Article

The Difficulties in Symbol Grounding Problem and the Direction for Solving It

Jianhui Li [1,2,*] **and Haohao Mao** [3]

1   Academic Registry, BNU-HKBU United International College, Zhuhai 519000, China
2   Research Center for Value and Culture, School of Philosophy, Beijing Normal University, Beijing 100875, China
3   School of Marxism, Anhui University of Science and Technology, Hefei 232001, China
*   Correspondence: jianhuili@uic.edu.cn or ljh@bnu.edu.cn

Abstract: The symbol grounding problem (SGP) proposed by Stevan Harnad in 1990, originates from Searle's "Chinese Room Argument" and refers to the problem of how a pure symbolic system acquires its meaning. While many solutions to this problem have been proposed, all of them have encountered inconsistencies to different extents. A recent approach for resolving the problem is to divide the SGP into hard and easy problems echoing the distinction between hard and easy problems for resolving the enigma of consciousness. This however turns out not to be an ideal strategy: Everything related to consciousness that cannot be well-explained by present theories can be categorized as a hard problem which as a consequence would doom the SGP to irresolvability. We therefore argue that the SGP can be regarded as a general problem of how an AI system can have intentionality, and develop a theoretical direction for its solution.

Keywords: symbol grounding problem; SGP; meaning; artificial intelligence; Chinese Room Argument

1. Introduction

The symbol grounding problem originates from the "Chinese Room Argument" formulated by John Searle [1]. According to the original thought experiment and its reformulation in later presentations [2,3], the argument profoundly attacks computationalism or the notion of strong AI (Artificial Intelligence). Searle believed that all computers merely process inputs and outputs consisting of strings of symbols, while the human mind not only is manipulating symbols but also has the capacity of understanding their meanings. He thus insisted that artificial intelligence systems cannot have human-like intelligence owing to their inability to understand. If someone wanted to claim that artificial intelligence had human like intelligence and could understand meanings the way human beings do, he/she would need to demonstrate that artificial intelligence could understand the meaning of the symbols it processed. As Stevan Harnad postulated, "How can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meanings in our heads? How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols?" [4] (p. 335). This problem is called the Symbol Grounding Problem (SGP for abbreviation).

Harnad attended to the problem raised by Searle's argument and refined it as SGP. In this context it is noteworthy that The HOW question in fact presupposed that this problem could be solved in some way, while Searle's argument suggests that it is impossible to obtain meanings for AI which only manipulates symbols. Without a more in depth discussion about the philosophical premise of his theory, Harnad proposed a strategy of a hybrid symbolic/sensorimotor system, in which meanings coming from the external world would enter a perceptual system to obtain perceptually invariant features. These features would be classified into different categories, the names of which are the basic symbols which refer to the features of the categories. He argued that this would present a possible solution due to

its ability to break up the closed system in the "Chinese Room", in which symbols could only be transferred into other symbols. According to his strategy, the information of the external world is introduced to the system as the source of meanings. Mapping his approach against causal theories of mental content supported for instance by Jerry Fodor [5], Fred Dretske [6] and others, in which the relations between representations and their underlying content are determined by a causal processes, we can conclude that Harnad's strategy heavily relies on causal processes as well: the sensorimotor system is a typical causal system which receives information from the world that triggers a causal effect (interacting with the world to adjust its meaning-symbol relations). In fact, as Harnad focused on the ability of AI to interact with the world, his strategy could be included in the "robot reply" to Searle's argument, and Harnad himself introduced the term "robotic functionalism" to oppose symbolic functionalism [7]. The core idea is that the non-symbolic functions (e.g., sensory, motor functions) which can directly interact with the world are primary, by which the symbolic system can be grounded. Therefore, he did have a functional premise: If a symbol system can be in a causal relation with the external world, it can access its meaning.

In consequence this led to the following intricacy: If causal theories form the presupposition of the hybrid system, the SGP can be classified as an "easy problem" due to the fact that any system that supplies a causal access between meanings and symbols can solve the SGP. But we have to take into account the term understand in Searle's argument is explicitly different from merely causal relations. Therefore, it is indispensable to determine first whether the problem at hand is a Searle-style problem, i.e., "how can the AI system understand", in order to develop an appropriate solution.

