentence. The inner sentence is not the quotation of the whole. It is only quotation of the sentence we had before the substitution – these two are equal just like $G(x) = G(x)$. But in order to get a false statement we would have to complete the new (doubled) sentence with its own quotation. This is beginning of the infinite regress. In the moment of substitution the previous identity disappears. The referring and the referred are no longer the same because we get a completely new formula. It is the reason for which $x(x)$ could not have been called self-referential.

But the case of $G(G(x))$ is not that straightforward. Its direct object (visible in the formal notation) is $G(x)$ but the formula contains a kind of rule for how to obtain a new object from this direct one. It says: "Take the direct object (x) and substitute its own GN for the free variable which must occur in x. The object (formula) you get in this way is not provable." It is non-provableness of this new object which is asserted by the formula. The direct object serves only as an intermediary step. If we take $G(x)$ itself to be this middle term (that is, if we say $G(G(x))$) we will then have to substitute its GN for x. The resulting object will be $G(G(x))$. The whole point of the trick is the smart use of $G(x)$ within $G(G(x))$ as the means for reconstruction of the whole $G(G(x))$ as its own indirect object.

We may think about the following situation. I am talking to a person P and I am actually the only one who is talking to him. Saying "the person speaking to P" I refer directly to P but I mention him only in order to point to myself. If I were not able to speak directly about myself ("I") such a procedure would be an indispensable tool for compensation of this lack. If, instead of $G(x)$ we substitute another formula for x (say $F(x)$) the procedure will be the same. The resulting $G(F(x))$ will say that $F(F(x))$ is not provable. Self-reference is lost in this case. $F(F(x))$ is as little self-referential as is $x(x)$. $G(F(x))$ refers directly to $F(x)$ and indirectly to $F(F(x))$. But the way we obtain this final object $F(F(x))$ from the direct one $F(x)$ is exactly the same as obtaining $G(G(x))$ from $G(x)$. If we imagine that except me and P there are two other persons, A and B, talking to each other, then the reference made through the expression "the person speaking to x" will depend on the actual substitution for the free variable x. If we put in B, the reference is made to A, if we put in P – to myself. The mode of reference is the same in both cases – it is determined by the manipulation with substitutions. But, unlike the universal TM which failed to distinguish itself from another universal TM, the relation between $G(G(x))$ and its object $G(G(x))$ is totally different from the relation between $G(F(x))$ and its object $F(F(x))$. The difference comes from the fact that the formulae (as numbers) can be treated by the theory in which the whole construction has been carried on. The axiom of identity, which is valid uniformly for all the objects of the domain (numbers), can be now applied to the formulae as well. That "the theory can recognize the fact" means that this fact can be proved within the theory. And the theory can prove, on the one hand, that $G(G(x))$ (as the referring) is equal to $G(G(x))$ (as the referred) and, on the other hand, that $G(F(x)) \neq F(F(x))$ (the identity and difference meaning the identity and difference of the corresponding GNs). Thus it is not only we who know that $G(G(x) = G(G(x))$. The formal system itself can recognize the fact yielding a formula referring to itself.

The constructed self-referential structure can be summarized in four points:

1. mixing of the two levels: the referring and the referred, by codifying the formulae as numbers (cf. point 1 at the end of section 2.1);
2. achieved self-reference is indirect;
3. the self-reference is a special case of usual reference, obtained as a result of particular substitution (this follows from 1 above);
4. the identity of the referring and the referred is now proved within the system (cf. point 3 in 2.1).

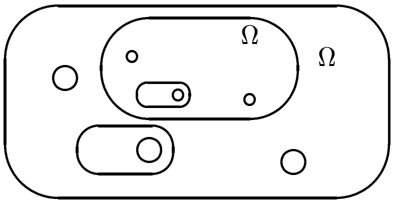
The next example is not going to add anything essentially new to this scheme. The points 1, 3 and 4 give us the vital features of the concept of self-reference. I will only illustrate them in another context making, by the way, some improvements to 2.

2.3 A self-containing set

The section presents a formal notion which seems to diverge (if not contradict) very strongly from the usual way of thinking about the mathematical objects and their definitions. Such thinking is based on the fundamental elements from which the whole study begins. All complicated and interesting structures are composed of these "elementary particles". An input tape is a sequence of letters from a given alphabet (ultimately the zeros and ones). Machine does nothing more than recognizes the single symbols and puts them together into bigger units – bytes, words, commands. Mathematics investigates collections of some well defined elements. Groups, categories, sets are gatherings of points, numbers, functions, other sets etc. Any mathematical structure is built from basic elements and their interrelations. The set theory expresses this feature in the axiom of foundation (AF): "The membership relation is well-founded." Less formally, it means what has just been said – every collection must be constructed out of some fundamental, irreducible entities. Given a set, we can pick some of its elements. If these are not the simple entities, we can again pick some of the elements from each of them. According to the AF the process of such picking the elements of the elements of the elements of ... must terminate after a finite number of steps. And it terminates when we reach the level of the fundamental, irreducible and indivisible elements. Each set is a hierarchical arrangement of such elements, their collections, collections of these collections etc.

The direct consequence of the AF is, that there is no element x which is a member of itself. If there existed an x such that $x \in x$, it would contradict the AF because it could not be built from any fundamental elements. We would have an infinite decreasing chain of elements which never reaches the bottom: $x \ni x \ni x \ni x \ni x$... We do not need to interpret the membership relation as "reference" (although I will allow myself to call it so) in order to make it relevant for our subject. A self-referential being is something which, in a sense, comprehends itself within itself, which contains itself like a subject thinking itself contains itself as an object. In the previous section we saw that there is no need for any mystery of this kind. Self-reference was achieved indirectly or, if one likes, implicitly. But if we had an x such that $x \in x$, then (having also $x = x$ as an instance of the identity axiom) it would amount to the same as a direct, immediate self-reference. Recently, some mathematicians have dispensed with the primary ground of the set universe. Denying the AF, they allow the sets to have no least elements. In a sense, one does not have control over such sets. We can not simply look at them and say "here it begins". We can not construct them because they have no beginning. We can only know that assuming their existence, we do not contradict the remaining axioms of the set theory.

In this new version (let us call it "not-well-founded" – NWF) there is a set satisfying $x \in x$. Unfortunately (or perhaps very significantly) it is the whole universe itself (denoted Ω). It is a member of itself in the same way as any other set is a member of it. Within the universe Ω where all the sets reside, there is also, as one of them, the whole universe itself. From Ω we do not need to proceed in any special way, we do not need to construct any new object in order to return back to Ω. NWF gives an Ω which contains itself entirely as itself – not only as an allusion to itself. It is rather an unjustified procedure, but if we want to appeal to our spatial intuition, it is like making the following absurd possible



where the inner circle is known to be the same as the outer one. We may observe here the analogous features to these possessed by the Gödel formula. Ω is its own element in exactly the same way as are its other members – the same membership relation applies to all cases ($\Omega \in \Omega$ and $a \in \Omega$ where $a \neq \Omega$). The same relation describes also the connections between other sets ($a \in b$). This amounts to saying that self-reference here, as in the previous example, is a special case of the usual reference. The identity $\Omega = \Omega$ is granted in the same way – by means of the identity axiom.

The points corresponding to 1-4 from the end of section 2.2 can be now expressed thus:

1. juxtaposition of the referring and the referred is implicit in the fact that Ω is homogenous, that is it contains only sets, sets of sets etc. Consequently the sets can only refer to (contain) other sets;
2. the direct self-reference is achieved by allowing the "groundless" sets;
3. this self-reference is again a particular case of the reference which is available to all (also not self-containing) sets;
4. the "self" of self-reference, that is the identity of Ω as referring and as referred is granted not inwardly, by Ω itself, but externally by means of the axiom of identity.

2.4 Self-reference in general

We can summarize the main ideas about self-reference. The following paragraphs correspond, respectively, to the points 1-4 from sections 2.2 and 2.3.

2.4.1

The point of departure is a system with two components: the domain of objects and the formulae or, more generally, the entities which can be referred to and the ones which can refer. The sharp distinction is visualized by the fact that only the former are the objects of interest for the theory – the latter provide only the means to carry out the inquiry. [1] Accordingly all the relations and statements are interpreted in terms of the objects. In particular, the identity axiom is valid for all the objects of the domain and only for them. It does not make any sense when applied to the formulae. These can be equivalent (like the machines M and N were) but their objective identity is not recognizable within the system. In order to reach self-reference:

1. there must be an object (not only a formula) capable of making reference and
2. the referring object must be proved to be equal (since it is an object and not just a formula such a proof makes sense) to the referred one.

Both conditions are independent in the sense that the objects can refer to other objects without being equal to them and, on the other hand, the identity of the objects ($x = x$ or $x = y$) does not imply their referential ability.

2.4.2

The construction of the referring objects (1 in 2.4.1 above) is carried out in the process called by us "the mixing of levels". In 2.1 it is coding the algorithms on the input tape, in 2.2 – the Gödel arithmetization (coding formulae as numbers). In 2.3 this step is immediate because the sets can only refer to other sets and can be referred only by other sets - there is no original heterogeneity of the referring and the referred. In this last example, the elimination of heterogeneity is the essential move. Without it the condition 2 from 2.4.1 could not be satisfied.

In this process the old objects do not acquire the referential ability. But all the referring agents can now be looked at as the usual objects. The ordinary relations preserve their meaning with respect to the old objects and acquire it with respect to the new ones. The latter can be equal, greater etc. in the same sense as the old ones – both groups make in this sense one. I want to emphasize this levelling (one can almost say collapsing), because we will later see the relevance of this point for distinguishing the character of self-reference and self-consciousness.

[1] We do not need to worry about the lack of this distinction in the last example. The point is that the referring can "return to itself" only if it becomes an usual object, only by being submitted to the ordinary relation of reference and identity – see 2.4.2.

2.4.3

Having collapsed the two levels, what remains is to find an object satisfying the condition 2 from 2.4.1. In the formulation the phrase "must be proved" was used. I have also been speaking about "the system recognizing identity" and the like. These expressions are to indicate that the argument of the kind "But it is a human being who in the last instance does all this, who proves, constructs, etc.." is not relevant here. It is true, and in the next section it will appear as the important observation, that in the case of self-awarness it is the very subject who himself recognizes his own identity and, consequently, that his self-identity is, so to speak, a living and lived identity – not a pure, external, axiomatic identity 1=1.

