Article

Metacomputable

Piotr Bołtuć

Department of Philosophy, University of Illinois, Springfield, IL 62703, USA;
epetebolt@gmail.com; Tel.: +1-217-206-7422

Abstract: The paper introduces the notion of "metacomputable" processes as those which are the product of computable processes. This notion seems interesting in the instance when metacomputable processes may not be computable themselves, but are produced by computable ones. The notion of computability used here relies on Turing computability. When we talk about something being non-computable, this can be viewed as computation that incorporates Turing's *oracle*, maybe a true randomizer (perhaps a quantum one). The notions of "processes" is used broadly, so that it also covers "objects" under the functional description; for the sake of this paper an object is seen as computable if processes that fully describe relevant aspects of its functioning are computable. The paper also introduces a distinction between phenomenal content and the epistemic subject which holds that content. The distinction provides an application of the notion of the metacomputable. In accordance with the functional definition of computable objects, sketched out above, it is possible to think of objects, such as brains, as being computable. If we take the functionality of brains relevant for consideration to be their supposed ability to generate first-person consciousness, and if they were computable in this regard, it would mean that brains, as generators of consciousness, could be described, straightforwardly, by Turing-computable mathematical functions. If there were other, maybe artificial, generators of first-person consciousness, then we could hope to design those as Turing-computable machines as well. However, thinking of such *generators of consciousness* as computable does not preclude the stream of consciousness being non-computable. This is the main point of this article—computable processes, including functionally described machines, may be able to generate incomputable products. Those processes, while not computable, are metacomputable—by regulative definition introduced in this article. Another example of a metacomputable process that is not also computable would be a true randomizer, if we were able to build one. Presumably, it would be built according to a computable design, e.g., by a machine designed using AutoCAD, that could be programmed into an industrial robot. Yet, its product—a perfect randomizer—would be incomputable. The last point I need to make belongs to ontology in the theory of computability. The claim that computable objects, or processes, may produce incomputable ones does not commit us to what I call *computational monism*—the idea that non-computable processes may, strictly speaking, be transformed into computable ones. Metacomputable objects, or processes, may originate from computable systems (systems will be understood here as complex, structured objects or processes) that have non-computable admixtures. Such processes are computable as long as those non-computable admixtures are latent, or otherwise irrelevant for a given functionality, and they are non-computable if the admixtures become active and relevant. Ontology, in which computational processes, or objects, can produce non-computable processes, or objects, iff the former ones have non-computable components, may be termed *computational dualism*. Such objects or processes may be computable despite containing non-computable elements, in particular if there is an on and off switch of those non-computable processes, and it is off. One kind of such a switch is provided, in biology, by latent genes that become active only in specific environmental situations, or at a given age. Both ontologies, informational dualism and informational monism, are compatible with some non-computable processes being metacomputable.

Goals

The paper has four main goals. Firstly, we define the notion of metacomputable. A process or a functionally defined object (from now on we just say "process"), is metacomputable if it originates from another computable process; yet, itself, such a process may or may not be computable.

A process is meta-metacomputable if the process from which it originates comes from a computable process (this is similar to the relation to one's grandparents). For example, a certain process, may come from a non-computable systems (e.g., a process that includes a Turing's oracle) but that second process comes from a computable process of higher order (A non-computable process may function in such system that, while being non-computable, its outcomes are not completely random). The notion of the meta-metacomputable seems more controversial than the metacomputable, since it touches on the issue whether Turing-non-computable processes can produce anything (I presume, reliably enough); this is not an issue I take up here. The main argument of this paper does not depend on the notion of the meta-metacomputable; the main claim is that things, or processes, that are non-Turing-computable may be metacomputable.

Secondly, we develop an example of metacomputable processes that seems interesting in its own terms. This is one way to show that the issue of metacomputable processes is also somewhat interesting, outside of computability theory. We show that the stream of consciousness may be meta (or perhaps, meta-meta) computable. Here is the argument:

A. If the consciousness-stream is metacomputable, it may be incomputable, but projectors of consciousness may still be computable. In particular, those projectors must be *effectively computable in operation* (to use a broad analogy, they may be like projectors of holograms instantiated in brain matter or, maybe, in something else).

B. If consciousness-stream is meta-metacomputable then it can be non-computable, and projectors of consciousness (such as brains) may be non-computable as well; yet, machines able to produce projectors of consciousness can still be computable. For instance, such machines could be designed in AutoCad.

Thirdly, we develop the notion of first-person non-reductive consciousness within a physicalist framework. This looks like overkill in a paper on meta-computability. However, without a somewhat thorough discussion of this point readers are likely to be confused what notion of the stream of consciousness is being used. The argument requires an analysis of consciousness that goes beyond the level of complexity usually adopted in such arguments and the following synopsis may be difficult to follow before the reader engages with the argument presented later in this paper:

C. Consciousness can be viewed in the following manner: (a) functional consciousness is a higher functional level of cognition; and (b) first-person phenomenal consciousness can be grasped solely from the epistemic first-person viewpoint. This view of consciousness seems to avoid the collapse of non-reductive naturalism (physicalism) into Cartesian dualism. Introduction of first-person epistemic consciousness may be facilitated by early Russelian monism viewed as a complementary approach to the ontic and epistemic viewpoints. It was later toyed with by Tom Nagel in *The View from Nowhere* and in his even earlier essay. It is important to be clear that our point is not to explain out non-reductive consciousness—what Chalmers calls the Hard Problem of Consciousness—the point is to provide a naturalistic account of how we can design a machine, very broadly understood, able to produce first-person consciousness. The notion of the metacomputable makes this distinction clearer, thus elucidating the Engineering Thesis in Machine Consciousness, to which we proceed in the last section of the article.

The fourth and final point is the application of the findings of Part I and II within the Engineering Thesis in Machine Consciousness:

D.   The Engineering Thesis in Machine Consciousness is an argument that if we know how first-person consciousness is generated in the brain, we should be able to engineer it. In this article it is re-formulated in the context of meta-computability (points A and B above) and in the context of *early* Russelian monism.

1. Introduction

In general terms, the Engineering Thesis in Machine Consciousness can be presented in the following way:

1.   Assumptions: AI should be able, eventually, to replicate functionalities of a conscious human being. Within non-reductive physicalism first-person consciousness is one such functionality. AI should also be able to replicate the content of conscious experience; produce a movie of what one sees, feels, and thinks. This is often identified, by philosophers, with phenomenal consciousness. However, if we record the content of one's visual experiences on a tape, e.g., based on fMRI of one's visual cortex [1,2] and other parts of the brain, this replicates the content of one's inner experiences without facing the so-called *hard problem of consciousness* [3].

2.   A philosophical claim: Since phenomenal content can be read from visual cortex (thus undermining the so-called *privileged access* claim, important to some philosophers), then the problem of privileged access is not the gist of the *hard problem of consciousness*. Moreover, contra David Chalmers, the creator of the notion of the Hard Problem [3,4], the Hard Problem is not the problem of experience, of its phenomenal qualities. The gist of the actual hard problem of consciousness, which lends plausibility to Chalmers' broadly appreciated argument, is the problem of the epistemic subject to whom those experiences appear. The very possibility of the epistemic subject is the gist of the hard problem, not specificity of phenomenal qualia that are (features of) internal objects in one's mind. Those internal objects, as Nagel, Husserl, and classical German philosophy observed, are necessary elements co-constituting the epistemic relationship of subject and object [5].

