Knowledge Argument versus Bundle Theory according to Derek Parfit

TOMASZ HUZAREK
Uniwersytet Mikołaja Kopernika, Toruń
tomhuzar@umk.pl

Summary. According to constitutive reductionism of Derek Parfit, a subject/person is not a separate existing being but his existence consists in the existence of a brain and body, performance of actions, thinking and occurrence of other physical and mental events. The identity of the subject in time comes down only to “Relation R” – mental consistency and/or connectedness – elicited by appropriate reasons. In the following article, I will try, relying on Frank Johnson’s Knowledge Argument, to argue in favour of the following conclusions: (1) a person/subject is a “fact” irreducible to body and physical relations with the environment and (2) a subject is something/“fact” non-reducible to mental occurrences.

Keywords: subject; person; identity; qualia; consciousness; self-consciousness.

Epistemological turn in European philosophy started by Descartes – although we could distinguish more such turns in the history of human thought – starts a process in which the study of the world’s reality is substituted with the study of human experience of this reality. In further consequence, these studies focus on human subjectivity, consciousness or mind. In this way, it seems that the study of consciousness and subjectivity takes a central place.
in philosophical investigation. Evidently, it does not happen in separation from rapidly expanding empirical studies which pose further challenges to philosophical reflection on subjectivity.

Self-consciousness is a term that expresses a self-concept of consciousness. Self-consciousness in its primary meaning is an integral feature of consciousness in a sense that there is no self-consciousness without consciousness and vice versa. In a more narrow sense, self-consciousness is a feature of a higher form of consciousness in which the sense of “I” is emphasized (Ziemińska 2012, 395–398). The question of subjectivity, consciousness, self-consciousness and self-cognition are important philosophical problems and they are widely discussed. At the same time, the experience of self-consciousness is something so obvious to everyone that it becomes a subject of consideration only when there are difficulties in its normal functioning and when it causes some kind of tension. On the one hand, the universality of this phenomenon generates a belief that studying and defining of what consciousness and self-consciousness are and how they affect human identity should not cause greater problems. On the other hand, this issue has been engaging philosophers for centuries and the development of knowledge is, as it seems, insignificant and undoubtedly unsatisfying, abounding in many divergent positions.

The positions concerning the possibilities of consciousness cognition are divided into three groups. Mysterianistic positions are supported by some thinkers who claim that consciousness is a “secret/mystery” that is not the subject of rational studies. In contrast, those who support transnaturalistic positions claim that consciousness is such a specific subject that the explanation of this phenomenon is out of reach for today’s competence and knowledge and that it requires the creation of a new field of science. And finally, there are naturalistic positions, which dominate today, and strive to explain consciousness with the help of science (Klawiter 2012, 353–356).

The question of consciousness and self-consciousness is associated with the issue of personal identity in time. In *Critique of Pure Reason* Kant uses such an analogy: let us imagine a springy ball that hits an identical ball in a straight line therefore giving it all its movement and all its state.
By analogy, we can comprehend self and subjectivity as a substance from which one would “pour” into another its concepts/images together with their consciousness. In this way, we can imagine the whole range of such substances: the first one would pass the other its own state and the state of the previous one and the third one would similarly pass the state to the next one with the states of the two previous ones. The last substance in a row would be the consciousness of all states and its own state because all the previous ones together with their consciousness would be passed into this one. That substance, however, would not be the same person in all these states (Kant 2001, 331).

Derek Parfit, a British philosopher, calls his strategy of dealing with the “troublesome issue” of person’s identity in time constitutive reductionism. According to this theory – although a person varies from his body and from series of thoughts and experiences – however, “his existence consists only of these elements, it is constituted by them.” (Parfit 1988, 440). In Parfit’s opinion personal identity should not play any role in practical reasoning. For he assumes that criteria of personal identity in time are unreachable and redundant (Scruton 2009, 384) and continuity of person’s existence directly depends on continuity of his mental/psychological life of which brain is the foundation. Parfit reasons in favor of the following conclusions: “(1) We are not separately existing entities, apart from our brains and bodies, and various interrelated physical and mental events. Our existence just involves the existence of our brains and bodies, and the doing of our deeds, and the thinking of our thoughts, and the occurrence of certain other physical and mental events. Our identity over time just involves (a) Relation R – psychological connectedness and/or psychological continuity – with the right kind of cause, provided (b) that this relation does not take a ‘branching’ form, holding between one person and two different future people” (Parfit 1987, 216).1