We can not help that whatever we do is done and recognized by us. When we write a computer program which produces its own copy, we see the identity of both and it suffices to call it self-repreoducing. But even if it happens within a bigger formal system which is able to see this identity (as in the examples from 2.2 and 2.3), it would sound strange if anybody claimed from this very program to identify itself with its product. Equally meaningless would be to want a formula itself to prove that it is equal to its object. If we want some object to recognize something as equal to itself, it must in advance have given its own identity. It can not see itself in another unless it can see itself in itself. But in a formal system it is not the object which provides us with any results – it is the system itself. It is the system which recognizes (proves) the facts about the objects. Accordingly, the immediate identity is granted by the axiom. The axiom, however, is not given through the objects. These, in themselves, are not identical. Our intuition and the knowledge of this intuition make us introduce the suitable axioms and relations. We can study the objects with the first order predicates but no such predicate can serve as a substitute or a foundation for the identity. Identity is not logically definable or, in the formulation of Frege "Since every definition is an identity, the identity itself cannot be defined."

In the case of $M(M)$ we saw the lack of the formal ground for granting the subject-object identity. To do this the universal TM was ... too universal. In both other examples the work was done by an instance of the identity axiom, that is, by the system itself. It explains also why the difference between the direct (Ω in 2.3) and the indirect (Gödel's formula in 2.2) self-reference is not significant from the formal point of view. Identity is simply identity and nobody is going to distinguish the two cases:

- where we start with $G(G(x))$, take the inner $G(x)$, substitute its GN for x getting $G(G(x))$ and then, comparing the point of departure with the point of arrival, realize that the relation between both is an instance of the identity axiom, and
- where we start with a single object, Ω, and perceive immediately that $\Omega = \Omega$, the fact which is then adjoined to the other one, namely that $\Omega \in \Omega$.

2.4.4

Thus a self-referring structure emerges from the simple reference. It is by no means something basic – on the contrary, it is a vary sophisticated case of reference. The self-reference is made in exactly the same way as reference to a distinct object. The externally granted identity completes the return to itself. Nobody will probably argue that there is a simple transition from self-reference to self-consciousness. Yet, the distinguished character of the former, the similarity of expressions, the particular quality which self-reference seems to have in comparison to reference, on the one hand, and the apparently similar quality of self-consciousness in its relation to consciousness on the other, all this can suggest at least some analogy between both. As I believe, not only there is no single step from self-reference to self-consciousness but any attempt to describe the latter in a way somehow parallel to the description of the former is an unjustified procedure. There is no significant or meaningful analogy.

3 Consciousness and self-consciousness

3.1 Between self-reference and self-consciousness

Before entering the disussion of self-consciousness, a few provisional remarks may be in order.

3.1.1 Self and self-consciousness

Speaking about self-reference, an implicit assumption was made that there was no Self or consciousness. There were entities like formulae or sets and as a special case of their referential ability they were referring to themselves. Now, when we are beginning to speak about the living subjects, there appears the possibility of confusion whether we simply mean reference to itself or to some, further determined Self. We may distinguish the two meanings of "self-reference":

i) there is some Self, I, subject and reference is made to it. Both the subject itself and some external agent can perform Self-reference in this meaning;

ii) the accent is put not on the Self understood as something active or conscious but on the fact that some x refers to itself – x can be a living subject or a formal object.

Combination of these two gives us the following more exact meanings:

iii) i) excluding ii) – an agent refers to some external Self, a modification which belongs under the headline "being-with-others" rather than to our present subject;

iv) ii) excluding i) – some formal object refers to itself, as the cases from the previous section;

v) i) intersected with ii) – some Self [2] refers to itself, differentiating itself from itself it remains, in this difference, its self-identity. This is the meaning which is closest to our subject.

3.1.2 Non-reduciblity of self-consciousness

Self is very closely related to the original self-reference (self-awarness) – not necessarily as standing beyond or above it, but rather as this very self-awarness or as its side-product (depending on the further specification of the meaning we assign to "Self"). In any case I am going to concentrate on the meaning of the whole expression, not only on its first component. This remark is important to realize that it is not even the meaning v) which will be applied. The meanings iv) and v) share at least one significant feature. They say: there is something (a thing, a Self) which refers to *itself.* In the first part it was always a determinate object $(G(G(x)), \Omega)$ which was capable of reference and self-reference. Other possibility is hardly imaginable in a theory which is in principle a theory about some objects.

But the real self-reference (self-awarness) seems to lie beyond any objective determinations – it lacks a definite substantiality. If we here, in a similar way, try to isolate some object, some Self which should subsequently become self-aware, we run into the trouble known to all empirical and positivistic approaches: problems of expressing consciousness in terms of external relations and of reducing self-consciousness to the consciousness of the second order. One can begin with impressions and their images to admit afterwards "when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade (...). I never can catch myself at any time without a perception, and never can observe anything but perception" [ToHN] and give up at this point. It is then the question about honesty whether he is able to add "but upon a more strict review of (...) personal identity, I find myself in such a labyrinth, that, I must confess, I neither know how to correct my former opinions, nor how to render them consistent. If this be not a good reason for scepticism, 'tis at least a sufficient one for me to entertain a diffidence and modesty in all my decisions." [ToHN]

[2] I am not writing "some sort of Self" as if there were different kinds of selves, but I am writing "some Self" to indicate: "Self, whatever we can mean by that".

Somehow worse is an attitude which, in the face of the impossibility of reduction labels other approaches as meaningless. It can seem that some possibility lies in taking the departure in the achievements of the contemporary logic and linguistics and proceeding along the way of the strictly scientific thinking towards the expected explanation. The reducibility of self-reference to reference can then serve as a basis for an analogous understanding of the relation between self-consciousness and consciousness.

In my opinion, the most appropriate way is to accept the fact as it presents itself and try to find a description, a language which suits best the phenomenon. The attempts which follow the arbitrary program of reducing everything to the more palpable and simple elements explain the whole away instead of explaining it.

3.1.3 Immediate self-awarness

The example of the indirect self-reference (2.2) was illustrated by the conversation between two persons P and R. Using the expression "the person speaking to P" R obviously refers to himself. If there are other people listening to the conversation, they realize that the person referred by this description is R. They play then the same role as the whole formal system plays when recognizing the self-referential character (self-identity) of $G(G(x))$ or Ω – they assert externally the identity of the referring and the referred. Since the only identity in a formal theory is the one proved from its axioms, the indirectness of reference is no obstacle in ascribing the predicate "self-referential".

The person R, on the other hand, would not necessarily refer to himself, if he did not know that "the person speaking to P" is, in fact, he himself. Even if others think that he does refer to himself, it does not need to be so. R may, for instance, have a confused notion of a person and not regard himself as a person but, let's say, as a physical medium through which some mysterious person is speaking. Or he may regard himself as somebody beyond any relations to others and perceive the whole conversation with its actors as a kind of meeting of the particle-like persons in the physical space. Nobody understands conversation in this way and it is easy to agree that saying "the person speaking to P", R usually means himself entirely and totally as a whole subject. It is only a descriptive way of self-reference which can be used in order to emphasize some aspect of myself, of the present situation etc. Here it can be a part of the meaning "The person speaking to you (P) likes to be listen carefully to." This mode of self-reference would not be self-reference at all, if it was not based on the immediate acquaintance which a subject has with himself, that is, on the original self-awarness.

3.1.4 Self-reflection

One remark concerning vocabulary should be made before we go further. I am rather freely exchanging "self-consciousness" and "self-awarness". The reason is that, in what follows, "self-consciousness" will denote the immediate self-awarness, not the reflective, mature consciousness of one's Self, being and goals. Not all people are interested in making their person the theme of reflection. The deep, spiritual life understood as an inquiry into foundation of one's own being and personality may be a very rare event. Even if such an inquiry is the fundamental concern of philosophy, the scope of this paper does not permit for its adequate considerations.

On the other hand, every human nature, reflexive or immediate, spiritually rich or poor, is perfectly aware of itself. We do not need to do anything in order to experience ourselves in our immediate being and activities. As soon as we are, we are self-aware although in our thematic thinking we usually pass by this fact. Phenomenology of consciousness as found in the philosophies of Fichte and Sartre centers upon this observation. The self-reflexive relation to oneself can appear subsequently but self-awarness is its necessary condition. "Self- consciousness" will be used as it is used by Fichte and Sartre – to refer to the immediate self-awarness. If some reference to the reflective mode is made, I will use "self-reflection".

3.1.5 Self-reference and self-reflection

In the formal self-reference two aspects were easily separated:

1. the referential ability of an object which, at the same time, can be referred to by another object and
2. the ability to recognize the identity of the referring and the referred.

The two are also found in the reflective self-consciousness (self-reflection):

1. the thinking subject posits itself as an object for its thinking – it grasps itself, if not entirely in its full concreteness, so at least as a whole. The grasped (reflected) is, in a sense, a part of this whole, one of the possible objects for the grasping (reflecting).
2. On the other hand, this part is also the whole. If the identity of the reflected and the reflecting was not recognized, the difference in 1 would not yield self-reflection.

Thinking about self-consciousness as self-reflection helps to pick up this analogy. Consequently, the effort can be made to explain self-consciousness in terms of self-reference. The only genuine aspect of such an effort is the recognition that the abstract reflection which perceives clearly itself both as the object and as itself is, in its formally conceived structure, very close to the structure of self-reference. But, like the latter lacked the internal self-identity, so for the former the subject-object identity remains a mystery. And this is so because this identity is not to be looked for in the external relation between the thinking subject and its object, but in the immediate self-awarness. Reflexive detachment is only one among the possibilities contained in this primordial self-awarness.

3.1.6 Not language but being

Finally, there is a difference between an utterance of self-consciousness and self-consciousness itself. And the latter depends on the former no more than the former is possible without the latter. If an utterance is to express self-consciousness, this must be present beforehand. Otherwise the result may be at most self-reference.

The original mode of our referring to ourselves is self-awarness. The linguistic expression does not bring anything qualitatively new. When a child begins to use the word "I" it has already long ago known its meaning. Firstly it referred to itself by means of "you" or its proper name – the words which others used when referring to it. The correct apprehension of the language opens us on the new sphere of linguistic experience and the new dimension of being with others but is hardly necessary for our original self-awarness – in the worst case it conceals it.