3.   The claim that consciousness is more like hardware than software points out that the carrier of the content of consciousness is not *more content*; it is more like a carrier-wave. In this sense, it is more like a piece of hardware than software; such software constitutes a condition of first-person epistemicity.

4.   The possibility of non-reductive physicalism: Due to feuding philosophical approaches to first-person consciousness, ranging from eliminative materialism [6] to substance-dualism [7], the issue of the first-person consciousness, and its *locus*, has not been formulated well enough for AI to tackle. Many philosophers do not distinguish between dualism and non-reductionism of the subject to the object, a deficiency which would make non-reductive physicalism conceptually impossible to even formulate.

5.   A brief version of the argument Engineering Thesis in Machine Consciousness: Having adopted the framework of broad non-reductive physicalism, I argue the following:

If we understand how the stream of first-person consciousness is generated in animals, including humans, we would be able to present the process generating it at the level of mathematical equations [8,9]. Contra to most forms of computationalism, consciousness is not just such equations, like a weather front is not just its computational simulation [10]. Yet, once we have the equations that describe correctly and sufficiently the way first-person consciousness gets generated, we would be able to use them to engineer a stream of consciousness [8,9,11,12]. Like in engineering of an artificial heart, whose function we understand already, we need to inlay those mathematically-grasped functionalities in the right kind of matter [8], which is not identical as running them in a program. This last point shows the difference between the simulation

of a process (e.g., running a computational simulation of a nuclear explosion) and the actual physical process (producing a nuclear explosion). In order for a process to be designed, either as a mathematical simulation, or as an experiment in nature, there needs to be a way to establish some kind of effectively computable conditions for such a process to occur [13]. Once we have such computation, and material science allows us to inlay the program for first-person consciousness in physical substance, we should be able to design a projector of first-person non-reductive consciousness. Such a projector may, or may not, be able to be inlayed only in organic matter. This naturalistic view on the subject of first-person consciousness seems to run counter to the point that the subject of consciousness is something akin to a soul, or at least the gist of a conscious living being [14]. My understanding is that identification of the gist of an animal, or a human being, with the locus of consciousness—is unfortunate and unnecessary. Maybe there is a gist of a living being, or actually a soul, but its identification with any specific functionality—such as having a beating hearth, breathing, thinking, speaking, or perceiving phenomenal qualia—is unnecessary and unfortunate. Those functionalities, if properly and narrowly defined, may be necessary for a living being like us, but they can also be engineered outside those beings. They may also function there, should such an artificial environment allow for their functioning. Having first-person consciousness, which we share with rats and frogs, seems to be one such functionality. One has to have some capacities to make any use of it, but there is no obvious need to assume that it is the gist of one's existence, whatever this last term would mean.

2. Part I

In Part I we introduce the notion of metacomputable. We also discuss computational dualism and computational monism.

2.1. Metacomputable

Incomputable processes are of two kinds. The first kind, their starting conditions are computable. The second kind, their starting conditions are non-computable. At this point in the argument I do not make claims whether any of those categories is, or is not, empty-satisfied. We call processes of the first kind metacomputable. We call processes of the second kind meta-incomputable. Machines can be computable if relevant features of their operation can be described, straightforwardly, as computable processes; this remark is important when one considers some of the physical interpretations of the Church-Turing thesis.

2.1.1. Metacomputable Holograms

A machine is metacomputable if it can be built by another machine that follows computable instructions, e.g., in CAD or a 3D printer. Natural processes, including the stream of consciousness in animals, can perhaps be characterized as metacomputable machines, of a biological kind. A process is metacomputable if it can be designed reliably enough (the clause that it has to be designed "reliably" would be a bit strong since it denies meta-computability to those processes that can be designed with limited reliability; yet, a good level of reliability seems required to bring them high enough above random noise).

There is a persuasive claim in favor of what I shall call "computational dualism"—the claim that, under no condition could a given incomputable process be designed by a computable system. If computational dualism was the case, meta-computability would only be possible through an admixture of incomputable causal chains with computable ones. For instance, an incomputable process can be produced by a computable machine and a Tuting oracle "found in nature".

Yet, I think the term we use should be open to the alternative, that a computable process can produce (reliably enough) a system that is incomputable. This option should be terminologically allowed as long as the question is open in a strong sense (as long as nobody presented an impossibility

proof, and interpretation of such a proof becomes uncontroversial). On the other hand, computational dualism seems uncontroversial: it is quite clear that at least some of the incomputable processes are metacomputable—one may plausibly claim that we can always find an oracle (an element of "nature"), for instance a random quantum occurrence, that introduces this incomputability, but the issue does not seem decided and closed one way or the other.

Let us shift from mathematical functions to events and processes (I made such shifts in the introduction, but now is the time for an explanation). What allows us to make the shift is an assumption, based on a strong interpretation of Church-Turing thesis [15], that all events can be described by discreet or continuous mathematical functions. This does not imply that we view Deutsch-style narratives as correct interpretations of Church-Turing thesis, but we view them as persuasive ontological claims that go beyond it.

The second important point is that there is an important difference between a computer running a program, as a mathematical simulation of an event, and that event happening in physical space outside of the computer. Obviously, a machine, such as a crane or industrial robot, guided by such a program, can produce the latter outcome; we call it to inlay the mathematical function in nature. Industrial robots do it all the time: they use a design, e.g., in AutoCAD, to make a physical object, by shaping some substance (e.g., iron) into an object, such as a car engine, a light bulb, or a projector of holograms. The last two examples, light bulbs and hologram projectors, have something in common—they can be conceptualized as computable machines, engineered to produce certain processes. Trivially, light bulbs produce some light. Also, hologram projectors also produce some light, but it results in 3D holographic projections. Now, if one views light as computable then a light bulb produces a process that is both computable and metacomputable; yet, if we were to view light as a wave-corpuscular entity that is incomputable, then the light produced by such a bulb would still be metacomputable (produced by a light-bulb, which is a computable device), while also being non-computable. The same goes for projectors of holograms, and machines that produce all kinds of other processes that can be at least conceived of as incomputable.

To sum up, an event, such as an electric discharge, can be effectively engineered; it can also be used to produce something beyond that process (e.g., an electric discharge used for producing a carving on a metal surface). We define a process, or event, as metacomputable if it can be designed reliably by a computable machine. It is meta-metacomputable if it can be designed reliably by a machine that is computable and itself effectively designable by a computable system. A meta-metacomputable process may or may not be metacomputable since, in theory, some process can be designed by some incomputable entity (say, a machine with an oracle), which itself would be designed by a computable one.

2.1.2. Metacomputable Consciousness

In this paper we are toying with the idea that consciousness, specifically the stream of consciousness, may be viewed as something like light, or a hologram, that can be projected. It could be produced by projectors of consciousness that are computable. In particular, brains may be computable—if so, the stream of consciousness would be metacomputable. Thus, the stream of consciousness can be a process that is not computable, yet metacomputable.

Here is a truly controversial thesis to consider: If the source of consciousness (e.g., brain) turned out not to be computable (for whatever reason, such as the inclusion of Turing's oracles somehow found in nature, quantum effects, or even just excessive complexity) but we still learned the brain's detailed blueprint—it would still be physically possible that something like it could be reliably built (e.g., bioengineered) by a system that is computable in operation. Hence, even if brains were not computable, consciousness could be meta-metacomputable. This point is not central to the argument of this article but it is relevant in discussion with those who claim that projectors of consciousness (brains or otherwise) are non-computable—both for relatively trivial (excessive complexity) or non-trivial reasons.