Unfortunately, many other scholars who have studied the SGP seem to have skipped this problem like Harnad and continued their research on unclarified premises. Mariarosaria Taddeo and Luciano Floridi reviewed eight solutions to the SGP and classified these strategies into three categories: representationalism, semi-representationalism and non-representationalism [8]. The majority of these solutions is functional and does not provide complete explanations to Searle's problem.

Regardless of the growing amount of academic work on the matter, new solutions and theories continue to be put forward. Obvious inconsistencies in the SGP have been addressed by several scholars, who however failed to isolate the crucial point. The status quo remains that in spite of many solution proposals for the SGP, there is no concluding discussion on whether the problem has been solved or not which eventually stimulated new scholarly studies. In the article "The Symbol Grounding Problem has been solved. So what's next?" Luc Steels drew on Charles Sanders Peirce's semiotic ideas and methods and claimed that the problem was solved [9], whereas Selmer Bringsjord argued in the article "The symbol grounding problem...remains unsolved" that a robot must have similar semantic knowledge as humans to be able to ground its symbols, so that according to the current state of the art no solution was available [10].

However, as they did not touch on the most controversial point, the issue is mired in bias. In these ongoing discussions, recently a clear direction has emerged, namely that of dividing the SGP in a manner similar to the division between the "hard" and "easy" problems of consciousness proposed by David Chalmers [11]. This is reasonable to some extent because the ultimate goal of solving the SGP is to enable AI to understand their language; in Searle's argument the ability to understand requires the AI to obtain intrinsic intentionality, which is closely related to conscious experience. As a result of compromise, they had to admit that there were two kinds of SGP: the solvable one which only related to derived intentionality and the unsolvable one that related to intrinsic intentionality. Maybe due to the fact that Harnad included the term "feeling" in some of his articles [12], or because scholars just found the hard/easy division of the consciousness problem was similar to the situation of the SGP, they applied the hard/easy division to the SGP, which under closer scrutiny turns out not to be appropriate.

But before explaining why this kind of division is not appropriate and proposing an alternative pathway to solving the SGP, we need to review the main strategies of SGP solutions to see how and where difficulties arise.

2. Main Solutions of the Symbol Grounding Problem and Their Problems

So far, multiple solutions to the SGP have been proposed. Following Floridi's review, we will discuss the two main strategies of solving the SGP: Harnad's hybrid symbolic/sensorimotor system [4]; and the semiotics strategy, also called "physical symbol grounding", proposed by Paul Vogt and Luc Steels [9,13]. Additionally, "the physical grounding hypothesis" proposed by Rodney Brooks [14] was fairly influential, but the SGP has rather been avoided than solved due to the lack of a symbol system, hence we will not evaluate it in more detail. Moreover, we will discuss the "zero semantical commitment" proposed by Floridi and Taddeo to examine these strategies [8] as well as the arguments against it.

2.1. Harnad's Hybrid Symbolic/Sensorimotor System

In the "Chinese Room Argument", Searle noted that semantics cannot be produced by a computer program that exclusively relies on syntactics. Harnad endorsed Searle's standpoint but differed from Searle in arguing that if a non-symbolic system that can obtain real semantics from the external world can be combined with the symbolic system, the meaning of these symbols can be accessed by the system. For example, the image database of apples may be fed to a neural network, and by extracting the "perceptually invariant features" (e.g., the redness and roundness of apples), machines are able to form a category of apples, and the name of the category is the symbol which refers to apples. In this system, the images of apples are called iconic representations; the categories with perceptually invariant features are called categorical representations and the names of categories are called symbolic representations.

(1)   Iconic representations: They are the projections of distal objects onto proximal perceptual organs such as images or sounds. For example, the data could be "the many shapes of an apple projected onto our retina." These raw data are further processed and abstracted as categorical representations.
(2)   Categorical representations: They are features extracted from iconic representations. They are common perceptual features shared of the data. For example, for all apples, the features of a red or cyan color and approximate roundness are common to all apples. Categorical representations are the basic units of meanings and the names of these categories are the basic symbols of the symbol system.
(3)   Symbolic representations: They are composed of basic symbols that designate various categories. For example, a symbolic representation such as "zebra" is composed of two basic symbols: "horse" and "stripes." The meaning of "horse" and that of "stripes" are derived from iconic and categorical representations.