3.2 Fichte and Sartre

There are many points at which thinking of Fichte and Sartre diverge. As the most general and transparent, one can mention the metaphysical idealism which is the standard label of the former but which can hardly be found in the latter. The difference may be characterized in the following way. Fichte operates with an epistemological notion of thing-in-itself. For Sartre it has primarily an ontological meaning. For Fichte it is a kind of limit of our knowledge. Since reality inaccessible to the consciousness is a contradictory concept, Fichte solves the problem by letting the transcendental subject "posit" the in-itself. It is a negative, inexperienced and yet necessary correlate of the consciousness. Sartre, on the other hand, begins his whole exposition with securing an ontological basis for fully positive transcendence which exists independently from the consciousness. For him it is the consciousness which is the negative element.

Notwithstanding the different ontological status of the in-itself, we will see that the ultimate transcendence is an equally important, even the constitutive moment of self-consciousness according to both Fichte and Sartre. We can extract the phenomenology of consciousness leaving the ontological question for the time being. Self-awarness occupies then the central position in both systems. It appears as an irreducible factor, independent from the concrete determinations of other experienced phenomena. It is characterized as an absolute relation, not to some predetermined being but to itself. There are also further similarities which concern the relation between self-consciousness and consciousness, its finitude and limitations. To indicate the directions in which the parallell interpretations will be developed and to avoid the possible confusions caused by the differences of language, I make a short review of the fundamental notions.

A. The starting point for Fichte is an absolute, self-identical ego. For Sartre it is "a being which is not what it is and which is what it is not". The principle of ego's identity is certainly essential for Fichte's concept of I. This principle seems to be completely missing in the Sartrean "for-itself" which is never itself. Already the use of the words "I" (Fichte) and "for-itself" (Sartre) indicates the difference. "For-itself" is something which never reaches itself, never rests in itself, which never being itself is only *for* itself. "I", on the other hand, in its presence to itself is primary itself, is self-governing and undisturbed by any lack of satisfaction with itself. I think that it is not only a superficial impression but the real difference in the mood and the meaning of both philosophies. Nevertheless, as far as the phenomenological analysis is concerned, the difference reduces to the replacement of the emphasis within the same structure. The constitutive qualities of self-consciousness as described in the one account are paralleled in the other.

The identity "I=I" of Fichte's system can not be understood as a dead, formal identity. The self-difference of Sartrean for-itself is here expressed as ego's presence to itself. This self-presence is the ground of Fichtean thinking. In order that I might be present to itself, it must not only be itself but also know its self-identity. The "I=I" appears as a moment in a self-conscious process of ego's being for itself.

B. For Sartre the for-itself is in an original opposition to in-itself – Fichte introduces non-ego only at the second step of his deduction from the unity of the transcendental ego. But, for the first, the order of the philosophical deduction does not imply any precedence in time. On the contrary, that non-ego follows from ego means that the latter is impossible without the former. The philosophical analysis introduces distinctions of reason which separate the elements of the real totality. For the second, the non-dual aspect of the original self-consciousness is an essential point for Sartre. Self-consciousness is not a juxtaposition of subject and object but an unified state of original indivisibility.

C. Fichte calls his ego "subject" while the same word means for Sartre only one, narrowed and derivative possibility of for-itself. I will use this word for something (self, consciousness) which is confronted with (not necessarily opposed to) objects. It can then cover both Fichte's ego – non-ego relationship and Sartre's fundamental characteristic of consciousness as intentionality. Another difference of language: the "I" is for Sartre only a reflective achievement of for-itself - for Fichte it is the transcendental subject. Sartre's "I" corresponds closest to Fichte's "empirical ego". Both for-itself and Fichte's ego are impersonal and, in the sense which will be specified later, transcendental. Accordingly I will use "ego", "self", "I" to refer to such a subject. "For-itself", "consciousness", "self-consciousness" are the corresponding expressions from Sartre's vocabulary.

D. One can also point out that Sartre's for-itself is contingent – there is no justification for its being. Least of all can it be considered as its own foundation. The vocation of consciousness is too look for an approval of its existence and to be responsible for the being which does not originate from it. Fichte, on the other hand, speaks about the absoluteness of I. Its explanation does not go beyond its being, its presence to itself is the first principle not only of philosophy but of all intelligence. The ego finding itself is its own confirmation.

This difference touches upon the difference of mood referred to in **A**. It opens a perspective for consideration of the moral questions. In the 3rd chapter of "The Vocation of Man", for instance, it leads Fichte towards explication of the meaning of the moral law and the categorical imperative. In the posthumously edited "Notebooks with a View to Morals", Sartre develops it into the exposition of the few indications from "Being and Nothingness" about the possible salvation from "bad faith". Unfortunately, there is neither space nor reason for taking up these things here. I only notice briefly that a possible search for affinities can go in the following direction.

Speaking about contingency, Sartre considers the existential aspect of for-itself. There is no cause, no reason for its existence and no intelligible origin of it. Its appearance in the midst of in-itself is an "ontological act" without least explanation. But lack of any authorization for something can be the ground, in fact good enough, for considering it as an absolute. "A cause of itself" is a good name for such a case and only intuitive uneasiness can prevent us from beginning all other explanations from this, so obtrusive fact. Some of this meaning lies also hidden in the Sartrean "ontological act".

But there is also another aspect of this contingency. It is what Sartre calls "facticity". It is what happens to for-itself in such a way that for-itself is not the foundation of this happening. "The for-itself is in so far as it appears in a condition which it has not chosen." [B&N] (If there is anything corresponding in Fichte's exposition it must be what he calls the "bare empirical", see 3.4.1.E.) Now, consciousness appearing in the above condition, is never identical with it but has always already withdrawn from it into itself. Consciousness stands above its own contingency because the consciousness is for itself, because it is self-conscious. In this sense it is absolute. Absoluteness of consciousness means impossibility of reducing it to its facticity, to the given – the impossibility of approaching it as if from below. In other words, absoluteness of consciousness means that it is not a simple relation of two external elements but a doubled relation – a relation conscious of itself. These words can be easily matched to Fichte but it is actually Sartre who comes with nearly the same formulations. He also describes (self-)consciousness as a transcendental absolute and its being as consisting in a perpetual presence to itself.

The list could be extended both in length and in depth but its aim was only to sketch the main links between the two thinkers. Before I discuss the points of interest in more details I want to observe some aspects of the Fichtean idealism and its common reception. Although it does not belong directly to our subject, it is not only interesting but also, to some degree, relevant. As said before, it is phenomenology of consciousness which concerns us most – not the idealistic dimension of Fichte's philosophy. But even if the two can be judged by the independent standards, they still should be regarded as just two aspects of one system. Discovering some truth in one of them can help in appreciation of the other.

Idealism in general, and the idealism of Fichte in particular, has been often discredited as subjectivistic. The accusation is derived from misinterpreting idealism as if it regarded the particular ego of a person – which is perceived by this very person as limited, sometimes even as powerless – as the absolute source of truth and being. We read, for instance: "that all non-ego is *only* for ego; that this non-ego derives all the determinations of its apriori being only through its connection with the ego (...); that therefore whatever holds as logical truth for any intelligence which finite intelligence can think, is at the same time *real truth*, and that there is no truth other than that." [RoA] [my emphasis] Or: "to think and to determine an object is one and the same", "reason is absolutely independent. It is only for itself. And there is nothing more for it" [ITS], "consciousness is consciousness through and through. It can be limited only by itself." [B&N] If deliberately misunderstood, this can mean: reason (consciousness) lives in a world which is its own creation; its freedom is an unlimited power to create the objects; whatever it thinks, becomes real and is determined and limited only by consciousness itself. Unfortunately, the last quotation is from Sartre.

When it comes to the absoluteness of the reason, other statements should be added. "We can not think up anything absolute, create it through thinking. Only what is immediately intuited can be thought." [ITS] It is not the unlimited power of a finite reason which is the point of (Fichte's) idealism. It is rather the fact that this reason can not be taken as a pure mechanism reducible to the interplay of its determinate elements, to the association of empirical data. The "logical truth of an intelligence" must be taken as the real one because what can be grasped by hand or by a narrow thinking does not exhaust the whole sphere of reality. The truth about ourselves certainly can not be reduced to the reality of things. It is rather the reality of things which is founded upon our way of comporting ourselves toward them, upon our using and projecting them. An object must be thought of as an object for us, for our thinking. The only alternative is naive realism. Even if we are thinking a thing-in-itself it is we who think it. We can consider it as a reality independent from our consciousness but we can not say that we do not think or represent it. It must appear for us before it can get the status of an independent reality in our thinking. Like "the first undefined number" gets its definition in this very expression, so the "things existing independently from us" appear for us in the very moment we mention them – appear as independent. What is worst for the enemies of this idealistic reasoning is that "they can not ask about anything or talk about anything without thinking it" [ITS].

Although speaking about unknowable does not make much sense, it does not follow that any transcendent being is denied by the same token. Fichte was probably the first who reacted against the concept of the thing-in-itself as a totally unknowable entity. But he did not dispensed for this reason with everything which transcends the sphere of subjectivity. The transcendence of things is not known by means of the empirical intuition but according to a higher law of our constitution: "the absolute subject, the ego is not given in an empirical intuition, but is posited through an intellectual one; an absolute object, the non-ego, is what is posited in opposition to it. (...) But absolute subject (...) and absolute object, a thing-in-itself independent of all representation – of these one will never become conscious as something empirically given." [RoA] This indicates also how we should approach the subject of consciousness. It does not appear within the world, as one thing among the others. One must "let the thing appear for the eye of the thinking person." [ITS] The person itself remains different, though not separated, from the world. As such it can not be put into the conceptual network describing the reality of things. While a thing is for-us, we are not for the things but only for ourselves. Things are within the world – we have the world for ourselves. These two aspects will lead the further discussion:

- section 3.3 – being is always for-us which does not mean "created by us" but rather that our being and being of the world are never separated from each other;
- section 3.4 – the ego can not be considered as one of the things – it is an absolute, independent activity irreducible to the entities or their external interrelations.