Back to the main argument: Let us pose the claim that consciousness, (viewed as a stream of sorts) may very well turn out to be incomputable [13]. We could still maintain that projectors of consciousness are computable. On the other hand, if brains (and/or other potential projectors of consciousness) were incomputable, two avenues are still open: First, it may still be possible that computable projectors of consciousness could in principle be built. For instance, that we could find sub-components of brains that are computable and also sufficient for generation of the stream of first-person consciousness—or, we may perhaps be able to develop alternative cognitive architectures that are computable, even if the stream of first-person consciousness in nature is always produced by incomputable brains (this would presume that it is not an essential feature of consciousness that it is produced by non-computable a source); Second, it may be possible to produce non-computable projectors of consciousness by computable processes, for instance by a Turing-computable program run by an industrial robot. Even if brains and other projectors of consciousness were incomputable, they could be metacomputable, which would make consciousness meta-metacomputable. Consciousness could be metacomputable, if it could be produced by computable projectors of consciousness, or meta-incomputable, if it could originate only from incomputable sources. In both cases, consciousness could still be meta-metacomputable.

The claim that consciousness is meta-metacomputable, is going to be our fallback option when we get to re-formulate The Engineering Thesis in Machine Consciousness [8,9,11,12]. Yet, the claim that a machine is incomputable does not imply that it cannot be made and used by some procedure effective in operation. We may define a machine that is incomputable through and through as a machine that is not metacomputable, or meta-metacomputable or not produced by any meta-meta-meta . . . computable process that can be tracked. Such a machine, incomputable through and through, may still be effectively constructible and effectively operable. An argument for this claim goes far beyond this paper since it requires a conceptual apparatus being developed elsewhere [13].

2.1.3. Ontological Implications of Metacomputable Revisited

The notion of *metacomputable* helps us distinguish between two ontologies. The first ontology is a computability-dualism that consists in accepting two kinds of entities and processes that do not mix: Within this ontology, incomputable things (processes or events) always originate from incomputable causes. The computable ones originate from computable causes. Non-computability in metacomputable systems may only originate from admixture of some incomputable element added to the computable process. Imagine a river of computable processes that merges with another river, this one of incomputable processes. The resulting New River has some computable and some incomputable waters, which never mix entirely. Kind of like a Cartesian person is a mixture of body and soul. Just like Cartesian dualism, this ontology suffers from the problem of interaction: in this framework the non-computable are never ontologically metacomputable (though, functionally, the system they create with computable processes may be computable, thus, its products are metacomputable). If the products come from a system that is an amalgamation of computable elements mixed (perhaps in a structurally sophisticated way) with some incomputable ones, then the total set of causes would not be computable and, in this sense, the process would not be metacomputable.

However, such a dualistic system may be computable in another way. Namely, the incomputable ontological component may not be relevant to the functional description of such a system. This is similar to latent genes or any system with an on and off switch, when the switch is off.

The relation between non-computable physical processes and non-computable mathematical functions is interesting, though slightly off topic. It seems that physical processes may be non-computable in at least two different ways: (a) by following non-computable, or at least not Turing-computable, mathematical functions; or (b) by not being describable in mathematical terms at all. The latter instance implies that a function cannot be constructed, such that it covers a given physical process; certain actual or hypothetical physical processes, such as Turing's oracles, cannot be described, consistently, by mathematical functions. The former instance may not easily apply to

standard Newtonian physics, but it may very well apply to Chaitin's understanding of quantum mechanics and other fields that sort of escape consistent description in standard logic.

This opens a broader issue, to what extent standard logic, and predicative language based on it, are able to provide a consistent conceptual system—the so called *inconsistency theory of truth* [16,17]. The topic is different from the issues generated by Godel's second theorem since it is not just about the lack of completeness proof, and its impossibility. It is a positive claim, an inconsistency claim pertaining to the structure of the logic of the language we use. This point may seems like a philosophically extreme interpretation of ancient paradoxes (especially *The Liar*) that brings only speculative consequences. Yet, in the world where computability becomes preeminent, due to the ubiquitous role of AI, this seemingly abstract claim seems likely to have practical implications, even in the creation of advanced semantic level (e.g., semantic web and other semantic applications) for AI.

The second computability-based ontology is computability-monism. Within this framework computable processes (that include functional description of objects, such as machines or events) can reliably produce incomputable results. In this ontology at least some of the incomputable processes are metacomputable. In practice, this means that physical incomputables may be engineered, straightforwardly, by computable methods. The latter notion may seem quite trivial—as long as it entails the claim that computable systems may produce incomputable outcomes, e.g., one may give computable instructions of how to build a *true* randomizer (e.g., quantum a randomizer) *assuming such randomizers are incomputable*. Yet, the standard interpretation, based on Turing's oracles, seems to be that such non-computable effects need to be taken "from nature", which indicates computational dualism.

To sum up the current sub-section, we developed the crucial claim of this article: even if the stream of consciousness is incomputable it does not imply that it cannot be produced by computable entities. Specifically, it does not imply that if the stream of consciousness is incomputable and it is produced by brains, then brains have to be incomputable (in the sense defined here, or in some other relevant sense). It also fails to imply meta-computational dualism: the claim that, if we established with mathematical precision how human brains produce consciousness and, therefore, that the relevant function of brains is computable (hence, that brains are functionally computable in the relevant sense) [13]—that this would imply that consciousness is also computable through and through. We have shown that both, within computational monism (quite easily), and within computational dualism (despite some complexities) it is not true that computable machines would never be able to produce an incomputable stream of consciousness.

2.1.4. Conclusions of Part I and Heads Up to Part II

Cognitive architectures that are produced by computable processes, or that can be designed using computable programs, may or may not be computable. The most straightforward way to attain incomputables from computables is to incorporate Turing's *oracles*, non-computable elements found in nature, within such a new system. Consequently, the stream of consciousness does not need to be computable even if it originates from computable (or, metacomputable) entities.

Having discusses the issue of metacomputable systems, which include brains and other cognitive architectures, we now focus on this latter example. There are two reasons why we zero in on generators of consciousness, and in a rather thorough fashion at that:

1. The issue provides an interesting application of metacomputable. In fact, my work on metacomputable systems originated from a discussion of non-reductive consciousness; and
2. The topic of non-reductive consciousness is complex and quick and easy attempts to cover it lead to confusion. Thus, we are trying to explain our terms well enough to avoid any major misunderstandings.

3. Part II

In Part II we define the stream of consciousness. We also present the most reduced notion of he epistemological subject. It is needed to put all objects on one side of the equation that we are

trying to tackle [18,19]. Even phenomenal objects [20], and *qualia*, turn out to be objects of some sort. This allows us to see clearly what is left when objects are out of the picture—what is left is essential for non-reductive subjectivity.