Such a symbolic grounding process is a bottom-up, meaning-to-symbol process. Naturally, meanings from the external world could be transferred by iconic representations and categorical representations to symbolic representations. At the non-symbolic level, iconic and categorical representations are abstracted and produced by interaction between the world and the system. At the symbolic level, new symbols are formed by combining these non-symbolic categorical names.

To uncover the problem of this strategy, we should pay close attention to the problem posed earlier: How can the AI system understand? The answer is quite straightforward: It does not *understand* anything. Since we know that the parameters of neural networks can be edited and adjusted casually, and we still have to write a program from which we derive our intrinsic intentionality, the hybrid system itself does not possess intrinsic intentionality, hence no understanding. Harnad's framework in principle allows two options to tackle this issue. One way would be to redefine what "understand" is. For example, the behaviorist definition of understanding is that an AI can act like a human: the AI may see an apple,

send the image information to the processor, and then say the word "apple" by its sound generator. Alternatively, one could declare that the SGP is different from the Searle-style problem, say, the SGP is not relevant to the intrinsic intentionality, but this may make the SGP a superficial issue in which grounding is simply a causal problem.

Harnad adopted the first method, and proposed a hierarchy of Turing Tests, from toy functions (T1), which can only simulate fragments of human functions (i.e., visual functions), to total symbolic function (T2—the standard Turing Test), and to external sensorimotor (robotic) function (T3), and to T4 and T5 function, which are the simulations of our complete brain/body functions (the biological details such as how neurons work) and systems. An AI reaching the T5 level would be totally indistinguishable from a biological human being [7]. By defining these levels, his primary purpose is to defend that T3-level robots are immune to Searle's argument because they can interact with the world just like human-beings, so the SGP can be solved by creating a T3-level robot. For the T4 and T5 level, he believed they were overdetermined, hence not necessary.

However, it is hard to see how this method could succeed. As mentioned earlier, the term "understand" is closely related to intentionality, but the hierarchy of Turing Tests is per se a behaviorist method. And albeit extensive discussions about mind/body issues or other-minds problems, there is just no discussion of intentionality in the article [7]. If the T3-level robot can be immune to the "Chinese Room Argument", he must have adopted a behavioral definition of understanding rather than an intentional one. An extended conclusion following behaviorism is that Turing Tests should be regarded as scientific goals which cognitive scientists should pursue, and that T3 function is the appropriate target level for cognitive science. This can be seen as an outlook of future research, but the implicit idea is that the part of the SGP related to intentionality is unsolvable.

In conclusion, Harnad's strategy ended up falling back to behaviorism to avoid dealing with the Searle-style problem: How can the AI system understand? Judging from what he has published since then, it can be assumed that he is well aware that the intentionality problem is the key, but his attitude is rather pessimistic: as we have no means to be the robots themselves, we will never know if they are intentional or conscious. As a cognitive scientist, he came unsurprisingly to this conclusion, but from a philosophical point of view, there are multiple theories addressing intentionality. It is too early to give up.

2.2. Physical Symbol Grounding

Vogt and Steels proposed a new solution which they called physical symbol grounding [9,13], and which was compared to Brooks' physical grounding hypothesis [14], to indicate that a symbolic system is necessary and should rely on a physical basis. From a terminological perspective, this approach did not deliver too many new ideas, just as the hybrid system, the basic idea is that a symbolic system should be combined with a non-symbolic system, in their words, "they (robots) should be grounded by physical agents that interact with the real world" [13] (p. 435).

The Semiotic systems and guessing games are two essential concepts of physical symbol grounding. Semiotics (in accordance with Peirce) is a theory that focuses on sign processing, including the production, activity, and meaning of signs. Unlike symbols in a program, "semiosis" usually refers to symbols in a broader sense, including totems, signposts, and other things that have a referential nature.

According to Vogt, semiotics divides the definition of semiosis into three parts: (1) The formal element of semiosis. When semiosis is arbitrary and prescribed and depends only on the shape to distinguish itself, the form is the symbol itself. (2) The meaning element of semiosis. Meaning is the sense of the form, i.e., the interpretation of the symbol. (3) The object element of semiosis. This element is the object to which the semiosis refers. In semiotics, an instance of semiosis itself contains form, meaning, and object, so the symbol already contains meaning in its definition, a move which to some extent avoids the symbol grounding problem.