3.3 The dual (empirical) moment of self-consciousness

3.3.1 Fichte

A. Philosophical reflection and being of the ego

In the "Introduction to the Theory of Science" two points of view should be separated. The one is achieved by means of the philosophical reflection upon the immediate being of the ego. The other is that of the unreflected ego itself. Both are mixed together in a rather unsystematic manner but the main idea seems to be clear enough. The original ego is not thematically aware of its own being. It simply is and acts according to the laws of its being without knowing them explicitly. The philosophical reflection presents the ego to itself but does not affect its original structure – it makes only this structure explicit for the ego. In the course of the exposition Fichte emphasizes the real unity of all the elements which are distinguished only for the sake of the analysis. Philosopher "can make an abstraction, that is with the freedom of thought, he can separate what in the experience is joined together. In the experience, the thing which is to be determined independently from our freedom and according to which our knowledge must direct itself, and the intelligence which is to know are inseparably connected. (...) consciousness about this depends on the abstraction." [ITS]

Since there are two constituents, the non-ego and ego, abstraction can be made in the two directions. "If he [a philosopher] abstracts from the former [thing], he gets an intelligence in itself – he abstracted from its relation to experience. If he abstracts from the latter, he keeps then the thing in itself, that is, he abstracted from the fact that it is given in experience." [ITS] This latter possibility is rejected as ineffective: "you will never explain intelligence unless you think it as the first absolute." [ITS] The former, Fichte claims, can give the account for both sides of reality. "Apriori and aposteriori are, for a perfect idealism, not two different things but are exactly the same. It can be only treated from two different points of view." [ITS] In this struggle between dogmatism and idealism neither party can come with a decisive argument for its own position or with a decisive refutation of its opponent. But the choice is not only a matter of taste. It is also the matter of honesty to the whole of experience – dogmatism gives right only to one part of it.

B. Involvement in the world

Fichte begins his system with the principle "The ego is for itself" and claims to have deduced the whole experience from it. On the plane of the systematic reflection it does not mean that, given this formula, everything follows like the theorems of the classical geometry follow from the axioms of Euclid. The formula expresses the highest point of thinking, the ultimate unity and act (not simply fact) which can not be deduced from the other givens. If we do not begin with it, we will never reach it. Instead of concealing it with the confused formulations in the spirit of dogmatism, we must take it as the first and then show that "it is not possible unless something else happens; and that this "something else" is, in turn, impossible without something third. The process will go on in this way until the ultimate conditions for what has been shown at the beginning are completely exhausted." [ITS] The consequences are the *necessary conditions* for the adequate understanding of the antecedents. From self-consciousness we can conclude consciousness because the former is impossible without a transcendent being: "being of the ego is impossible without an immediate being outside it" [ITS].

Here we are coming to the real dimension of the deduction. The deduction is constructed by the philosopher but it reveals the reality of the non-philosophical I. On the level of the pre-philosophical experience it amounts to the simultaneity of all steps. Conclusion follows because it is a fact of our existence: "only on the ground that by the pure act of self-positing no consciousness appears, will we conclude another act which causes appearance of the non-ego for us", "as I is only for itself, so there appears necessarily a being outside the I" [ITS]. The being of I is self-consciousness and it is intuited (never given as an object) in the intellectual intuition: Equiprimordially, the being outside the consciousness is given to it. Neither is possible without another: "self-consciousness and consciousness about something which is not ourselves are necessarily joined together", "the intellectual intuition is always accompanied by the empirical one" [ITS]. As soon as self-consciousness is there, a being outside it appears to it – both are inseparable. But the point is also that self-consciousness has a superior character over empirical consciousness and, in the last instance, over unconscious being – superior in the sense that the assumption of the dogmatic thing-in-itself (not the Kantian one) can not serve as the ground for explaining self-consciousness. What is worse, it overlooks the activity of the subject which is present already in making such an assumption.

If we direct our attention towards self-consciousness, we notice its irreducibility to the world but also its necessary engagement in the world. Self-consciousness (not confused with self-reference) displays its essential feature as being-in-the-world (which is to be distinguished from thing's being-within-the-world). Speaking about some abstract Self in itself is, in the best case, a philosophical abstraction. And it is not only the thinking of the most recent times which has paid attention to this fact. Reversal of the dogmatic order of thinking implies the "for-us" of being. The mode of beings with which we are confronted is, in a sense, the function of the subject. We construct the objects although we do not create them. Even if we know the features of the things which we have "constructed", it does not mean that nothing remains veiled – for us. "For-us" does not mean that there is nothing unknown or that the world must behave according to our wishes. It means simply that we must see it as it appears to us and that we can not claim the absolutely objective knowledge about each concrete aspect of the world. It seems to express our limitation in a more genuine and intelligible way than the notion of the supposed inaccessible reality in itself.

C. The individual

We can further develop the assertion that the real being of an average consciousness interests Fichte at least as much as the possibility of grasping it in the philosophical reflection. Everybody can think himself, we make pronouncements about ourselves, "I", "me" so often and so easily that, concentrated on the world, we do not notice their particular significance. "I am kept and limited [by the actual object] and when I am thinking it, I forget my own thinking. (When it comes to the usual thinking this is what usually happens.)" [ITS] I forget my thinking, my own self but nevertheless it is this thinking which is the actual subject of my activity. "The pure I is the foundation of all thinking. So far everything goes mechanically." [ITS] Not like in the mechanics of Newton, of course, but simply without our purposive activity – we only are and our subjectivity, our ego is already there along with us. "But to notice this necessity (...), this is not accomplished by a mechanism." [ITS] Dogmatism, taking the point of departure in things, is not able to recognize the presence of this transcendental subject. To do this, one must freely choose it as the object of one's reflection. The act of this reflection is free – I am now deciding to reflect upon myself. But its result is revelation of the necessary truth about the perpetual activity of the ego. The schism dogmatism – idealism occurs only at the level of the philosophical discussion. In daily life everybody "bears his self with himself", is the I structure. "To think this or that depends on my self-consciousness. But I have not made the self in itself." [ITS] This structure, however, remains concealed for most of us – we are not fully conscious of ourselves. Lost in the world, we forget the necessary condition of our being, of having-the-world for ourselves. Common thinking will direct itself toward the common objects – consequently, most people are dogmatics.

The last quotation can be also used to illustrate how little plausible the accusations are which blame Fichte for "having created the world out of his head". "I have not made the self in itself." The "self in itself" refers to the transcendental ego which has little in common with the "I" referring to myself, to that concrete individual. The transcendental ego is the necessary structure of our being. The empirical ego is only a part (a necessary one) of that structure which relates it directly to the world. Within the scope of the empirical ego, there emerges the I as this individual person. It is this one which we usually mean saying "I". This does not give an explanation of the appearance of the non-ego (I finds it as already being there) nor of my thorough-going identity. It is only the multiplicity of the impressions and ideas. The real subject becomes for it something standing beyond that multitude but its nature remains misunderstood as long as we imagine it as one more object added to the perceptions or as a point of reference without any internal relation (for-us) to a different being.

D. Pure activity and the non-ego

The I becomes diversified in its empirical dimension. Once it is the I who hates, at the other time the I who buys a car and so on. In this multiplicity, it remains different also from itself. It is differentiated from itself as one unified subject. If it was not the one it could not be multiplied. Everybody knows this identity and yet, at each moment, he is cut off from it. The transcendental ego persists in spite of our awareness of it as a hidden, inexplicable unity. Our ignorance does not annihilate it. "I have not made the self in itself." It is there and I, the individual, can not help it. Properly speaking it is not me, the concatenation of the manifold experiences, who am conscious of myself. It is the transcendental ego which precedes me. Without it the "me" would only be one thing among others and would never become a self-conscious activity. "One can show in the experience of any person that the intellectual intuition is taking place each time he is conscious. I can not make a step, move my hand or foot without the intellectual intuition of my self-consciousness in these acts. Only by means of this intuition do I know that it is me who do all this (...). Everybody who ascribes himself an activity invokes this intuition. It is the source of life and without it there is only death." [ITS]

The ego is not an object but the source of activity directed towards the objects. Actually, it is not even the source but the activity itself: "intuition of the bare activity which is never fixed but always in a flow, not being but life." [ITS] Intelligence is an activity and nothing else. "One can not even call it the active (...)" [ITS] If it was a source, a fixed entity, we would get an external relation between it and the objects on the one hand, and the dead, numerical identity of being on the other. But as the activity, as "doing" itself, it is different from any being. Simultaneously, it is intimately involved in being, because, as activity, it is transcending being, it is a free working it out and working into it. And it could not be an activity directed toward some *different* being if it was not, at the same time, an activity directed toward itself.

3.3.2 Sartre

A. Philosphical reflecton and being of the ego

Emphasizing the significance of the empirical aspect for Fichte can go against usual opinions about (his) idealism. In the case of Sartre, the corresponding aspect of being-in-the-world can be easily recognized as having crucial importance. As remarked earlier, ego is for Sartre only a secondary phenomenon. It is not constituting but constituted. When I am in the street and want to reach the buss, what happens in the consciousness (not in *my* consciousness) is not "I must reach the buss over there". It is rather "consciousness of the streetcar-having-to-be-overtaken". [ToE] There is no "me" though there is an active consciousness. The "I" can appear for consciousness only like other objects – within the world, like Fichte'e individual. Sartre speaks then about the reflective (positional) and unreflective (non-positional) consciousness. "The-streetcar-having-to-be-overtaken" as an object of the immediate consciousness is the case of the positional consciousness of the car. In "I must reach the car over there" not only the car but also the I is such an object. The consciousness itself does not announce itself nor makes itself into the object. But in its activity (having-a-car-to-be-overtaken) it is immediately aware of itself. It knows itself without being its own object. It is the non-positional consciousness of consciousness.

The reflective and non-reflective do not correspond to Fichte's philosophical and non-philosophical. The positional consciousness of "I" is what Fichte would call the non-philosophical concept of I. The positional consciousness in general, corresponds to the Fichtean empirical consciousness. Both are characterized as having a definite object – material or ideal. The non-positional consciousness-of-consciousness displays the closest similarities to the transcendental ego. Both are immediately aware of themselves and can never be made into objects. Sartre seems to be more consistent in speaking about consciousness-of-consciousness and avoiding the distinction (at least verbal) made by Fichte between the intellectual intuition and the transcendental I. Anyway, being of the non-positional consciousness is constituted and exhausted in its appearance to itself and so does the being of Fichte's ego.

Both reflective and non-reflective consciousness remain at the level of everybody's being – it does not matter whether we know and express them explicitly or not. Philosophical analysis can describe the non-positional consciousness because the latter is always conscious of itself. But it can not make it into the *object* of reflection because it is the non-positional consciousness itself which would be responsible for the whole act of such a reflection. The non-positional transcendental consciousness is not only irreducible to the objective consciousness but is also its very condition. As it finds itself, and as the philosophical reflection finds it, it is always in the world and it can not be thought of otherwise. It is, so to speak, the doubled positional consciousness of the world. The investigation begins with the latter but it immediately finds the former as the condition both of the positional consciousness and of the investigation itself.