3.1. Stream of Consciousness and Complementarity of the Subject and Object

3.1.1. The Stream of Consciousness

Many authors define the stream of consciousness as the stream of experiences, or phenomenal content, or sense data. Such bundle theories identify the stream of consciousness with the stream of its content. This is a reasonable choice for those, like Hume, Price [21] or Parfit [22] who identify consciousness with whatever the content of conscious experiences is, at a given moment. Many authors who subscribe to this view accuse those who do not identify consciousness with its content of adhering to a dualism of sorts. Yet, in doing so they commit a tertium non datur fallacy. Indeed, Cartesian dualists [23,24] view souls, or non-material substance, as distinct from the content of experiences. However, first, not every dualism is Cartesian dualism burdened by the problem of interaction and other major infelicities. Second, and more importantly, not every non-reductive view collapses into dualism. I adopt the second approach, but the first one is also worth some attention. If non-material substance (whatever that means) is a substance nevertheless, could there be a science devoted to such a substance? This idea, hard to conceive within early Newtonian physics, may be more acceptable within some interpretations of contemporary science (e.g., quantum or string theory). The gist of dualism may still be preserved in any notion of the two substances that do not mesh with each other, but this may not lead to the conclusion that each of such substances remains outside of some broad realm of scientific investigation. Enough on non-Cartesian dualism, which is not the path I follow.

There are, however, non-dualistic ways to distinguish consciousness, or even the stream of consciousness, from its phenomenal content. This is the sole avenue open to non-reductive physicalism. One such way, popular in AI, is a mild version of reductionism—it is to view consciousness as a set of functionalities—from Jerry Fodor in philosophy to Stan Franklin in AI [25,26]. It would be a stretch to call such consciousness "a stream of consciousness" (except if one means a set of such functionalities executed in a temporarily ordered manner). Such functionality is not based upon phenomenal information, defined in this context as the sort of information that human beings view as phenomenal qualia (auditory, olfactory, tactile, visual, and other such impulses at the spectrum perceptible to most humans). What counts here is not the experience as phenomenal, but as a means of gaining the sort of data that animals gain from phenomenal experiences. This is because the first-person perspective, and related experiences, are not what matters in functionalist models of consciousness (they may not even be identifiable in that framework). The ability to gain informational content, and to integrate it into broader cognitive processes, is just one of the cognitive capacities that make a being conscious in the sense of functional consciousness. As mentioned above, in the functional definition, consciousness is an advanced form of cognition. It may either include: (1) some specific cognitive process, not available at the level of mere cognition; or (2) it may still have the same kind of cognitive processes that appear at the level of cognition, but a whole lot of them; or (3) it may be identical with (2) except that such processes are ordered in a meaningfully different way and, therefore, result in a new level of functionalities. Most likely, such processes have to be integrated in a specific advanced cognitive architecture. Functionalist views on consciousness distinguish consciousness from phenomenal content by using the third-person view on consciousness. This results in the lack of compelling reasons to give special consideration to phenomenal content. Obviously, functionalism is not phenomenalism.

On the other hand, functionalism lacks reasons to give special considerations to the locus of consciousness [27], or the first-person view [18,19], outside of both: functional content, as well

as the phenomenal one. This last issue is the hardest to demonstrate; we shall attempt such a demonstration now.

3.1.2. Subject and Object

Now we get to the view that is essential to this article: the distinction between phenomenal content (of consciousness) and the epistemic subject to which such content presents itself.

Let us begin with a simple point: Phenomenal features of consciousness seem like the gist of consciousness, at least within the first-person epistemic perspective. Yet, it is possible—and helpful—to distinguish phenomenal experiences from the epistemic subject. Counterintuitively, we may use Berkeley's extreme empiricism, which leads to ontological idealism, to give this idea a first approximation. Though secondary qualities are sustained in existence by the mind of God, for Berkeley [28] the mind of God is not just the stream of such qualities. This can be explained along the lines based on a re-interpretation of Cartesian dualism (God is the only non-phenomenal substance). However, a simpler model, and a bit more faithful to Berkeley, views God as the sole true subject, its thought constituting objects in the universe by way of secondary qualia without an independent ontological base (ontological status of human souls according to Berkeley goes beyond this paper).

German classical philosophy has brought this notion of subject a bit closer to the formulation that may be helpful in the contemporary debate. We rely to some degree on Leibniz, Kant [29,30], and Hegel, but to a larger degree on Fichte's Wissenschaftslehre [31,32] and the first book of Husserl's Ideas [33], elaborated by Ingarden [34]. The general point is that perceptions, including phenomenal qualia, are objects, at various levels of ontological generality—from sense-data or simple Gestalts, through ontologically under-defined objects of perceptions, and then objects in the world, all the way to the socio-cultural context, in which they acquire their always already social meaning. In order to make possible the basic epistemological reflection the object, at its most reduced, the minimal level is viewed as an atomic object (the most reduced notion of an object attainable at a given level of analysis) ([9], Section 6). This notion of the minimal object must be juxtaposed with the notion of the minimal epistemic subject. Such a subject has no ontology independent of its epistemic role of mirroring, reflecting (not "reflexing", as a reflection on the world, or self-reflection, as a higher intellectual function), putting within the epistemic realm the object of perception. Such a process is based upon the atomic objects as distinct from the minimal epistemic self, appearing as an atomic subject. No independent ontology of such a subject (no status as a mental substance), would be able to avoid Cartesian dualism—but such a heavy ontology of basic epistemic units seems unnecessary and counterintiuitve. An atomic epistemic subject is postulated as a non-substantial, merely epistemic, part of the primary (atomic) subject-object relationship.

While, as mentioned before, overzealous reductionists argue that any non-reductive epistemicity amounts to dualism, a non-reductive naturalist (or non-reductive physicalist) may, and needs to, accuse them of defining such non-reductive views out of existence. Pure subject, posed as the condition of epistemicity (of every phenomenal perception), does not entail mental substance within the Fichtean, or even Husserlian framework, nor within Tom Nagel's View from Nowhere [19]. The locus of epistemicity is posed as analogous to a mirror (or light, but the latter metaphor is overly complex) that is necessary for an object to appear as "an object of . . . ", be it an object of perception, of empirical knowledge, or of any epistemic function. This is due to the basic, and necessary, status of the subject-object relationship not just for epistemology, but also for ontology. Already Leibnitz noticed, rightly, that unknowable possible worlds are not ontologically possible—this is just because human ontology is always already also basically epistemic. We cannot predicate of ontologies that we have no direct or indirect epistemic access to. When we move from atomic objects to more complex epistemic structures we view epistemic subjects as a bit more complex system built of many objects that are able to enhance epistemicity (the latter may for instance be emergent, though basis subject-object relationship does not require emergentism. A bit later, we explain the complementary structure of subject and object perspectives, which is crucial to German Classical philosophy but also strong in

Anglo-American writers, especially Russell). In this sense the subject is more like hardware than software, more like the light produced by a light bulb than like information encoded in a program [35].

Information available to the person can be viewed, in a non-epistemic manner, as the transformation of ontological relationships; such an operation of matrices or maps in a brain was described by Damasio [36], and can also be cast as computational processes. Yet, when people talk of epistemic consciousness we focus on phenomenal content, always already anchored in the first-person experience—in particular in the subject-object relationship.

3.1.3. Complementary Monism: Russelian Analysis of Mind

Complementarity of subject and object is explained, by Russell's The Analysis of Mind [18]. Unlike Russell's more materialistically conceived neutral monism from The Analysis of Matter [37], this earlier idea can be named a 'complementary philosophy' [38,39]. Let us introduce it through the historical context of neutral monism: According to Spinoza [40], we have two ways to access the world—subjective and objective. Before coming back to Spinoza's neutralism, let us focus on those two viewpoints:

There are two completely different ways to view the world: From the subjective perspective, so aptly presented by Descartes in his Meditations [23], we see the world always as objects of phenomenal experiences. Moving beyond Spinoza, the open question, for all versions of phenomenalism, is the ontological status of those perceptions: Do they represent anything and, if so, what, and how can we know this? Ontological standing of represented objects, from such an epistemic viewpoint, is always an open question, known merely through inductive reasons.