To apply semiotics theory to their practice, Vogt and Steels introduced the guessing game, which may be influenced by Wittgenstein's language games. The game involves at least two robots. One of the robots is the speaker that relates arbitrary symbols to the objects it observes. The other robot is the hearer who guesses which object corresponds to the word spoken by the speaker. The game aims to allow the hearer and speaker to agree on an "object-symbol" relationship, thus forming a public language. For example, in a color guessing game, a number of different color blocks are presented on the table. A robot randomly becomes the speaker, picks one of the blocks as the target, and defines it as belonging to a category. The other robot is the hearer that accepts the name spoken by the talker and guesses which color block the name indicates. The game is successful if the robots agree on the relationship between the color block and name.

According to Vogt and Steels, compared to a simple hybrid system, applying the semiotics system to the guessing game renders that the game is not a static system but an adaptive, dynamic, social, and evolutionary system. The major advantage of the game is that it does not depend on human interventions to form a public language, and the relations between symbols and objects are generated autonomously. If the language of the robots becomes mature, people less familiar with the experiment will not even be able to understand their language. This fact means that the robots can produce their own language without relying on humans. These features make the language of robots more human-like and the process of language generation more autonomous.

However, the advantages mentioned above are hardly tenable and semiotics may be misapplied to solve the SGP. Firstly, still, they didn't answer the Searle-style question: How can the AI system understand? What they focused on was how to solve the literal SGP problem: how can a symbolic system obtain its meanings? In addition, the term *meaning* is also defined inappropriately in their theories. They correctly found that meaning should be explained more clearly in the SGP, but they applied semiotic theory to the SGP in a rather unsuitable way.

The scope of semiotics is so large that all signs are the subjects under study and semiotic theories try to give a general explanation of them, but the SGP does not need a general theory of all signs. Signs are different from symbols. All meaningful things including words, images, sounds, gestures are signs, as long as they can be interpreted, but symbols, especially the symbols in the SGP refer in particular to the signs in AI's programs.

It may be possible to make a general theory applicable to a special problem, i.e., how a subset (symbols in AI) of signs can obtain its meaning, but at least Vogt's usage of the Peircean model of signs is problematic. In Peirce's semiotic triangle, a major feature is that the meaning is dynamic. He wrote that "the meaning of a sign is the sign it has to be translated to" and that "a sign is not a sign unless it translates itself into another sign in which it is more fully developed" [15] (p. 33). So, it is an ongoing process, the interpretant of a sign could be counted as a new sign, therefore producing a new triadic relation; since any sign requires an interpretant, an infinite chain would become necessary. Apparently, in the SGP, there is no such feature of infinite regress and conversely the meanings of symbols in the SGP should not be generated infinitely, otherwise it would cause the disjunction problem or trivialization problem analyzed in detail by Krystyna Bielecka [16].

Another important feature of Peirce's theory is that the interpretant is necessary for making meaning of signs. The meaning of a sign is not simply contained by the "representamen"—the sign vehicle, but is generated dynamically through the triadic relation in which the interpretant should be active. Though the term interpretant may not have a precise definition in Peirce's model, as he defined the interpretant as "something created in the Mind of the interpreter" or the "effect of the sign" [17] (p. 43), social community and convention can serve as the candidates [15] (p. 35). However, the SGP only requires the connection between symbols and their content without mentioning whether there should be an interpretant or not. Appealing to an interpretant can even be confusing for solving the SGP, since human-beings can be the interpretant to explain the meanings of symbols

in AI programs, and this brings Searle's argument back: If the meanings are parasitic on people's mind, AI cannot understand.

Based on this analysis we can conclude that Vogt and Steels' application of Peirce's semiotic theory is ineffective. Perhaps out of convenience, they have utilized only parts of the theory to support their theory, but it does not make a big difference for solving the SGP. The only advance of it is that they make the hybrid system work in a real robot successfully. Another way of applying Peirce's semiotic theory to the SGP is based on Terrence Deacon's model of symbol emergence [18]. Instead of focusing on the triadic relations, this model is mostly built on the classification of signs following Peirce: icons, indexes and symbols. The icons are the signs that share common properties with its referent, as a map resembles its territory. The indexes have causal or physical relationships with its object. For example, a weathercock is the index of the direction of the wind. The symbols are based on social convention or habit. In Deacon's theory, there is a hierarchic structure between these three forms of references: "symbolic relationships are composed of indexical relationships between sets of indices and indexical relationships are composed of iconic relationships between sets of icons" [19]. (p. 75)