B. Involvement in the world

"Consciousness is consciousness of something. That means that there is no consciousness which is not a positing of a transcendent object; consciousness is a positional consciousness of the world." [B&N] It is this consciousness about which we are usually conscious – our daily awareness is directed toward the world. And it is not any negative evaluation. It is the fundamental quality of consciousness, even its definition: "consciousness is defined by intentionality." [B&N] Fichte admitted the principle of intentionality: "all thinking must be directed toward an object." [ITS] Reformulated into "as I is only for itself, so there appears necessarily a being outside the I" [ITS] it served as a step in his deduction. Was it a kind of ontological proof? Sartre did not hesitate: "Consciousness is consciousness of something. This means that transcendence is the constitutive structure of consciousness; that is, that consciousness is born supported by a being which is not itself. This is what we call the ontological proof." "Not only does pure subjectivity, if initially given, fails to transcend itself to posit the objective; a "pure" subjectivity disappears." [B&N]

We think that we can imagine a pure I which is not infected with any being outside itself. It is such an image which accompanies the common-sense refutations of idealism. The intention of Sartre, as well as of Fichte, is that it is only an image, a thought – at best, an abstraction. This version of the ontological proof does not start with an ego, with some "inside", in order to deduce some being outside it. It starts with an ego which can not be thought of without some transcendent being.

The consciousness is always, per definition, engaged in being. Of what king is this being? "There is no being which is not the being of a certain mode of being, none which can not be apprehended through the mode of being which manifests being and veils it at the same time." [B&N] That is: the being as being is not given to our apprehension directly – we are always confronted with some mode, some particular meaning of it. "Consciousness can always pass beyond the existent, not toward its being, but toward the meaning of its being. (...) The meaning of being of the existent in so far as it reveals itself to consciousness is the phenomenon of being." [B&N] The particular modes of being can be grasped by consciousness in their meaning. Needless to say – meaning for-us. As the consciousness is essentially with the being, so the being is essentially for the consciousness. Consciousness which has directed itself toward being has also made this being into being-for-consciousness. Being-in-itself is independent from consciousness in so far as the consciousness is not a being but, as Sartre expresses it, is nothingness. But its independence is constituted for consciousness – it is the phenomenon of being. "Only in so far as it reveals itself to consciousness is it the phenomenon of being." [B&N] Actually, Sartre is very cautious to point out that it is the being of phenomenon and not only the phenomenon of being which gives us the ultimate access to being. A phenomenon is only a phenomenon but being of the phenomenon is already being. In any case, the consciousness is the necessary condition because the being of the phenomenon is exactly the being which, so to say, happens between consciousness and its transcendent object.

C. The individual

Sarte is quite explicit about the fundamental object of his interest. It is "la réalite humaine" – the human reality. The theory of consciousness occupies only a little introductory part of [B&N]. The rest of the book describes different sides of the fundamental project of the men, its contradictory premises and resulting "bad faith".

The human reality is being-in-the-world. Everybody is in the world as consciousness but also as self-consciousness. "Every positional consciousness of an object is at the same time a non-positional consciousness of itself. (...) This self-consciousness we ought to consider not as a new consciousness but as *the only mode of existence* which is possible for a consciousness of something." [B&N] In spite of its omnipresence the non-positional (self-)consciousness remains veiled. Instead there appears its empirical correlate. "There exists an immanent unity (...): the flux of consciousness constituting itself as the unity of itself. And there is a transcendent [not transcendental!] unity: states and actions." [ToE] This transcendent unity is the ego. "It is outside, in the world. It is a being of the world." [ToE] Multiplicity of states, feelings, situations does not clarify the intuited self-identity. On the contrary, the ego is only an "ideal unity", like a horizon of all objects which, in principle, is not different from the world. But both the ego and the world "are two objects for absolute, impersonal consciousness and it is in virtue of this consciousness that they are connected." [ToE] In other words, without the transcendental, absolute self-consciousness the ego and the world fall apart and there remains only the confused "I" confronted with the inaccessible reality of the thing-in-itself of dogmatism.

Sartre is well aware of the possible misunderstandings and he writes: "The world has not created me; the me has not created the world." [ToE] The me has not created the world – and the individual ego of Fichte has not done it either. The ego is a projection of the transcendental consciousness into the world. It is not the activity of ego which unfolds itself in the world. The ego itself is as dead and determined object as any thing. The absolute consciousness, on the other hand, "when purified of the I, no longer has anything of the subject. It is a first condition and an absolute source of existence. And the relation of interdependence [between the me and the world] is established by this absolute consciousness." [ToE] It may serve as a good hint for understanding the Fichtean deduction from the transcendental ego to the opposition between ego and non-ego. The deduction which is based on the phenomenological, rather than logical analysis.

D. Pure activity and the non-ego

The transcendental consciousness ("for-itself" of [B&N]) escapes any kind of determination. As the transcendental ego of Fichte, it is always in-the-world and never falls into this world. It is essentially concerned with being-in-itself (ontological proof) – "the phenomenon of being, like every primary phenomenon is immediately disclosed to consciousness." [B&N] And for this very reason, that consciousness has it disclosed and is concerned with it, the consciousness is what the in-itself is not. In-itself neither has itself nor is concerned with itself – it simply is. On the other hand, as an activity, for-itself always transcends not only in-itself but also itself, its present situation, "it is what it is not" – "not being but life". "The being which is not what it is and which is what it is not" is always self-differentiating and self-differentiated. It is never frost in unambiguous determinations – it is subjectivity without the subject, activity without the active.

We may ask what, if anything at all, corresponds to the Sartrean in-itself in the system of Fichte's. Sartre makes a distinction between the world and the in-itself. The world is a definite object and is opposed to the me (an object as well) but is present to the transcendental consciousness which is the condition of both. In-itself is one step above the world – it is an ontological quality equiprimordial with the for-itself. As argued before, the reality to which consciousness has no possible access is, for Fichte, a contradictory concept. But consciousness exists, according to him, *necessarily as limited.* This limit (which is never given in the empirical intuition like, for instance, the world is, and which is consequently, as Fichte puts it, posited by the subject) is the non-ego. Deduction from self-consciousness to the opposition between ego and non-ego means exactly this necessity. The non-ego must be included into the adequate notion of self-consciousness. It is this being which consciousness is not and which is not consciousness. As the result of the next step of the deduction this "empty" non-ego acquires determinations of definite objects. The world as the totality of concrete (though still diffuse) things "is posited" when the transcendent-as-such has already appeared. Thus we can trace the distinction in-itself – world in these two subsequent steps of the deduction.

To push the analogy still further, one more characteristic common to the Fichtean non-ego and Sartrean in-itself can be mentioned. The first truth about the pure I is that it is for itself. From it Fichte passes to the examination of its relation to the non-ego. The latter appears only as an object upon which consciousness exercises its activity. The non-ego, now understood as being, is a total passivity, devoid of any relation to itself and posited as an opposition to (correlate of) the active ego. Now, Sartre writes: "if being is in itself, this means that it does not refer to itself as self-consciousness does. It is this self. It is itself so completely that the perpetual reflection which constitutes the self is dissolved in identity." [B&N] The in-itself is what it is – a dead identity opposed to the restless for-itself. As such it can not act upon consciousness. It is only the consciousness, reaching beyond itself, which can act upon in-itself. Being(-in-itself) is finite and "dead". It rests in itself indifferent to the world. It is there merely as a "limitation of the free activity". The transcendental ego, the free activity itself limits and determines the passivity of the objects giving to them some meaning.

Not only the division lines in both cases seem to be parallel. The relations between the two spheres separated by these lines and the spheres themselves are also very similar. (We should not, however, forget about the difference mentioned at the very beginning concerning the ontological status of the in-itself.)

3.3.3 Conclusions of the section

1. Consciousness is in the world – not as one thing among others but as *having* the world *for itself.* It is so because consciousness is essentially self-conscious and only as self-conscious is it an activity transcending in one act both being and itself.

2. In general, from the assumption of one being, there does not need to follow that there is also another being. But in the case of the consciousness, it is not only the fact of its existence, but it is also the essential feature without which consciousness would not be consciousness at all. We can abstract from it, we can dissolve the original relation of consciousness to being into its components and hope to get the whole again. But if such an analytical procedure forgets the synthesis which gave it its origin, it will never reach the genuine truth about its object.

3. As the relation to the transcendent being is constitutive for consciousness, so there is also an essential dispersion of consciousness. The consciousness is confronted not only with the world as such but also with the multitude of experiences within the world through which the I is diversified. It is always in some particular state, and the states, actions and feelings pass through it all the time. In this self-multiplicity, there persists the immanent consciousness of the one subject hiding behind the façade. The experiencing I feels itself to be different from the actual one. In the diversity of feelings and perceptions there is no unifying principle. And yet the "non-actual" and the actually experiencing I's are identical. If the latter was not the one being it could not recognize

its different states as its own. The world would be a collection of instantaneous, point-like events without any continuity or – there would be no world at all.

But, on the other hand, this transcendental unity can appear only in the world. The ego "needs" the empirical me if it is to be self-conscious. In the next section we will see why it is reasonable to say that not only consciousness implies being other that itself, but also that beginning with pure self-consciousness one should immediately pass to such a being. Only in the face of the world one can, in a reflective act, return to itself, ask about the principle of one's unity and find it. And not only the reflective, but also the immediate self-consciousness would be impossible without the world being there. The distinction into the aspects of differentiation and unity is here justified by the care for the presentation. In reality these aspects are not only intimately connected but are nothing more than manifestations of one, not total but simple, not unified but indivisible phenomenon.

3.4 The transcendental unity of self-consciousness

3.4.1 Fichte

A. Self-presence as the transcendental spontaneity

Since the transcendental ego makes its appearance in all usual situations one may wonder what its transcendental character means. It has nothing to do with the infinity of the regress "I think, that I think, that I think, that ..." Such a mediated, reflective difference between I-subject and I-object, typical for self-reflection, is only one possible modification of self-consciousness.