Alternatively, we can see the world from the objective (or rather, objective) perspective, as a set of all the objects, often reducible to the set of all material objects (the way reistic materialism does). From this viewpoint the first-person consciousness is never fully explainable—though most authors who adopt this perspective deny the need for such an explanation (outside of explaining it away). This is the dominant view of the late 20th century, crowned with Armstrong's The Nature of Mind [41].

The refreshing aspect of Spinoza's position is that he is pre-Kantian, and anticipating Kant, the idea that those two viewpoints are not coming from nature of what there is to know (from the world). Instead, they come from human ramifications, and in fact limitations, of perceiving the world. Spinoza's speculation that angels perceive the world through a large, even infinite, number of such modes of perception, may—at first glance—strike us as a purely theological speculation. Yet, it should be viewed as a superb analysis of possible worlds that radically transcend human capabilities. This is an important point, which seems superior to the view of those who insist on dogmatically putting humans as the pinnacle of the universe. There is nothing scientific, or theological, about such human shauvinism. Within the Spinozian framework we may conceive not only of divine beings, but also of robots, extraterrestrials, animals, or humans of special abilities, from Leonardo da Vinci to Savulescu's people with engineered enhancements, who would access radically different ways of grasping the world than the epistemic (first-person) and ontological (third-person) perspective. Maybe the second-person perspective—Buber's I-Thou [42]—is such a third dimension that allows us for an essentially axiological viewpoint [43]. However, there could, in principle, be many more.

Russell's starting point was more straightforward and maybe less controversial—it is complementarity of ontology and epistemology (quite similar, in structure, to the Hegelian dialectics of subject and object). Let us develop this Russellian perspective a bit further. Russell [18] accepts and develops *the two perspectives view*. I agree with Robert Tulley [44] that this early version of Russell's neutralism received insufficient attention compared to the later version of his neutral monism. It is helpful to further clarify this position, maybe going beyond Russell's own framework. The main analytical point is that there is no re-translation among those genuinely different and complementary frameworks, one of which is essentially a first-person epistemology; yet, they are both ontologies in a broad sense—an epistemically based ontology of phenomenalism and a reistic ontology.

When we start our investigation at the level of objects, reistic materialism [45,46] becomes the most obvious option. Justification of any status of the first-person, phenomenal experience becomes a problem that can never be completely settled (or, it becomes a no-problem, if one misreads well-known Wittgenstein's remarks "Whereof one cannot speak clearly, thereof must one remain silent" as an endorsement of radical verificationism [47]).

When we start from the epistemic viewpoint we begin with phenomena. The phenomena are obviously all we know for sure, while the ontological status (or basis) of the content of phenomenal experience, is a matter of hypotheses—both in terms of its emergence base and veridical status (or, representational value). This sounds a lot like Hume's empiricism—but in Russellian monism of 1921, Hume has just a half-truth, one side of the complementary explanation. Interestingly, also the ontological status of the perceiving subject (not to be confused with its epistemic status) is far from obvious. Even philosophical beginners point out that Descartes is mistaken claiming that Cogito ergo sum [23]. Descartes made a hasty move assuming that phenomenal content allows us, in any direct manner, to assume that I am a thinking thing. Instead, I may be no 'thing' at all! That is how first-person epistemic certainty fails to translate onto any ontological substance whatsoever. Hume's conclusion that we are just the bundle of phenomena is an absolutization of the epistemic view, its infringement into ontology—similar, in its one sidedness, to Descartes' error. The Fichte-Husserl view on pure epistemic subject [31,33] seems like a better option—just because it is less ontologically heavy than Descartes' dualism, while being open to the uniqueness of first-person subjectivity advocated by Hume.

The above discussion brings about two complementary perspectives [38,39]: the first-person epistemic and the third-person ontic viewpoint. A philosopher is well-advised to shift between those complementary perspectives, just like a person with some forms of strabismus able to use only one eye at a time. The point is that we cannot do any better (a similar approach is obvious in the wave-corpuscular theory of light) . When we start with objects we use the ontological eye; when we start with phenomenal experience we use the epistemic eye. There seems to be no full retranslation between those views—a radically different justification-base, but not merely phenomenological "attitude" (in the Spinozian sense) is involved in those domains.

3.1.4. Summary of Section 3

We can address the issue of consciousness in either of the two frameworks. In the ontological framework the most persuasive answer is materialistic functionalism. Within this framework consciousness can best be established by its functionalities. Hence, the difference between consciousness and cognition is reduced to the difference in quantity or quality of performance. The issue of first-person consciousness can perhaps be addressed in this framework, if cast as the question of first-person experiences and their neural correlates. Such research is based on individual testimonies quantifiable among many research subjects and verifiable by correlation between those testimonies (on feelings and experiences) and states of the central nervous system (established through fMRI or other empirical methods), as well as behaviors of large numbers of individuals. Such answers fit, broadly speaking, with various stripes of broad reductive physicalism.

In the epistemic framework the answer about the question of first-person consciousness can be addressed more directly; phenomenal content is a given. The framework is already essentially first-person (even the fact that somebody may be lying about her phenomenal content is a very remote concern—trivial lies are of limited philosophical interest and lie-detectors seem to catch most of those. A somewhat more central concern is that people may have different, and confusing, ways of expressing and communicating about those phenomenal experiences, but this is not the main point either). One does not focus on the intersubjective verification of first-person content of experience in the epistemic framework. Such content is by far more obvious, and in this local sense, better grounded, than any form of inter-subjectivity. To put it paradoxically, it is more objective. This is not ontic objectivity but epistemic undeniability of first-person experience. First-person content leaves room for all kinds of ontological extrapolation about the existence of

external objects, other persons, the society, etc., which are not given, but always already constituted by (or even constructed from) epistemic experiences (one example of such constructivism, overly complex for the present task, is Kant's transcendental idealism, with the mind co-constituting the world [30]). The epistemic framework is essentially non-reductive. Importantly, solely acceptance of such epistemic framework—complementary with the ontological one—allows for true non-reductive views on consciousness.

This is important for non-reductive physicalism. There can be no truly non-reductive subject, if it is not presumed at the very beginning of one's philosophical explanations. Neither could there be a true way out of essential, though not necessarily practical, skepticism about the existence of the external world unless it is already presumed. I say that the only version of non-reductive naturalism relies on the very basic complementarity of subject and object.

4. The Engineering Thesis in Machine Consciousness Revisited

What are projectors of the stream of consciousness? Under the minimal view they are animal brains [14]. Yet, the whole area of research in artificial general intelligence, including biologically-inspired cognitive architectures, indicates that all functionalities of animals, including the functionalities of CNR, should eventually be transferable into advanced artificial cognitive architectures [48]. This is one way of attaining artificial general intelligence [49,50]. The project is not limited to the informational content, but also to other functionalities.

Is the first-person cognitive stream one such functionality? It would be quite troubling for materialism if it weren't.

4.1. Consciousness as the Epistemic Locus

The question is: How do we engineer consciousness understood as a locus of individual epistemicity [26] that can be addressed the way Russell [18], Nagel [19], as well as Husserl [33] describe the epistemic (or pure transcendental) subject?