According to the hierarchy, the grounding process in an AI system is conceptualized as a continuous process. The iconic relationships should be obtained at first followed by the indexical relationships and symbolic relationships. This method is similar in style to the hybrid system of Harnad, in which the iconic representations should be formed at first and then the categorical and symbolic representations. Though the definition of index and symbol in Deacon's theory are different from the categorical and symbolic representations in the hybrid system, the core idea is still to ground symbols through already grounded signs as other solvers did. The approach is coherent on the background of Deacon's theory on the evolution of language, but again solving the SGP needs to address the intentionality problem. Naturalizing intentionality based on Deacon's theory might be a possible way forward but needs further exploration.

2.3. Floridi and Taddeo's "Zero Semantical Commitment"

In 2005, Floridi and Taddeo summarized all existing SGP solutions at that time and claimed that none of those solutions truly allowed the AI to acquire semantics autonomously, thus proposing a "zero semantical commitment" to strengthen the requirements necessary to solve the SGP [8]: (1) Any form of innatism is disallowed. Semantic resources should not be presupposed in AI; (2) Any form of externalism is disallowed. Semantic resources should not be loaded into the AI from the external world. For example, the built-in preference or feature detectors of neural networks are usually set manually and installed in the AI as intrinsic semantic resources. Cases classified as instances of manually supervised instruction provide semantics to the AI from the outside, i.e., from external semantic resources. These semantic resources are disallowed.

The so-called semantic resources, irrelevant whether internal or external, could be seen as resources that people feed to AI to let them learn how to interact with the world and how to categorize the environment correctly. For instance, a neural network can be fed by a large image database to categorize the faces of men and women. The resources include people's experience of constructing a suitable neural network like how to tune the parameters of the network, which is internal to AI, and the labels of the database, which is external to it.

Floridi argued that thus far (as of 2005), no solutions meet the zero semantical commitment, including the hybrid system and physical symbol grounding. For the hybrid system, Floridi said:

> Unfortunately, the hybrid model does not satisfy the Z condition. The problem concerns the way in which the hybrid system is supposed to find the invariant features of its sensory projections that allow it to categorize and identify objects correctly . . . Neural networks can be used to find structures (if they exist) in the data space, such as patterns of data points. However, if they are supervised,

e.g., through back propagation, they are trained by means of a pre-selected training set and repeated feedback, so whatever grounding they can provide is entirely extrinsic. If they are unsupervised, then the networks implement training algorithms that do not use desired output data but rely only on input data to try to find structures in the data input space. Units in the same layer compete with each other to be activated. However, they still need to have built-in biases and feature-detectors in order to reach the desired output. [8] (p. 423)

In response to Vogt and Steels' physical symbol grounding, Floridi first criticized the semiotics scheme:

Suppose we have a set of finite strings of signs—e.g., 0s and 1s—elaborated by an AA. The strings may satisfy the semiotic definition—they may have a form, a meaning and a referent—only if they are interpreted by an AA that already has a semantics for that vocabulary. This was also Peirce's view. Signs are meaningful symbols only in the eyes of the interpreter. But the AA cannot be assumed to qualify as an interpreter without begging the question. Given that the semiotic definition of symbols is already semantically committed, it cannot provide a strategy for the solution of the SGP. [8] (p. 435)

Later, he criticized the guessing game:

Unfortunately, as Vogt himself acknowledges, the guess game cannot and indeed it is not meant to ground the symbols. The guess game assumes that the AAs manipulate previously grounded symbols, in order to show how two AAs can come to make explicit and share the same grounded vocabulary by means of an iterated process of communication. Using Harnad's example, multiplying the number of people who need to learn Chinese as their first language by using only a Chinese-Chinese dictionary does not make things any better. [8] (pp. 435–436)

Based on their zero semantical commitment, these arguments are consistent. The hybrid system did involve internal and external semantics and in Vogt's strategy the semantics have been presupposed without any explanation of its resources—they should clearly state what is the interpretant in the triadic relations. Floridi and Taddeo argued that these solutions all failed since they violated the zero semantical commitment, and therefore the way to solve the SGP is to develop an AI that does not have any semantic resources.