1 Moreover, Parfit (1987, 259–260) claims: “(2) It is not true that our identity is always determinate. I can always ask. ‘Am I about to die?’ But it is not true that, in every case, this question must have an answer, which must be either Yes or No. In some cases this would be an empty question; (3) There are two unities to be explained: the unity of conscious-
In Parfit’s opinion there is no answer to the question of personal identity in time, that is why, this question should be replaced by a question about Relation R status. The continuity of a person’s existence depends directly on his mental/psychological life continuity of which brain is the foundation. This continuity does not have to include the person’s identity: in many cases the answer to the question whether X in \( t_1 \) time is identical to Y in \( t_2 \) time is not possible because psychological connectedness and/or psychological continuity between the two people is only partial or because that answer to the question about identical nature will be the case of authoritarian decision (Huzarek 2016, 108).

Fundamentally important, according to Parfit, is Relation R: reasoned psychological connectedness and/or psychological continuity. If the personal identity is based exclusively on psychological continuity, the person/subject is not an independent or separately existing “fact” regardless of body and physical relations with the environment as well as a person/subject is not “something” existing regardless of mental occurrences.

Using the implications of the so-called knowledge argument I would like to reason in favor of the following conclusions: (1) a person/subject is a “fact” irreducible to body and physical occurrences with the environment and (2) a subject is something/“fact” non-reducible to mental occurrences.

**The Knowledge Argument against Constitutive Reductionism**

The knowledge argument is a thought experiment to argue against physicalism and functionalism. It is based on the story of an imagined brilliant

...
neuro-physiologist Mary who spends her entire life in a room where she can only see black and white and her only connection with the outside world is a black and white monitor. Despite the fact that she becomes the best neuro-physiologist in the world and knows everything there is to know about physical structure of a brain and its functionality, has all the information about what colours are and, what follows, she has all physical information about what happens in retina when we see colours as well as the rules of how the light refracts. One day Mary is released from her prison and having a sensory experience of red – let us say when seeing a red strawberry – she says: “So this is a red colour” or “This is what experiencing/seeing red is.” Frank Jackson, who was the first to phrase this thought experiment, deduces: Despite her brilliant knowledge as a neuro-physiologist Mary did not know what it is like to have a sensory experience of red. Hence even complete knowledge about physical facts, that relate to sensory perception and accurate brain activities still misses something vitally important (Jackson 1982, 127–136).

The extended version of knowledge argument is as follows (Nida-Rumeli 2010):

Premise P 1: Mary has all the knowledge about human mechanism of seeing colours.

Consequence C 1: Mary knows all the physical facts about seeing colours before her release from her black and white room.

Premise P 2: There exists some kind of knowledge about seeing colours that Mary does not have.

Consequence C 2: There are certain facts about seeing colours which Mary did not have before her release.

Consequence C 3: There are non-physical facts about seeing colours. Non-physical facts about seeing colours (or other sensory experiences which are accompanied by characteristic features of mental experience) are called qualia.

There are numerous studies on the question of qualia as well as abundant philosophical dispute on their nature and properties, etc. The aim of this article is not to concentrate on this issue. My intention is to adopt minimal
assumptions: provided that there exist nonphysical facts concerning the perception of colours, we have to assume that there exists the subject that is different from those nonphysical facts.

**Conclusion 1** – a person/subject is a “fact” which cannot be reduced to body and physical occurrences with the environment

Consciousness appearing during sensory experiences and derivatives regarding these experiences, reminders or sensations based on these images is the so-called phenomenal consciousness.\(^2\) Klawiter describes this kind of consciousness in such a manner.

I look at the red strawberry. I recognize that it is red and identification of this feature is accompanied by a conscious experience of red. This recognition of red (detection) and conscious experience of red are two different processes. The first one is called *senseption*. A man is equipped with a “biological device” to recognize colours – in other words, using a technical language – a subsystem of sensory optic system.\(^3\) The recognition of stimulus feature