The Engineering Thesis

Let us reformulate the above question in the following way: under what conditions could machines become first-person conscious in the sense often described as non-reductive consciousness? I think physicalism (and also panpsychism, as David Chalmers pointed out to me), entail that such conditions must exist. Whether such conditions can be satisfied in a world like ours is an open question, but we should adopt at least an attitude of moderate optimism about it.

The argument:

Step I.　If (1) humans have non-reductive consciousness, and if (2) science can, in principle, explain the world, then (3) science should, in principle, be able to explain non-reductive consciousness.

Step II.　To explain some process scientifically, in a strict sense, means to provide a mathematical description of that process (4). As it happens, such a description constitutes an engineering blueprint for creating that process (5).

Step III.　Hence, if some day science explains how non-reductive consciousness is generated, it would thereby provide an engineering blueprint of non-reductive consciousness (6). It would be a matter of standard engineering to turn such a blueprint into the engineering product (7).

Step IV.　If engineering non-reductive machine consciousness does not solve, or butcher, the hard problem of consciousness—this is because we assume the "black box approach". We should be able to build projectors of consciousness even if we are unable to explain it in an ontological sense; this is a practical advantage of the engineering approach over more philosophical attempts (8). The ontological status of non-reductive consciousness (whatever this means), is distinct from the way it can be engineered. If it is engineered properly, the ontological status of engineered first-person consciousness would be relevantly similar to such a 'natural'

consciousness, though its genesis and social role may be different from that of a biological and social being.

Step V. People raise an epistemic problem how would we know that a being, e.g., a machine, has non-reductive consciousness? This is a problem since there are reasons to believe that functional consciousness can be engineered without first-person non-reductive phenomenal consciousness. This may be answered in terms of hard- and soft-AI, and even more clearly by strong physical interpretation of the Church-Turing thesis [15]. Yet, the main answer is simpler: this is a special version of the problem of other minds. Few, even among philosophers, seriously doubt the existence of other minds anymore. If we have a good engineering blueprint we should have decent epistemic reasons to believe that, by following it, we would attain first-person consciousness. A philosophically more interesting answer would be based on a version of Chalmers' 'dancing qualia' argument [51]. Having established what functionality makes a human being first-person aware, a future scientist surgically removes that part (say, a thalamus) and replaces it with the artificial one, then records behavior, which includes self-reporting and neural correlates in the rest of CNR. Finally, the experimenter removes the artificial implant and implants back the natural part of the brain. If there is no significant difference in self-reporting, behavior, the main measurements of neural correlates, and the feel based on the memory of the periods when the natural and, the other time when the artificial generator of consciousness was in use—then there are good reasons to believe that the experiment of creating a generator of consciousness has succeeded. This point is valid provided that the shift among the centers of first-person perspective in the revised dancing qualia thought experiment does not cause the changes in the memory of events lived with different projectors of consciousness.

An important issue is whether non-reductive consciousness is epiphenomenal. Even if it is, we may have reasons to care one way or the other [52], but in this paper let us assume that it is not. While the issue goes beyond this article [13], let us pose that, in our possible world, the emergence base of first-person consciousness may be in a causal one-one relationship with important functional characteristics. This opens some room for Baars' phenomenal markers view on the role of consciousness [53]—however, if this approach was adopted qualia would be a functional characteristic. This can be seen as a reason not to view qualia (an aspect of the content of consciousness) as the carrier of pure subjectivity. I agree with Baars in part, that qualia provide phenomenal markets, and disagree in part, that this point provides an explanation of first-person consciousness. There is part of the problem that Baars does not even touch on, and that is the locus of the epistemic first-person viewpoint.

The stream of consciousness that carries phenomenal content may also be the carrier of subjectivity—but this would be its other, distinct function. It is what I called earlier in the paper the carrier wave. This carrier wave could be measured [54], based on the knowledge in the area of medical anesthesia. Tononi provides five different strengths (kinds) of consciousness that follow the standard levels of anesthetic intervention. This is a good beginning of such measurements, even though mathematical specifics of Tononi's Phi function remain controversial. If the stream of consciousness can be measured, it implies that first-person consciousness is detectible in some way. However, anesthesia (e.g., through fMRI) measures the functioning of consciousness, and its correlates in the patient's brain, It does not directly measure, or detect, the first-person non-reductive epistemic subject.

The carrier wave of phenomenal consciousness, viewed from the first-person perspective, provides the *locus of first-person epistemicity* [27]. *The carrier wave of consciousness* should be viewed as a natural process that provides ontological underpinning of both, phenomenal content and first-person consciousness. The proposed view is not a double aspect theory, but rather "a triple aspect theory" that distinguishes among: (1) phenomenal content; (2) the carrier wave that holds this content (a bit like a stream of photons carries visual information); and (3) the non-reductive *epistemic subject* of first-person experiences. Such a carrier wave could be effectively engineered, via computable

instructions given to a machine [9,11], even if the wave was incomputable. It is sufficient for the carrier wave of phenomenal consciousness to be metacomputable (constructible by a computable machine), or even meta-metacomputable, for the process to be effective. This is one way to view non-reductive physicalism, though this article contains only a sketch of an argument to this effect.

There is one more argument of how measurement of first-person machine consciousness should be possible through scientific means. It seems conceptually possible to measure the stream of consciousness apart of its neural correlates and the regular functionalities of conscious beings. This can be done if there exist direct consequences of conscious activity. In his upcoming article Darmos [55] tries to test "if something has consciousness" by using the Copenhagen interpretation of quantum theory. If, as Wiener claims, conscious observation was influencing quantum stream in a measurable way, then it would be measurable whether a supposed projector of consciousness (a living brain or machine) were in fact projecting consciousness. Hence, objective readings of quantum stream would be a test of what is conscious. The interpretation of quantum effects as depending on consciousness has been largely rejected in physics decades ago. Yet, the fact that it used to be viewed as a strong hypothesis (and a few quantum physicists, such as Stapp, still seem to endorse it) points to at least a conceptual possibility of such experimental evidence.

The next points provide the review of the main issues of the article.

4.2. Stream of First-Person Consciousness and Computability

As we remember from Part I of this article, the stream of consciousness may not be computable, but reliably metacomputable. Brains can be computable even if the stream of consciousness is not. Based on a very broad take on the physical interpretation of the Church-Turing thesis [15], one would be inclined to view brains as the kind of objects that are computable. One way to describe future developments of neuroscience is to view it as providing more and more accurate understanding of the computations made by brain processes. Those computations are of course inlayed in nature [8], they have to be performed by some physical device; they are instantiated in such a device [11].

As the Engineering Thesis presumes, someday we should be able to understand the specific computations within the brain that, if instantiated in the processor that can run them, result in a stream of consciousness. The point might imply computationism, except for 'the right kind of substance' clause, which needs to be taken rather seriously. Consciousness is not just a mathematical process but a natural process—while it can be simulated mathematically, such simulation is not its instantiation. We would possess full technical understanding of how the stream of consciousness is to be created; yet, only certain materials can run a given process. These computations should produce a stream of consciousness, if inlayed in nature (instantiated in a material able to run them—not just simulated). The difference between running a process and simulating it may sometimes be difficult to see. It becomes clear when we take into account the difference between simulating a tsunami and an actual tsunami taking place in the physical environment.

First-person conscious processes are metacomputable if brains or brain-like structures are computable (less likely, structures able to produce a first-person stream may themselves be incomputable, but they would be meta-meta computable if they could be built through a computable, and somewhat predictable, process.)