However, their commitment is not consistent with Harnad's interpretation of the SGP and inappropriately proposed a too strict condition for AI. The key mistake is that they confused the term "intrinsic" and "autonomous", and the latter is what they added based on Harnad's definition of the SGP:

Usually, the symbols constituting a symbolic system neither resemble nor are causally linked to their corresponding meanings. They are merely part of a formal, notational convention agreed upon by its users. One may then wonder whether an AA (or indeed a population of them) may ever be able to develop an autonomous, semantic capacity to connect its symbols with the environment in which the AA is embedded interactively. This is the SGP. [8] (p. 420)

In their explanations, autonomous AI just has access to its own semantic resources and goals without people's intervention such as feeding semantic resources to the AI. But in Harnad's interpretation, the term "intrinsic" refers to intrinsic intentionality as in Searle's argument. In a late publication, Harnad also concluded that consciousness or feelings which related to intrinsic intentionality are the unsolvable parts of the SGP [20]. But in some places, Taddeo and Floridi just equate "autonomous" and "intrinsic", arguing that the autonomous AI has intrinsic semantics. Replacing intrinsic intentionality with autonomous AI is just making a conceptual shift which does not contribute to solving the SGP.

Focusing on autonomous AI is also behavioristic, since researchers only focus on whether the AI has been trained by people and has the ability to interact with the world, without exploring whether machines can understand, feel, or experience. Their strategy

is equivalent to adding a strict condition to Harnad's T3-level robot, but even from a behavioristic perspective, this condition is not reasonable and almost impossible for AI. Education in human society is an obvious counter-example where a system can obtain meanings with external semantic resources. Like training an AI system, in our society, it is common that adults use databases with labels to train children. For example, parents usually use a booklet with pictures and categories of animals to help children learn to recognize them. Therefore, zero semantical commitment is not a necessary condition for solving the SGP or making AI understand their symbols.

Criticisms of the commitment have arisen since 2007, when Floridi and Taddeo proposed a "praxical strategy" in line with the zero semantical commitment [21]. The core idea of the strategy is to develop a robot without any goals so it can act in a random way at first. In their words, "the initial generation of meanings is teleologically free" [21] (p. 372). The corresponding internal states including the sensors and effectors' states induced by the random action are regarded as the semantical resources and then connected to a symbol system, so as to avoid external biases. But a robot without any purpose is meaningless since it does not have any functions. Unsurprisingly they introduced evolutional theory and Hebb's rule to make their strategy convincing. However, they just give a brief introduction of them and do not provide a clear explanation on how evolutionary theory can be teleologically free.

They have been following this misleading road so long that many people noticed the obvious mistake in their strategy. Among these criticisms, some argued that AI without any purpose is inappropriate. For example, Bielecka has attacked the praxical strategy by stating: "Just because there is no teleology assumed in their account of the agent's actions, the easy disjunction problem is unsolvable—actions are individuated just like responses in early behaviorism" [16] (p. 84). In addition to that, solutions to the SGP face the same problems as those faced by causal theories of reference, and Floridi's praxical strategy additionally suffers from a severe trivialization problem: if there is no purpose, all actions are meaningful, and anything can be represented. This inherent problem is closely related to the difficulty in the SGP since almost all strategies presuppose causal theories but do not mention anything about them. Bielecka's analysis could be regarded as a supplement for these strategies about causal theories and their problems, but still, neither did she discuss the relations between intrinsic intentionality and the SGP nor answer the question whether solving the SGP requires more than just causal theories.

Vincent Müller likewise argued that semantics and goals are in a binding relationship, and there were no semantics without goals. Semantic content requires normativity for using the symbol "correctly" or "incorrectly," and if there is no success or failure, the semantics are incomplete [22]. But the explanation in his article is unclear. It would be necessary to analyze the relationship. Why should semantics be with goals and what does the normativity mean? He may want to refer to teleological theories in which representations in a system are normative so that it could represent the content wrong or successful, but it should be clearly stated. And the term *goal* may be understood as a proper function in teleological theories, which are what they were selected for, determined by the history of evolution.

This kind of criticism is appropriate, for without functions or teleology, AI cannot even seem to be intelligent, not to mention understand symbols. Though Floridi and Taddeo have borrowed evolutionary theory to rationalize it, a suitable environment for evolution must be constructed by people, which introduces the external semantical resources again, according to the zero semantical commitment.