---

\(^2\) Klawiter (2012, 356–372) distinguishes phenomenal consciousness from constitutive consciousness. Phenomenal consciousness is a “reaction” to the direct brain answer on what it registers from the surrounding (it is not a direct reaction to the state of the world but a direct reaction (known as experience) to the state of mind caused by non-objective reality). Constitutive consciousness is a complicated creation of the mind created from components that, among others, are effects of its previous activity. Chalmers (2010, 64–73) distinguished phenomenal consciousness and psychological consciousness. Psychological consciousness and consciousness are functional terms. According to Chalmers, consciousness is a state in which a subject has access to some information that it can use to control behaviour. Conscious mental actions include, among others, being in reality (an ability to cope with surroundings), ability to relate, self-consciousness, attention, gained knowledge. Ziemińska (2012, 393–426) distinguishes introspective consciousness (act/reflexive consciousness) apart from accompanying consciousness (the latter one is called *great-consciousness* by Husserl, Ingarden calls it an *experience intuition* and Judycki uses a term *simple consciousness*). Lieberman (et al. 2009, 201–251), however, relying on a naturalized concept of qualia presented by Ranamachandran and Hirstein, creates a theory of bi-process model – instinctive system-X and reflexive system-C. In a way, these distinctions overlap.

\(^3\) Starting from retinal cones, in other words, photosensitive receptors of the retina which absorb light in different range of waves length, to V4 field in visual cortex that analyzes colors and contrasts and V5 field which is responsible for general movement perception and its direction.
as a red colour is the effect of the *senseption* process, namely the reception and detection of physical stimulus features which reach the organism’s receptors and consequently they reach certain neural changes in the brain. *Senseption* differs from perception because *senseption’s* task is to identify a stimulus feature, whereas perception’s task is to gather information about *features of the object* which is the source of the stimulus.

Alongside *senseption*, another process takes place, namely a conscious sense that takes place when “the biological colour detector” is in a certain state. A conscious experience of red coexists with the final stage of *senseption* process. It is a specific mental state which relies on the feeling that “I am red” (this phrase is an analogue to “I am sick” or “I am cold”). This state does not refer to the object’s surface but to the way it is seen, to the process of seeing. Therefore by saying “I am red” (“I feel red”) I am saying nothing about the state of the light, I am saying something about a certain feature of experiencing a colour that I am currently fulfilling. The attempt to describe such a state using language turns out to be impossible: a conscious experience of red appears, which cannot be described in a way so that it can be apprehended on what “*it is like to feel red*”.

Therefore it can be said that a conscious experience accompanies – as it seems – a well-known and able to be described in psycho-physical language sensory process. So a conscious experience of red is not independent from this physical base. The experience itself though – because of its specific and subjective character – has no physical character and eludes any physical description. The fundamental problem is not whether what is physical and what is conscious affect each other. The problem in incomprehensibility of

---

4 Physical state coexists with conscious state. In a way, the study of consciousness is to determine a correlation between physical and mental states.

5 Francis Crick and Christof Koch (2008, 302–323) claim that with a current state of knowledge we are not able to define consciousness. It is, however, possible to find a neural consciousness correlate, which will be defined as a “minimal set of neural occurrences that cause a specific aspect of aware percept”. Vilajanur S. Ramachandran and William Hirstein (2008, 324–363) propose a theory that “neural location of qualia is the set of neurons in lower, primary sensory area because it is them that send connections to frontal lobes where so many of the so-called higher functions happen.” They also claim that most of activity connected with conscious experience takes place in frontal lobes which are an actual interface between perception and action.
this relation, the way what is unaware and is an object of observation can be identical or remains in correlation with something I do not know from observation but I simply feel. The existence of a conscious experience rests on being-a-close-known (Frank 2008, 132–140).

Colin McGinn says about “cognitive closure” that it is based on a rule that because of some primary reasons I cannot determine the element that connects neural and mental processes. “A type of mind $M$ is cognitively closed with respect to a property $P$ (or theory $T$) if and only if the concept-forming procedures at $M$’s disposal cannot extend to a grasp of $P$ (or an understanding of $T$)” (McGinn 1989, 350). According to McGinn, we learn nothing about brain from the consciousness fact. On the other hand, brain studies do not allow us to determine what feature is responsible for emerging of consciousness and introduction of a new term to explain another term must require the criterion of homogeneousness to be obeyed. However, the term of what is conscious and the term of what is physical does not fulfill the criterion.

Without settling the dispute about correlation of physical and mental changes it can be said that quality and therefore individuality and uniqueness of conscious experiences are in a way depended on the quality of sensory process, unique sensual sensitivity and also previous experiences in life history (Koch 2008, 251).  