4.3. Consciousness as Hardware and Epistemicity

I argued above that the stream of consciousness is like a stream of light, or like a number of streams of light that create a hologram [19]. Since light is an anomalous physical object, this has led me to believe that the stream of consciousness is more like hardware than software.

This approach is a good antidote against post-Humean excessive focus on phenomenal content of the stream of consciousness and insufficient attention given to the carrier of such a stream. Consciousness carries phenomenal content just like energy carries an electromagnetic waves with informational content inlayed in them. However, this is not the main issue. What matters is the

problem of first-person consciousness, the problem of the self as the basis of epistemicity is raised by Nagel's early works [19] and by early Russell [18]. Russell in his, now more read, *Analysis of Matter* [37], and Nagel in his 2012 surprisingly unenlightening book [56], reversed themselves in the ways I find unfortunate.

The self is a subject that is not an object—it is merely the base of non-derivative epistemicity (or, if put somewhat ambiguously, *intentionality*). This is the gist of the problem of the non-reductive self. Even at this most essential level, unless one endorses a mysterian view on first-person consciousness, epistemicity should be something that could, in principle, be engineered in biologically-inspired cognitive architectures. This result, negligible from the (always functional) perspective of the old-school AI, philosophically stands as essential. It is essential in ontology of artificial subjects and their epistemology. Epistemic subject would become a locus of not merely functional consciousness, but the condition of epistemicity—of any first-person view. The result is non-trivial for applied philosophy and ethics; for example, the argument that artificial companions cannot engage in real relationships with people since they do not have first-person epistemic perspective (or intentionality) the way human beings and animals do [57], would not apply to artificial beings that satisfy the Engineering Thesis.

Rejection of the Engineering Thesis looks, in this light, like a very weak view. Such a rejection follows from (as well as leads towards) one of the two extremes: either mysterianism or reductivism about first-person consciousness. Within mysterianism on consciousness, epistemicity (what Searle calls non-instrumental intentionality [14]) turns out to be non-transferable to non-animal projectors of consciousness. Such a view would be mysterious since it places animal minds beyond the scope of engineering sciences. However, why would we be able to engineer, in multiple ways, functioning animal hearths and would we not be able—when the technology develops further—to engineer functioning animal brains, and in particular their centers of first-person consciousness?

A somewhat overlapping negative conclusion may be reached by reductive materialism. Here the first-person self is denied by an overactive Ockham's razor; namely, an overly strong requirement of intersubjective verification of one's first-person experience. When we replace post-Wittgensteinian verificationism (clearly visible not only in the work of Ryle, but also Dennett [58]) with inference to the best explanation (Harman [59–62]), we can talk of non-reductive naturalism (or, physicalism or materialism). Such non-reductive naturalism is able to navigate between the Scylla of mysterianism on consciousness and the Charybdis of reductivism. The claim that eventually we should be able to engineer projectors of first-person consciousness is a consequence of this approach [8,9,11].

5. Conclusions

In this article the author adds two points to the Engineering Thesis in Machine Consciousness, which had been proposed about ten years earlier. First, we show, in Part I, that projectors of consciousness may be computable (or metacomputable), while consciousness does not need to be so. Hence, such projectors should be designable and constructible when quantifiable knowledge of functioning of the brain and the right engineering means become available. This is true even if we were to find out that the stream of consciousness, itself, was non-computable, as some of the followers of quantum theories of consciousness seem to believe (Goertzel, Hameroff).

In Part II we show that non-reductive naturalism, which allows for the stream of first-person consciousness, requires a complementary view on subject and object. Each of those categories must be posed to be irreducible to the other. The subject is viewed as the source of epistemicity; the object as the source of reistic objecthood. Early Russellian monism allows for non-reductive naturalism, which helps us formulate the conversation about truly human-like consciousness for machines.

Within the framework of non-reductive physicalism, we maintain that artificial subjects may create their own internal *loci* of epistemicity [63,64]. In the world of artificial projectors of first person consciousness, there is something that it is like for fully-fledged artificial subjects to be them. They are spectators, epistemic *loci*—not merely the content. Reflection, or self-reflection, of the stream of consciousness is an advanced functionality that relates to a complex cognitive system—such systems

are objects in the world. The source of epistemicity presents itself as a feature of cognitive systems. Viewed as a subject that is not an object, it does not contain any identifiable structure; the structure of cognitive systems is always already the structure of objects, namely, machines that produce such a cognitive architecture. Thanks to pure epistemic subjects, first-person conscious organisms (and perhaps future machines), should not be seen as just information, or functionally-conscious zombies, but as subjects of first-person consciousness.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Nishimoto, S.; Vu, A.T.; Naselaris, T.; Benjamini, Y.; Yu, B.; Gallant, J.L. Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. *Curr. Biol.* **2011**, *21*, 1641–1646. [CrossRef] [PubMed]
2. Kay, K.N.; Naselaris, T.; Prenger, R.J.; Gallant, J.L. Identifying natural images from human brain activity. *Nature* **2008**, *452*, 352–355. [CrossRef] [PubMed]
3. Chalmers, D. Facing Up to the Problem of Consciousness. *J. Conscious. Stud.* **1995**, *2*, 200–219.
4. Chalmers, D. Moving Forward on the Problem of Consciousness. *J. Conscious. Stud.* **1997**, *4*, 3–46.
5. Boltuc, P. Reductionism and Qualia. *Epistemologia* **1998**, *4*, 111–130.
6. Wilkes, K. Losing Consciousness. In *Consciousness and Experience*; Metzinger, T., Ed.; Ferdinand Schoningh: Meppen, Germany, 1995.
7. Lowe, E.J. Non-Cartesian substance dualism and the problem of mental causation. *Erkenntnis* **2006**, *65*, 5–23. [CrossRef]
8. Boltuc, N.; Boltuc, P. Replication of the Hard Problem of Consciousness. In *AI and Consciousness: Theoretical Foundations and Current Approaches*; Chella, A., Manzotti, R., Eds.; AAAI Press: Merlo Park, CA, USA, 2007; pp. 24–29.
9. Boltuc, P. The Philosophical Issue in Machine Consciousness. *Int. J. Mach. Conscious.* **2009**, *1*, 155–176. [CrossRef]
10. Piccinini, G. The Resilience of Computationalism. *Philos. Sci.* **2010**, *77*, 852–861. [CrossRef]
11. Boltuc, P. The Engineering Thesis in Machine Consciousness. *Techne Res. Philos. Technol.* **2012**, *16*, 187–207. [CrossRef]
12. Boltuc, P. A Philosopher's Take on Machine Consciousness. In *Philosophy of Engineering and the Artifact in the Digital Age*; Guliciuc, V.E., Ed.; Cambridge Scholar's Press: Cambridge, UK, 2010; pp. 49–66.
13. Copeland, J.; Boltuc, P. Three Senses of 'Effective'. **2017**, in press.
14. Searle, J. *Mind, a Brief Introduction*; Oxford University Press: Oxford, UK, 2004.
15. Deutsch, D. Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer. *Proc. R. Soc. Ser. A* **1985**, *400*, 97–117. [CrossRef]
16. Chihara, C. The Semantic Paradoxes: A Diagnostic Investigation. *Philos. Rev.* **1979**, *88*, 590–596. [CrossRef]
17. Barker, J. Truth and Inconsistent Concepts. *APA Newslett. Philos. Comput.* **2013**, 2–12.
18. Russell, B. *The Analysis of Mind*; Allen and Unwin: Crows Nest, Australia, 1921.
19. Nagel, T. *The View from Nowhere*; Oxford University Press: Oxford, UK, 1986.
20. Block, N. On a Confusion about a Function of Consciousness. *Behav. Brain Sci.* **1995**, *18*, 227–287. [CrossRef]
21. Price, H.H. *Perception*; Methuen & Company Limited: London, UK, 1932.