"If you – I am sorry for the example but I still find it excellent – were to sew or cut the cloths to a person having these cloths on him and if you, without warning, pricked him, then he would cry out: Auu, that is *me*, you are pricking *me*!" [ITS] I too, find this example excellent (recall the street-car-having-to-be-overtaken). Fichte uses it to illustrate the difference between the I and the individual. And it is the former which is referred to, not the latter. "What will he say with it [..that is me..]? Not that he is this given person and not somebody else – this is perfectly clear to us – but that what we have pricked is not an inanimate garment without feelings but his living and feeling self." [ITS] One can say – well, but it is only the empirical I which is pricked – you can not prick the transcendental ego. But the point is that the conceptual separation of the two does not correspond to the real one. Abstraction is the effect of the philosophical reflection – the empirical can not be separated from the transcendental. If such a separation was possible, the empirical ego would be a thing unable to say "you are pricking *me*". The whole example is maybe an unusual context to put the transcendental ego into, but Fichte ... apologizes.

The transcendental character of the I does not lie simply in the fact that it is a synthetic unity of apperception and, as such, the condition of the possibility of any experience which itself remains beyond any experience. That character is not only epistemological but first of all ontological. Ego is the living source of existence. Discovery of the original being of the I and of our immediate access to this being which is incomparable to any objective being has been expressed by Fichte: "the logical truth of an intelligence is also its real truth". Kantian logical truth of the intelligence was the unity of apperception. Fichte's real truth – the immediate presence of I to itself. Presence which is never posited as an object, which is not perceived nor intuited like a thing and which, consequently, can not be called a real truth of an object but, in contradistinction from it – a logical truth of the intelligence. It was the greatest discovery of Fichte – that this last truth can be shown not only through reasoning as the necessary condition of our knowledge but also in a peculiar mode of intuition, in the intellectual intuition of every conscious being.

It has been said that the I "gets lost" in the everyday life, that in its concern with the world it does not notice itself. But this statement must be taken with care. I does not notice itself in the way it notices the objects within the world. And since the latter is the most usual and natural mode of consciousness, it does not know that it *must* notice itself, that it *does* notice itself all the time. Whenever I try to say something about myself I give only a partial description, some actual determination: I am fat, I am short of money, I am greedy etc. But the I which makes all these declarations is never exhausted in them. That I which is all that and always something else remains hidden beyond such descriptions. It can not be grasped in the limited concepts nor can it be proved from them: "What this activity is can only be intuited, it can never be derived from the concepts or communicated through the concepts." [ITS] It can never be given to us as an object because every object is given to this transcendental I. And yet, it is always present to itself. Everybody is fully aware of being himself. And he is so irrespectively to any concrete situation. He is given to himself in an immediate intuition which does not present to him any limited, well defined being. This immediate presence to itself is called by Fichte the "intellectual intuition".

Inaccessibility of this self-presence makes it into the transcendental phenomenon. It can not be reached by a successive summation of the particular observations although it is always an I with such limited determinations in time and space. "Intellectual intuition can not be proved from the concepts" but it "is always bound together with the empirical one". [ITS] I cannot make the totality of my being into an object for myself but I am for myself – it is the being of the transcendental ego.

And thus, transcendental idealism can not disprove the claims of dogmatism. If one does not freely admit self-consciousness of the transcendental character, then there is no way to explain to him its necessity. He will stay with the notion of one consciousness separated from the other and thought by this other which, in turn, is thought by the third which ... "The pure I lies at the ground of all thinking. (...) But to realize this necessity (...), this is not accomplished by a mechanism." [ITS]

B. From ego to individual

Fichte "protested that he had never intended to say that the creative ego is the individual finite self. "People have generally misunderstood the Theory of Science as attributing to the individual effects which could not be ascribe to it, such as production of the whole material world. (...): it is not the individual but the one immediate spiritual life which is the creator of all phenomena." " [HoP] The question whether there is only one such transcendental ego or not is not very important here. "From the phenomenological point of view talk about "the transcendental ego" no more commits us to saying that there is one and only one such ego than a medical writer's generalizations about "the stomach" commit him to holding that there is one and only one stomach." [HoP] Identity of indiscernibles of the transcendental ego(s) need not imply their numerical identity. But the difference brought in by the empirical do imply differentiation into many individuals.

"The "I think" is one and the same in all consciousness; that is – not determinable by something accidental in consciousness; "I" is determined exclusively by itself and is determined absolutely." [ITS] Being self-conscious is shared by all. But self-consciousness is necessarily engaged in the non-ego, not only as the world of things but also as the presence of other selves. The ego is an individual. "The I and the individuality are two completely different notions (...), the reason alone is in itself, individual is accidental." [ITS] That is: it is not accidental that the transcendental ego is also an individual but there is no necessity in the concrete character of this individual. The individual as such is an empirical phenomenon which can be understood in the terms of Hume. But like the Humean I, it lacks the fundamental quality which grants its self.-identity – it is not for itself.

The questions "How many transcendental egos are there?" seems to originate in forgetting that the transcendental ego is simultaneous with the empirical, and that they can be separated and considered in themselves only by a philosopher. The particular characteristics of the empirical ego are accidental but the unity of the universal and the individual is necessary. Neither can exist without the other.

C. The ego is for itself – the 1st principle

The transcendental ego can not be grasped or described as the empirical ego can be. This lack constitutes its transcendental character. It does not allow conceptualization because in each case it is this very I which makes something into the concept. It is the experience of the ever escaping nature of the self. But this does not mean that one can not meaningfully speak about that I – only that such speaking requires concepts which are not found among the objective beings. The transcendental I, as opposed to such beings, is a "pure activity". Activity is directed toward the entities but what makes it into an activity (and not simply a movement) is the fact that it is aware of itself, that, as an activity exercised on being, it is at the same time directed toward itself. "The I has no other feature except that it returns to itself. (...) The I and the act returning to itself are exactly identical notions." [ITS] The I is not only directed toward itself – it is nothing more than this directness, than self-consciousness of its own acts. There is no being, no entity behind this activity: "intelligence is an activity and nothing else." [ITS] Being appears only in so far as ego directs itself toward something other than itself. It is not created by this activity but comes forth as its limitation.

The lack of any definite substantiality in the self-conscious activity can be made more transparent by referring it to the nature of this activity – it is "an immediate unity of being and appearance". [ITS] Self-consciousness is *only in so far* as it appears to itself. In this sense it does not add anything to being. The being is. It is only for consciousness and the consciousness in the very act of being conscious of something is conscious of itself. It neither is a being nor supplies the being with something. It merely lets the being emerge for it and in any instance of this activity it knows itself exclusively by means of being this activity. It can not simply be, like a tree or a table. As soon as it is, it appears to itself and reveals some being.

The first principle of the Fichtean deduction is then "The ego is for itself", understood not as if there was an I which subsequently or by its very nature acquired the property of being for itself. The self presence, the for-itself exhaust its whole being. Its transcendental character consists in the immediacy with which it accompanies the appearance of being and which is always passed by on the way towards this being. Its absolute character consists in the fact that self-consciousness of this activity can not be reduced to the element of being. Since any being appears for it, it must be thought of as preceding all being and as never caught by it. "You will never explain intelligence unless you think it as the first absolute." [ITS] Because the being does not imply consciousness and because consciousness is self-conscious through and through, it "is absolutely independent. It is only for itself." [ITS] The modes and the changes of the world do not affect this being which is its own appearing. The differences in the actual content of the consciousness are always doubled by the sameness of the consciousness of there being a content at all and, on the other hand, of there being the consciousness. The transcendental I is always one and the same because it is "only" self-consciousness.

D. Self-consciousness as consciousness of transcendence – the 2nd principle

As the pure activity the I is not dual. It is the ultimate unity of oneself with oneself traced back to the pre-personal sphere. Properly speaking, it is even not the unity but an unbroken identity. "Everybody can think himself." expresses the possibility of positing oneself as an object. "Everybody intuits himself" expresses the genuine, perpetual activity which, without being different from itself, knows itself intimately at each moment. We can also say that this self-conscious activity is the non-dual aspect of consciousness. Consciousness, having always a distinct object, divided, on the one hand, into its object and itself and, on the other hand, into the multiplicity of these objects, does not give any ground for self-identity. No identical I follows from the continuous flux of perceptions. Self-consciousness can not be deduced from the content of consciousness. Can it be deduced from the consciousness itself?

To be conscious of something means to be aware of there being an object, to be aware of this object-not-being-myself and, what comes to the same thing, of *myself*-not-being-this-object. To be conscious is to be aware of this difference. When the difference is present, it is I who see both the object and myself. If I did not notice myself the object would not appear to me."Possibility of any consciousness is conditioned by the possibility of the I or the pure self-consciousness." [ITS] If I were not self-conscious I would not be conscious either. Only in so far as a subject is aware of himself, can it be conscious of an object, that is, know it as an object: "all consciousness depends on self-consciousness. (...) One must think the movement back to itself before any other act of consciousness." [ITS]

Self-consciousness is the necessary condition of consciousness. It is already much more than the usual concept of self-consciousness as the consciousness of the higher level, reflection of the second order can offer us. But Fichte claims even more. According to him we can begin with self-consciousness alone and show that the empirical ego, non-ego and their mutual limitation follow from this assumption. Since Fichte insisted on the strict validity of this order of the deduction I prefer to make an attempt to reconstruct that step rather than blame Fichte for insufficient understanding of logic or his own philosophy. This attempt, if successful, would mean that self-consciousness is not only the necessary but also the sufficient condition of consciousness or that self-consciousness is impossible without the non-ego, without some being outside itself. But firstly let us look at one more aspect of the transcendental ego which has some bearing on the relation between consciousness and transcendent being. The deduction from self-consciousness to the non-ego will be taken up along with the review of Sartre's ideas in the next section (3.4.2.D).

E. Mutual limitation – the 3rd principle

So far I have referred only to the two fundamental principles of Fichte: the 1st – "The I is for itself" and the 2nd – "as I is only for itself, so there appears necessarily a being outside the I". The 3rd principle states the mutual limitation of the ego and the non-ego – "equally sure is that I posit myself as limited (...). I am finite." [ITS] It can be difficult to see how the pure presence to itself can imply firstly the non-ego and secondly its own limitation. Reading [ITS], one can get the impression that the pure I comes firstly, then the non-ego appears, and finally the one limits the other and the determined content of what is thus limited comes forth. When it comes to the order of concepts the impression is correct. The consequents are implied by the antecedents, being in this way their necessary conditions and clarifications. But the impression is incorrect if the real time order is concerned. Here the deduction describes only the simultaneousness of the necessary connections. It is the analysis of one single act which is the genre of all our acts.