There is no way to tell, as Parfit does, that “our existence just involves the existence of our brains and bodies, and the doing of our deeds, and the thinking of our thoughts, and the occurrence of certain other physical and mental events.” (Parfit 1984, 216). The qualities of qualia are not independent from sensory body sensitivity but they are not limited to it. The existence of non-physical facts about seeing colors as a phenomenal quality of experiences and their unique subjectivity indicates a “fact” that

---

6 It seems that here McGinn assumes a hidden layer of awareness, we do not have access to, which is a certain indirect state between what is physical and what is mental.

7 Thomas (STh I-II, q 63, a 1) says: “Because of body adjustment some are better or worse prepared to certain virtues along with when actions of certain sensual strengths are partially connected to certain body parts which adjustments help or bother those strengths and rational strengths which use sensual strengths to perform proper actions”.

---
a unique subject, which cannot be reduced to body and physical occurrences with the environment.

**Conclusion 2** – subject is “something/someone” non-reducible to mental states

If nonphysical facts exist, then there must be a qualia’s subject, a subject with certain properties.

Consciousness has a time stretch and it occurs in time: the past, present and future. The future is an object of plans, expectations which do not “really” exist yet. The present exists but – if it can be expressed in this way – to a limited extent because it fades into the past very quickly. The most important dimension of time stretch is the past. The existence of the present, expecting the future and experiencing, living the past are incidents that occur for someone or “something” who can make temporary dimensions of stream of consciousness a subject of their reflection. Such a subject defines future plans, has self-consciousness in the moment of planning the future and can reflect on the experience of itself in the past. *Intentional referring to different temporary dimensions of stream of consciousness requires a subject which has certain conditions, some things it credits, some denounces, sometimes it becomes aware of something, sometimes it phrases some points, etc.* (Judycki 2004, 310).

Self-consciousness cannot be a simple cognition of consciousness according to an object-subject model. For in this model it is assumed that conscious states of mind differ from unconscious by the presence of a certain higher level of state which makes them a cognition object. This, however,

---

8 We may somehow broaden the argument from knowledge. Let us image Mary who has been blind since childhood and has never seen any colors, the beauty of the world or subtle shades of colors. Despite that she becomes a world-famous neurophysiologist and knows everything about the physical structure of the brain, possesses all the physical information about what happens in the retina when a man sees colors. Mary can not tell what the experiencing of red or blue is. One day someone gives her a hot stone to hold in the right hand saying “this is what red is like” and a cold stone to hold in the left hand saying: “this is a blue color.” The process of senseption described by Klawiter does not take place, however, may we say that a specific experiencing of red/blue is formed by Mary?
must lead to regresum ad infinitum. That is why, as it seems, the only solution is to acknowledge primary self-consciousness, to assume that consciousness assumes self-consciousness (Frank 2002, 119–120).

As a person/subject, I have a predicative knowledge about myself: I know how old I am, what my height is, I know I like red wine, I know what features, properties of character I possess. This attribution of any properties to oneself assumes the knowledge of to whom I attribute them. The predicative knowledge assumes different kind of knowledge about oneself: the knowledge that is in a way a direct, outside predicative.

Beyond any objectifying reflection a substantial subject has direct knowledge about itself. According to Judycki, intellectual clarity (different from sensual clarity) belongs to the direct experience of oneself as a subject. It is an experience of oneself as a being existing in time and centre from which different acts originate. Intellectual and clear perception of ourselves is expressed with the word "I". We are not subjects of this knowledge but the knowledge itself. It is an absolutely unique knowledge because it is a unity of idea and clarity of its own kind. This knowledge can only be fulfilled but it cannot be expressed conceptually. The substantial subject as a reflexive transcendence⁹ is both act and activity. Aquinas says that its action constitutes operatio per se. In cognition of oneself an experience of total undifferentiation between the subject (being) and cognition of this subject (being) is fulfilled. Simple consciousness is not based on the subject-object relation but can take place on one level. Cognition and its

⁹ It means that a subject comes "outside itself" in order to accept some subjective content (this content includes: sensual clarity, intellectual – abstract contents and contents of inner experience) and simultaneously a subject expresses this content, "comes back into itself" to assimilate expressed contents whilst keeping its unity and simplicity. This subject identifies with none of the components of its stream of consciousness, none of its profound values; it has its own experiences but it does not identify itself with this experience, it is not this experience; a subject is not an additional stream of consciousness that stretches in time. According to Judycki (2004, 316–327), a subject is characterized by the following properties: non-temporality (it is not timeless – it experiences time but does not identify itself with any features of its character nor with the contents of its memories) and non-spatiality; simple unity and unique simplicity; unidentical with any of the results stretched in time.
subject are blended together due to the lack of mediation through notions (Judycki 2004, 227–239).