22. Parfit, D. *Reasons and Person*; Clarendon Press: Oxford, UK, 1984.

23. Descartes, R. *Discourse on the Method, Optics, Geometry and Meteorology*, Revised edition; Olscamp, P.J., Ed.; Hackett: Indianapolis, IN, USA, 2001.

24. Descartes, R. *Meditations on First Philosophy*; Cottingham, J., Ed.; Cambridge University Press: Cambridge, UK, 1996.

25. Fodor, J. *The Modularity of Mind*; MIT Press: Cambridge, MA, USA, 1983.

26. Franklin, S.; Baars, B.; Ramamurthy, U. A Phenomenally Conscious Robot. *APA Newslett.* **2008**, *8*, 2–4.

27. Shalom, A. *Body/Mind Conceptual Framework and the Problem of Personal Identity: Some Theories in Philosophy, Psychoanalysis and Neurology*; Prometheus Books: Amherst, NY, USA, 1989.

28. Berkeley, G. *Treatise Concerning the Principles of Human Knowledge*; Jacob Tonson: London, UK, 1734.

29. *The Philosophical Works of Leibnitz*; Duncan, G.M., Ed.; Tutle Morehouse Taylor: New Heaven, CT, USA, 1890.

30. Kant, I. *Critique of Pure Reason*; Cambridge University Press: Cambridge, UK, 1781.

31. Fichte, J.G. *The Science of Knowledge: With the First and Second Introductions*; Heath, P., Lachs, J., Eds.; Cambridge University Press: Cambridge, UK, 1970.

32. Siemek, M.J. *Die Idee des Transzendentalismus bei Fichte und Kant*; Felix Meiner Verlag: Hamburg, Germany, 1984.

33. Husserl, E. *Ideas: General Introduction to Pure Phenomenology*; Routledge: Abingdon, UK, 1931.

34. Ingarden, R. *Der Streit um die Existenz der Welt: Existentialontologie*; Niemeyer, M., Ed.; The University of California: Oakland, CA, USA, 1964.

35. Boltuc, P. First-Person Consciousness as Hardware. *APA Newslett. Philos. Comput.* **2015**, *14*, 11–15.

36. Damasio, A. *Self Comes to Mind: Constructing the Conscious Brain*; Vintage Books: New York, NY, USA, 2010.

37. Russell, B. *The Analysis of Matter*; Spokesman Books: Nottingham, UK, 1927.

38. Boltuc, P. *Ideas of the Complementary Philosophy. (Pol: Idee Filozoff Komplementarnej)*; Warsaw University: Warsaw, Poland, 1984; pp. 4–8.

39. Boltuc, P. Introduction to the Complementary Philosophy. (Wprowadzenie do filozofii komplementarnej). *Colloquia Communia* **1987**, *4*, 221–246.

40. Spinoza, B. *Ethics*; Curley, E., Ed.; Penguin Classics: London, UK, 2005.

41. Armstrong, D. *The Nature of Mind*; University of Queensland Press: Brisbane, Australia, 1966; pp. 37–48.

42. Buber, M. *I and Thou*; Kaufmann, W., Ed.; Charles Scribner and Sons: New York, NY, USA, 1970.

43. Boltuc, P. Is There an Inherent Moral Value in the Second-Person Relationships? In *Inherent and Instrumental Value*; Abbarno, G.J., Ed.; University Press of America: Lanham, MD, USA, 2014; pp. 45–61.

44. Tully, R. Russell's Neutral Monism. *J. Bertrand Russell Stud.* **1988**, *8*, 209–224. [CrossRef]

45. Kotarbiński, T. Elementy teorii poznania, logiki formalnej i metodologii nauk, Lwow: Ossolineum, 1929; English translation (with several appendixes concerning reism), Gnosiology. In *The Scientific Approach to the Theory of Knowledge, Trans*; Wojtasiewicz, O., Ed.; Pergamon Press: Oxford, UK, 1966.

46. Quinton, A. *The Nature of Things*; Routledge: London, UK, 1973.

47. Wittgenstein, L. Logisch-Philosophische Abhandlung. In *Annalen der Naturphilosophie*; Wilhelm, O., Ed.; Verlag von Veit & Comp.: Leipzig, UK, 1921.

48. Samsonovich, A.V. On a roadmap for the BICA Challenge. *Biol. Inspired Cogn. Archit.* **2012**, *1*, 100–107. [CrossRef]

49. Goertzel, B.; Ikle, M.J.; Wigmore, J. The Architecture of Human-Like General Intelligence. In *Theoretical Foundations of Artificial General Intelligence*; Wang, P., Goertzel, B., Eds.; Atlantis Press: Paris, France, 2012.

50. Goertzel, B. *Mapping the Landscape of Human-Level Artificial General Intelligence*; AI Magazine: Menlo Park, CA, USA, 2015.

51. Chalmers, D. *The Conscious Mind: In Search of a Fundamental Theory*; Oxford University Press: Oxford, UK, 1996.

52. Boltuc, P. Church-Turing Lovers. In *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*; Lin, P., Abney, K., Jenkins, R., Eds.; Oxford University Press: Oxford, UK, 2017; pp. 214–228.

53. Baars, B. *A Cognitive Theory of Consciousness*; Cambridge University Press: Cambridge, UK, 1993.

54. Tononi, G. Integrated information theory of consciousness: An updated account. *Arch. Ital. Biol.* **2012**, *150*, 290–326.

55. Darmos, S. Quantum Gravity and the Role of Consciousness. In *Physics: Resolving the Contradictions between Quantum Theory and Relativity*; CreateSpace Independent Publishing Platform: Hong Kong, China, 2014.

56. Nagel, T. *Mind and Cosmos*; Oxford University Press: Oxford, UK, 2012.

57. Guarini, M. Carebots and the Ties that Bind. *APA Newslett. Philos. Comput.* **2016**, *16*, 38–43.

58. Dennet, D. Quining Qualia. In *Consciousness in Modern Science*; Marcel, A., Bisiach, E., Eds.; Oxford University Press: Oxford, UK, 1988; pp. 42–77.

59. Harman, G. *Change in View: Principles of Reasoning*; M.I.T. Press/Bradford Books: Cambridge, MA, USA, 1986.

60. Harman, G. Can Science Understand the Mind? In *Conceptions of the Mind: Essays in Honor of George A. Miller*; Harman, G., Ed.; Lawrence Erlbaum: Hillside, NJ, USA, 1993; pp. 111–121.

61. Harman, G. Explaining an Explanatory Gap. American Philosophical Association Newsletter on Philosophy and Computers. Ph.D. Thesis, University of Arizona, Tucson, AZ, USA, 2007; pp. 2–3.

62. Harman, G. More on Explaining a Gap. *Am. Philos. Assoc. Newslett. Philos. Comput.* **2008**, *8*, 4–6.

63. Evans, R. A Kantian Cognitive Architecture. *Philos. Stud.* **2017**, in press.

64. Evans, R. Kant on Constituted Mental Activity. *APA Newslett. Philos. Comput.* **2017**, 41–54.