The 3rd principle says that the ego can not be for itself otherwise than as limited by the non-ego and, on the other hand, that the non-ego does not appear for the ego otherwise than as limited by its activity. The latter means for Fichte determination of the content. "All limitation is only in so far as it is intuited as being, by its very concept, a thoroughly determined limitation and not only a limitation as such. This limitation is absolutely contingent and constitutes the bare empirical in our knowledge." [ITS] A thing is limited in its perceived and conceptual content by the act of our activity – it becomes an object, appears *for us*. But we do not form this thing according to our wish. On the contrary, every act of perception is performed according to the necessary law which says that the object must appear as a transcendent one (as non-ego), that is as a limit, a goal, an obstacle for our activity. The limitation of the non-ego is then equivalent with the limitation of the ego. Considering some particular object, the ego is limited to think this and not that, to see these features and not others. This is the more concrete realization of the principle that the non-ego, in general, lies beyond the ego, that the world, objects transcend the ego. This "finitization" of the transcendental ego is, according to the order of the deduction, the necessary condition of self-consciousness. A mere limitation of the content or the notion of thing do not imply consciousness. We may think of things being there without any consciousness. But when we say "self-consciousness" we must mean and say "consciousness" and then – "the transcendent being". There is no self-consciousness not engaged in the limiting and limited, determining and determined being.

3.4.2 Sartre

A. Self-presence as the transcendental spontaneity

It can be difficult to find the first principle of Sartre's philosophy. He is not interested in looking for such a principle. According to the phenomenological method, he dispenses with a deductive hierarchy of concepts. Instead, the analysis of the totality of appearances as they originally show themselves occupies the central position in the philosophical procedure. Consequently, Sartre does not need to stress all the time the real interdependencies of the described phenomena – their interconnections make up an essential part of the exposition. Activity of the for-itself, its independence from the world, are equiprimordial with its restless transcendence towards the objects, with its intentionality and specific dependence on the in-itself. There is no question about reducing some characteristics to others, all are just aspects of one being. In this way the immediate access of the consciousness to itself becomes not only one important phenomenon among others but the fundamental methodological assumption. The phenomenological description does not begin with the phenomena but with this assumption of self-consciousness.

Self-conscousness is not only the methodological point of departure but the ultimate source of any activity whatsoever. Behind any experience, reflection, discovery of a determined object there is the transcendental sphere which "is the sphere of absolute existence, that is to say, a sphere of pure spontaneities which are never objects and which determine their own existence." [ToE] It never becomes the object because whenever I try to posit it, it is this transcendental field which already has made both me and itself into the object for itself. As soon as I think myself, it is the transcendental consciousness which has withdrawn and made the whole act. It is like the external principle of identity between the me-subject and the me-object. An attempt to objectify it leads to the infinite regress because it is always one level higher, above the subject-object distinction. As this ever-escaping activity it is not limited by anything but itself. It can direct itself to anything and, looking from above, make it into its object. "Consciousness is consciousness through and through. It can be limited only by itself." [B&N] And it is so because the only mode of access to it is through its inner self-awarness. Anything else, remaining beyond the consciousness, does not penetrate into its being – nothing can deprive it of its being for itself.

On the other hand, this conscious presence to itself is nothing more than consciousness of having consciousness of an object. Being confronted with an object, the transcendental being of consciousnss means nothing more than that it is confronted with itself as well. "The reason is absolutely independent. It is only for itself. And it has only itself for itself." [ITS] This absolute self-presence is what constitutes the transcendental character of self-consciousness.

Its relation to itself is exactly opposite to the relation which it has to an object, that is to intentionality. It is, as Sartre says, consciousness (of) consciousness. The parenthesis is to indicate that it is not simply a consciousness *of*, not an object for itself, but one single being whose being is to be its own identity through the self-differentiating presence to itself. Not something which is and can be grasped but something which, being the true subject (not in the Sartrean meaning) of any positional consciousness, transcends what is objectively posited. Being present to itself in this way, it is impenetrable for any other being.

B. From ego to individual

The transcendental consciousness, escaping any finite determination, escapes also the characterization as "I". The I is its object just like in-itself is. I appears as a transcendent (not transcendental) unity of experience. In any personal being there are then the two aspects: the transcendent and the transcendental. The familiar impossibility of determining oneself as this or that, of defining oneself entirely as a waiter, a millionaire, as hateful or weak is related to the latter. The ideal unity of states and qualities transcends any actual self-definition. Nevertheless, this unity is thought of in a positive manner as a collection of the firm, substantial attributes. But the being of consciousness is thoroughly negative – it lacks and negates any form of substantiality. It fulfils itself in the intentions directed toward being and transcending both this being and itself. The being, in principle, can not serve as a description of the transcendental activity. Self-consciousness is a transphenomenal being incapable of being represented.

The positional consciousness of the world has as its object some transcendent being. But the transcendence itself is not its object – it is passed by. This consciousness is immediately (non-positionally) aware of itself and as such it can reflect upon itself. In the act of reflection the whole situation - the consciousness and its transcendent object - becomes the object for the positional consciousness. The active aspect of consciousness, its states and actions in the face of some other being, obtain a thematic representation. The unity of the active side is called "I" – its opposite constitutes the world. But the whole is found in the reflected consciousness. The reflecting one had only posited the ego and the non-ego and at once retired into itself retaining its non-personal, absolute character. "The transcendental field becomes impersonal; or, if you like, "pre-personal", without I" [ToE]. "Pre-personal" indicates something which is not included in "impersonal", namely, the possibility of subsequently becoming personal. This possibility is rather absent from the Sartre's philosophy. But to appreciate this meaning, we can only mention that, according to Fichte, the essence of the reason in its practical aspect is the constant striving to "bring to unity the ego which represents the intelligible [in the empirical intuition] and the self-positing [transcendental] ego" [RoA]. In other words: striving for the personal character of the transcendental being. Further illumination of this point would, however, lead us too far from our current subject.

C. The ego is for itself
The most striking similarity between Fichte ans Sartre is the following:

- the first principle for Fichte is "The I is for itself"
- the ultimate object of Sartre's analysis is "being-for-itself".

The similarity does not consistst only in almost the same expression or in the fact that self-consciousness is being expounded so much. It consists also in this that "for itself" has so high rank in both systems that it actually makes up the division line between the two areas of being: for-itself vs. in-itself (Sartre) and the transcendental ego vs. non-ego (Fichte). Consequently, also this other side of being acquires similar features in both systems. In the following quotation one can substitute "non-ego" for "in-itself" and "ego" for "self-consciousness" without violating the meaning of these words. "If being is in itself, this means that it does not refer to itself as self-consciousenss does. It is this self. It is itself so completely that the perpetual reflection which constitutes the self is dissolved in an identity." [B&N]

The Fichtean distinction between the empirical and the intellectual intuition is present in Sartre as the distinction between the positional consciousness of something and the non-positional consciousness (of) consciousness. This last consciousness "is non-positional, which is to say that it is not for itself its own object." [ToE] Thus, in both cases, on the one side of the division line there is an objective being, a limited in-itself which is itself completely and independently from the other beings. On the other side there is a being which can not be an object or pure identity, which always reaches toward the other being and, in its very independence, can not do without it. The being which is ontologically superior but existentially inferior to the in-itself. The being which has its essence and whole nature in the pure activity. And the concept of this being as an activity implies the concept of a being which is its object and limitation.

The activity as such has nothing substantial – there is no subject with some particular properties which is active. It is the activity alone which acts or, it is the consciousness alone which is conscious. It is conscious of a transcendent object but, at the same time, of itself. "Existence of consciousness is to be conscious of itself. And consciousness is aware of itself in so far as it is consciousness of a transcendent object (...), but consciousness is purely and simply consciousness of being consciousness of that object." [ToE] Not only "in so far as it is conscious of an object is it self-conscious" but also "is self-conscious in so far as it is conscious of a transcendent object". That is, not only consciousness, defined as intentionality, implies a being other than itself, but also self-consciousness does. It is the Sartrean parallel to the second principle of Fichte (cf. 3.4.1.E).

D. Self-consciousness as consciousness of transcendence
Revealing a being to itself, consciousness exists only as the consciousness of being consciousness of that being. The object of consciousness is consciousness and nothing else. "Consciousness has nothing substantial, it is pure "appearance" in the sense that it exists only to the degree to which it appears (...) – it is because of this identity of appearance and existence that it can be considered as absolute." [B&N] This is the aspect of non-duality. Consciousness does not need to posit itself as an object (as it does with in-itself) in order to appear to itself. It is not and can not be a positional consciousness of itself. And still it is self-conscious. It is self-conscious because its being is its appearing to itself and without this self-appearance it would not be at all.

According to Fichte this being for itself implies the non-ego in relation to which the empirical consciousness may emerge. Self-consciousness is for Sartre equivalent to consciousness: "the necessary and *sufficient* condition for a knowing consciousness to be knowledge of its object, is that it be consciousness of itself as being this knowledge." [B&N] It is the necessary condition for the reasons explained in 3.4.1.D. If it did not know itself as consciousness of an object, that is, if it did not know the object as *different* from itself it could not be *consciousness* of that object. "This is a sufficient condition because my being conscious (of) being conscious of that table suffices in fact for me to be conscious of it. That is of course not sufficient to permit me to affirm that this table exists in itself – but rather that it exists for me." [B&N] But it may sound a little confusing. It is not the pure self-consciousness Sartre is speaking about here. It is consciousness (of) being conscious of that table which at once gives consciousness of that table. And this is the crucial point in the whole concept of self-consciousness. Fichte states it by saying that "the intellectual intuition is always accompanied by the empirical one"; Sartre – by calling self-consciousness the consciousness (of) consciousness of an object. The transcendental consciousness does not exist otherwise than as a consciousness (of) consciousness of that table, that number etc., as a doubled empirical consciousness. Hence there is no problem of going out of the vicious circle of the activity directed toward itself or of deducing consciousness of the world from this activity. Consciousness of the world is the only being of self-consciousenss. That "it is indeed this transcendental consciousness which constitutes our empirical consciousness, our consciousness in the world" [ToE] or, to say the same with other words, that "the I is for itself" is the first principle means that the concept of empirical consciousness is included in the concept of self-consciousness. We can now reconstruct the way along which Fichte's thought was unfolding:

1. Even if we can think the thing-in-itself as inaccessible to consciousness
2. so such a concept is totally useless in explaining consciousness.
3. Consequently, consciousness appears for dogmatism as an usual object or does not appear at all – it is either stripped of its unique character or completely neglected.
4. Since dogmatism is unable to give the true account for being of consciousness, its conceptual framework must be abandoned.
5. Already when we look at 1 we see that in order to speak about thing-in-itself in this manner, there must be a subject who is doing that.
6. The fundamental quality of this subject is the possibility of performing such a reflection over its own activity – the possibility grounded in its immediate being for itself.
7. This being for itself, however, is nothing more than being in the face of the non-ego – from its concept the concept of another being follows with necessity because the absolute (irreducible) subject is dependent on (intentionally directed towards) the transcendent object.