Therefore, self-cognition assumes previous self-consciousness. This self-consciousness has, according to Manfred Frank, the following features: (1) it is not reducible to physicalism or naturalism;10 (2) as a de se attitude is not reducible to de dicto and de re"11; (3) self-consciousness is a kind of knowledge but a non-propositional (pre-propositional) knowledge, it is not a subjective knowledge, being, however, a real cognition. For it is a subjective knowledge that eludes in a significant way objective knowledge about the world (Frank 2008, 123. 353).12

It can be then said that a subject is an atemporal unity of existing (Judycki 2004, 317–320). In a way it exists out of time because the time and its changes do not touch the existence of the subject. Simultaneously, however, the subject is not “timeless” because it experiences time. The former “I”

10 The acceptance of non-reducible “subjective factor” must constrict a monopoly of exact sciences in explaining the world. But according to Frank (2008, 109–111), this step should be taken because of few reasons: 1. If we resign from treating ourselves as non-reducible subjects then we will not be able to study philosophy anymore. It will happen because what we mean when we address ourselves we could not separate from philosophical objects and explaining the phenomenon of human reality, its inner world, every kind of experience must remain only in exact sciences expertise; 2. Second intuition invokes a way of expression which is characteristic for philosophy, namely an argumentation. The argumentation differs fundamentally from science language as well as from everyday language of discourse. Arguments lead to certain beliefs using different way than a coercion supported by some kind of obviousness or a logical or causal coercion; 3. Serious ethical consequences, namely categorical imperative (Act in such a way that you treat humanity, whether in your own person or in the person of any other, never merely as a means to an end, but always at the same time as an end,) loses its addressee. Indeed – there is no such a thing as non-reducible subject, the implementation of this imperative lacks fundamentum in re.

11 Frank (2008, 121–124) explains that epistemic attitude towards oneself – the so-called “de se attitude” cannot be reduced to the de dicto attitude or the de re attitude. The de dicto attitude is an attitude towards certain dictum on account of what was said, on account of judgement content (proposition). The de re attitude is an attitude towards an object. In case of self-knowledge we deal neither with content-related knowledge nor with objective knowledge.

12 Frank adds: if a language is a whole of what can be made understandable, in relation to the content of opinion (propositionem), then self-consciousness is not a linguistic occurrence. Some kinds of our knowledge cannot be expressed in indicative opinions, but they are still knowledge.
is former only seemingly because it is the same “I” that lasts in a different time. In an introspective experience, sustained in different changes of our experiences, we can talk about some kind of the “core of existence.” The existence experienced by a subject is a foundation to experience time in general.

Parfit argues that “I” in T₁ time (in the past) does not have to be identical with “I” in T₂ time (in the present). However, this seems to be a wrong conception. The error has its source either in (1) confusion, mixing a subject with characteristic and habitual features which throughout the life could alter completely or in (2) mixing a subject with its own reminders.

However, a substantial subject is not the same as its features of character, possessions or memories. For these properties and memories can be objectified by a subject, can be a subject of reflection and therefore they do not equate with “I”. A subject is therefore “something” that does not identify with the result of states in time regardless of whether it will be a neural or physical state of brain or whether it will be a state of stream of consciousness or qualia as a phenomenal quality of experience. A subject is able to cross a material content of stream of consciousness, it is therefore “something” which cannot be reduced to metal states. It is characterized by an ability to analyze, synthesize and understand all kinds of data, structures, rules and mechanisms. Even if there is an unconscious dimension of our mind, namely a system of rules anchored in a brain to which a subject has no direct access – still then the subject is able to make effects of unconscious data processing a subject of reflection.

That way a subject does not identify with its stream of consciousness and does not identify with its brain. If through mind we are to understand a structure organized according to rules in some way anchored in a brain then subject is not a mind.

A substantial subject is a specific “crystal of unique structure” (Judycki 2004, 327) which focalizes data provided by senses and by a mind/brain that it possesses. It is a “crystal of unique structure” and because of that it receives data (both on a sensory level, and the level of sensory sensitivity as well as on the level of processing data by mind/brain ) in a characteristic way.
References