Whether this last principle is for Fichte only epistemological or, maybe, also ontological is the problem which I have already announced as surpassing the scope of this essay. Sartre at least, states it explicitly on the ontological ground. Not only the *concept* of the transcendent being follows with necessity from the *concept* of self- consciousness. "Consciousness is a *being* such that in its being, its being is in question in so far as this being *implies a being other than itself.*" [B&N]

Fichte avoids the very popular confusion of consciousness with something subjective, internal, enclosed and opposed to something objective, external, inaccessible. Expounding self-consciousness as the first condition indicates both inadequacy of such a dogmatic concept and impossibility of reaching from it the phenomenon of self-consciousness. Consciousness must mean self-consciousness. Neither can be separated from another. Consciousness implies its object – self-consciousness means only the transcendent character of this object.

E. Mutual limitation

Since self-consciousness is simply consciousness (of) consciousness and the latter, being governed by the principle of intentionality, is always consciousness of some object, we see how the Fichtean deduction is justified by Sartre. Deduction shows only in the concepts the real unity of the phenomena. In reality, for-itself is limited by in-itself in the sense that it is confronted with it as a transcendent being. And only finding itself in the transcendent world can for-itself be for itself. "If we were to seek for unreflected [transcendental] consciousness as analogue of what the ego is for consciousness of the second degree [reflected, positional] , we rather believe it would be necessary to think of the World" [ToE]. The transcendent character of the object means the same as unifying activity of self-consciousness. Each consciousness is a consciousness of something. That this something is transcendent for consciousness means that consciousness is aware of itself. “What can be properly called subjectivity is consciousness (of) consciousness. But this consciousness (of being) consciousness must be qualified in some way, and it can be qualified only as revealing intuition [of being]. Absolute subjectivity can be established only in the face of something revealed; immanence can be defined only within the apprehension of a transcendent.” [B&N] ”Consciousness is defined by intentionality. By intentionality consciousness transcends itself. It unifies itself by escaping from itself. (...) The object is transcendent to the consciousness which grasps it and it is in the object that the unity of the consciousness if found.” [ToE]

As consciousness is nothing more than consciousness of some object, so self-consciousness is nothing more than the consciousness of the transcendence of this object, or the consciousness of the difference between the consciousness and the object. A transcendent object is a necessary condition of self-consciousness. This is the strange way in which the ego finds its unity. The meaning of self-consciousness and self-identity is the same return to itself from the transcendent being. The transcendence of the world delivers me to myself. Only as self-identical can I be conscious of myself. And only as directed towards the transcendent, only as self-differentiated can I recognize this identity of mine.

4 Summary

The constitutive features of self-consciousness, summarised in the paragraphs 4.1-4.3 below, correspond to the respective paragraphs from section 2.4. The aspects of self-reference which we distinguished in 2.4 were:

2.4.1 the sharp distinction between the referring (formulae) and the referred (objects);

2.4.2 the mixing of these two levels so that the formulae could be regarded as usual objects;

2.4.3 the external (in relation to the self-referential structure) proof of identity of both elements from 2.4.1.

4.1

If consciousness is always consciousness of something then we can not analyze it in disconnection from its world. To say that consciousness is in-the-world does not, however, mean that it is within-the-world like a thing but only that it has-the-world-for-itself. Thus the world, too, should be seen only as a world for consciousness. If I say “I am conscious” I add implicitly “there is something I am conscious of”. If I say “the world” then there is already a consciousness for which this world has appeared. In this sense the world and the consciousness are equivalent – they are equiprimordial, simultaneous. But this equivalence does not mean identity. On the contrary, to be conscious of something is to know it as transcendent, as different from the subject. Consciousness is limited precisely in the sense that it is consciousness of a transcendent object.

But there is no problem of being within or outside consciousness because consciousness is not a box limited in space which can contain some things and not contain other. Consciousness is a way of comporting itself towards being – it is *only a relation to being*, or the way in which being is: on the one hand, the way in which consciousness is *with* and *toward* being and, on the other, the way in which being is *for* consciousness. Consciousness which is speaking about other kind of being makes the fundamental mistake. The difference between consciousness and its object does not correspond to the difference between the referring and the referred. The referred could have “collapsed” into the domain of the objects. Here there is no possibility of mixing the two levels. Consciousness which falls into the objective ceases to be consciousness. And no self-consciousness emerges from making the consciousness into the object. This can only result in the infinite regress with the perpetual lack of the subject-object identity.

4.2

The above equivalence is followed by another one. Consciousness and self-consciousness are impossible without each other. Consciousness reveals a transcendent being, that is, consciousness knows this being to be different from itself and only to that extent is it conscious. But this amounts to the same as saying that consciousness *knows itself to be different* from that being. By the very activity directed towards the being it refers to itself – transcending, it unifies itself.

If by consciousness we mean an external relation of two isolated elements touching each other at the boundary, we not only never explain the supposed emergence of self-consciousenss but we loose the genuine meaning of the consciousness itself. If we then take self-consciousness for a state beyond the world or for a purely formal relation abstracted from its proper context, we can not recognize in it any reality because we can not recognize in it self-consciousness at all. *Self-consciousness is simply consciousness of its own being conscious of some object.* It is nothing strange that, when dissociated, it becomes unintelligible. Consciousness is self-conscious only as *confronted* with the world. And it can be confronted with the world only as self-consciousness.

Thus the second equivalence is stronger than the first one. It states not only the simultaneousness of the two but the identity of their being. The two expressions cover only the different aspects of one being: consciousness – the aspect of its dependence, dispersion, self-differentiation and self-consciousness – the aspect of its absoluteness, identity and self-presence. It is the lack of the identity in the former equivalence which gives it back in the latter. Since the world is revealed only as transcendent, the consciousness, knowing its difference from the world, must immediately know itself. There is no need for proving self-identity in the assumed division into the subject and the object because the difference is not the original one. The identity is immediate – not like the axiomatic identity $x = x$ but as a self-aware one. Identity which is its own difference without the identical – activity whose oneness consists in the enduring being beyond itself.

4.3

The two equivalences give the third one: between self-consciousness and the world. Again, only the lack of the identity makes the equivalence possible. Only differentiating itself from the transcendent, only knowing itself not-to-be-the-world can consciousness be for itself. It does not rest in itself satisfied with the undisturbed immanence. Such an identity, lacking any resistance "from outside" is not able to recognize itself. To do so, the "strangeness" of something different is necessary. To be self-conscious means to find oneself in the face of something else, to affirm oneself in the opposition to the non-self. Self-consciousness must be "open" in this way for its own limitation, for its "outside", because it is this limit which encircles the totality of its being and delivers it to itself. And, properly speaking, it is not to be open but it is this very openness. It is a being which, not containing the non-ego, still, in a sense, comprehends it within itself. Only seeing itself from this "outside", "within non-ego" does it recognize itself. Otherwise it would only *be identical* with itself but could not *be its own identity.*

4.4

The above chain of equivalences is then the final result of the analysis of the concept of self-consciousness as well as of the Fichtean deduction. The equivalence means three things:

1. the simultaneousness of the antecedents and the consequents
2. validity of inferring the consequents from the antecedents and
3. necessity of looking at the consequent exclusively as the conclusion following from the antecedent.

3 means that the deduction from self-consciousness to the consciousness and to the transcendent being can not be reversed. I have explained the grounds for which speaking about the reality beyond the reach of consciousness is philosophically ineffective. But when it comes to the relation

between consciousness and self-consciousness, the impression can be left that they are in fact identical (cf. 4.2). It has been said that self-consciousness is both necessary and sufficient condition for consciousness. If we say that they are but two aspects of one being, then it should be justified to pass from one to the other in both directions. And it is correct in so far as the proper concept of consciousness is maintained. But if we focus on the more traditional notion which exhibits consciousness simply as an opposition between subject and object (I think there is at least some trace of such a notion in Fichte's "consciousness" and in Sartre's ego-world relationship) then we must admit that this opposition can appear only as posited by the higher being of self-consciousness. Self-consciousness in its immediate being conscious of the transcendent is not dual. But it bears with itself the possibility (necessity) of the separation. When it has taken place, there is no way of restating the original unity of the two posited poles. Subject separated from the object, ego opposed to the non-ego do not explain it. They only can not do without it. To realize the creative activity of the transcendental ego, one can not simply add one to one. The result will always be two. One can not try to catch oneself through the reflection which only carries further the gap between both. One needs an entirely new kind of intuition because self-consciousness is not a special case of consciousness. It does not emerge from the latter by the act in which consciousness recognizes its identity with one of its objects. Self-consciousness is the first condition of the distinction between subject and object, of the empirical consciousness. And it is so because self-consciousness is, by its very nature, confronted with the transcendent being. Thus the deduction should not be reversed at this step either. From the two external elements no genuine unity can be deduced. But from the concept of self-consciousness, the concept of the transcendent being follows immediately because self-consciousness is nothing more than the relation to such a transcendent being.

References

[ToE] J. P. Sartre, *The Transcendence of Ego*, Farror, Strauss and Giroux, New York.

[B&N] J. P. Sartre, *Being and Nothingness*, Washington Square Press, New York, (1956).

[ToHN] D. Hume, *Treatise of Human Nature*, Oxford, (1978).

[HoP] F.Coppelstone, *A History of Philosophy*, vol. VII, Image Books, New York, (1985).

[RoA] J. G. Fichte, *Review of Aenesidemus*, in "Between Kant and Hegel", Harris, di Giovani, eds., State University of New York Press, (1985).

[ITS] J. G.Fichte, *Introduction to the Theory of Science*, (a manuscript of Norwegian translation by E. Storheim)