Again others, who may go astray, start to invoke the consciousness problem. Müller is the first to suggest explicitly that similar to the problem of consciousness, there are "hard" and "easy" problems within the SGP [23]. Since then, scholars from various schools of disciplines have joined the discussion.

3. The Problem of Consciousness in the Symbol Grounding Problem

Early studies on the SGP focused on the design of engineering solutions such as hybrid systems, physical symbol grounding, and physical grounding. These solutions focused on how to solve the practical problem rather than on the philosophical presuppositions behind the problem or how to solve it philosophically. Since Floridi and Taddeo and others proposed the zero semantical commitment, criticisms and philosophical discussions followed. We have analyzed the first strand related to purposeless robots, and the other strand of debate relating to the question of whether the SGP involves the problems of consciousness. Müller [23] was the first to explore this issue, followed by Dairon Rodríguez [24] and Richard Cubek [25]. However, in fact, as early as 1993, Harnad had already realized the "unsolvable part" of the SGP.

It is hard to say that they were wrong in introducing the consciousness problem to the SGP, but dividing the SGP into hard/easy parts is pessimistic to philosophical research since all things that cannot be examined by existing scientific or philosophical theories could be included in the hard part, and therefore be declared to be unsolvable. We will first review why and how they approached the problem and then present a more optimistic way of reconciling it.

3.1. Harnad's Paradox

At the beginning, Harnad noticed that Searle's thought experiment is a closed system where symbols can only be related to other symbols. What he did was introducing something other than symbols to the system so that symbols can be grounded by it resulting in the hybrid system. In the hybrid system, symbols can indeed relate to meanings, but the key problem in Searle's argument is how the AI system can understand, or in other words, how can AI have intrinsic intentionality?

If the SGP is just a functional problem, the hybrid system is a perfect answer, but Harnad explicitly used the term "intrinsic meaning", which refers to Searle's intrinsic intentionality, and it is well-known that in Searle's publications, intrinsic intentionality is a kind of conscious intentionality. For example, in his article *Consciousness, explanatory inversion, and cognitive science*, Searle writes:

> Cognitive science typically postulates unconscious mental phenomena, computational or otherwise, to explain cognitive capacities. The mental phenomena in question are supposed to be inaccessible in principle to consciousness. I try to show that this is a mistake, because all unconscious intentionality must be accessible in principle to consciousness; we have no notion of intrinsic intentionality except in terms of its accessibility to consciousness. [26] (p. 585)

And if the intrinsic intentionality problem is an essential problem of the SGP, then strategies should deliver an explanation, but Harnad might misunderstand or just skip the discussion of it. The hybrid system is obviously a functional system as Harnad himself also called his method "robotic functionalism", but in his article, it seemed that the functional strategy can handle an intrinsic phenomenon:

> If both tests are passed, then the semantic interpretation of its symbols is "fixed" by the behavioral capacity of the dedicated symbol system . . . the symbol meanings are accordingly not just parasitic on the meanings in the head of the interpreter, but intrinsic to the dedicated symbol system itself. [4] (p. 345)

This leads immediately to the following difficulty: how can a functional system give an explanation for something intrinsic or something conscious? If Chalmers was right, the hard problem is hard because there is an explanatory gap between functions and experience. Therefore, if we follow Searle's division, there is also a gap between the functions of obtaining meanings and the experience of it, and then the hybrid system only solved the functional problem which is easy.

In his later publications, Harnad's attitude also becomes contradictory:

> The problem of meaning is in turn related to the problem of consciousness, or how it is that mental states are meaningful . . . But whether its symbols would have meaning rather than just grounding is something that even the robotic Turing Test—hence cognitive science itself—cannot determine, or explain. [27]

And we can see a lot of attitude swings elsewhere, in the article *Symbol grounding is an empirical problem: Neural nets are just a candidate component*, he even proposed a contradictory proposition that a grounded symbol system might have no intrinsic meanings:

> It is logically possible that an ungrounded symbol system has intrinsic meanings or that a grounded symbol system fails to have them. I'm merely betting (probabilistically, but with reasons) that T3-capacity is sufficient for having a mind and meaning. [28]

And not surprisingly, Harnad finally explicitly supported a division of hard/easy problem of the SGP in the article *Alan Turing and the "Hard" and "Easy" Problem of Cognition: Doing and Feeling*:

> Sensory-motor robotic capacities are necessary to ground some, at least, of the model's words, in what the robot can do with the things in the world that the words are about. But even grounding is not enough to guarantee that—nor to explain how and why–the model feels (if it does). That problem is much harder to solve (and perhaps insoluble). [20]

Harnad argued that the SGP should be distinguished from Searle's argument and regarded as an independent problem, but he did not give a valid position of its independence. If the intentionality problem is the key, then we can just treat the SGP as Searle's problem, and this would render it irrelevant to propose the SGP independently. But if the SGP is just a functional problem, it should just be researched by scientists and engineers, i.e., how can a system have the function of obtaining meanings, which is not a philosophical problem. The demand for independence puts the SGP in an awkward position, so the result of compromise is simply breaking it up into two parts.

A more optimistic way is to adjust the premise of the SGP. For example, instead of asking how can an AI system have intrinsic intentionality, we can ask the more general problem of how can an AI system have intentionality? The intrinsic intentionality problem needs a theory to explain consciousness but the answers to the general problem are pluralistic such as causal theories and teleological theories. This will keep the SGP independent because there is no need for a special theory to frame intrinsic intentionality and it is not simply a functional problem because the intentionality of AI still requires a more complete philosophical theory to explain. However, after Harnad, many other scholars just took the division of intentionality as granted, ending up putting the SGP into two parts either. We will have a brief review of their arguments and then give the details of the third road of the SGP.

3.2. The Arguments of Others

Harnad is representative for scholars who struggled with the consciousness problem which justifies the separate introduction. In addition to him, Paul Davidsson [29], Taddeo and Floridi [8], Müller [23], Cubek [25], Bringsjord [10], and others have all basically adopted the division of intentionality and referred to the problem of consciousness. It's worth noting that most of them major in computer science rather than philosophy. This may be one of the reasons why they did not engage in a deeper analysis of the intentionality problem. Examples are analyzed in the following.

Though Davidsson and Taddeo & Floridi did not explicitly segmentalize the SGP into two parts, as they followed Harnad's definition of the SGP and used the term *intrinsic*, incoherent arguments and contradictions will inevitably arise.

In the article *Toward a general solution to the symbol grounding problem: combining machine learning and computer vision*, Davidsson repeated Harnad's explanation of the SGP:

"The problem of concern is that the interpretations are made by the mind of an external interpreter rather than being intrinsic to the symbol manipulating system" [29] (p. 157).

By arguing interpretations should be intrinsic to a symbol system, his strategy failed to show how to generate intrinsic interpretations. For making the system learn to generate meanings by itself, he suggested two typical types of Machine Learning: learning by observation (supervised learning) and learning from examples (unsupervised learning). Despite the details, the main idea is still functional and entails the construction of a mechanism that enables AI to interact with the world and recognize or categorize the environment correctly, which is irrelevant regarding the question on how to obtain intrinsic intentionality.

And as mentioned earlier, Taddeo & Floridi employ similar statements: "This means that, as Harnad rightly emphasizes, 'the interpretation of the symbols must be intrinsic to the symbol system itself, it cannot be extrinsic, that is, parasitic on the fact that the symbols have meaning for, or are provided by, an interpreter" [8] (p. 5).

Rather than appealing to consciousness, the result of their struggling is the zero semantical commitment. But this is a more serious conceptual confusion. They argued that autonomous AI could obtain meaning without people's intervention and proposed a praxical strategy to realize it so as to solve the SGP. But similarly, the autonomous AI in their strategy is merely a functional concept which is even worse for not only does it not solve the intentionality problem, but it creates other problems about how can a purposeless AI be intelligent, which was widely criticized.

In recent years, more authors have become aware of the consciousness and intentionality issues and have articulated them more bluntly. After first proposing the division of the SGP in 2011, Müller followed up by writing the following in his 2015 article *Which Symbol Grounding Problem Should We Try to Solve?* [22]. He divided the hard and the easy problem again, in which the hard problem is "Why and how does physics give rise to conscious experience (to phenomenal consciousness, to 'what it is like')?", and the easy one is: "explanation of cognitive abilities and functions" of awareness (the ability to discriminate, integrate information, report mental states, focus attention, etc.)—computational or neural mechanisms [22].

Cubek echoes Müller's sentiments and writes in the article *A critical review on the symbol grounding problem as an issue of autonomous agents*:

> Several solutions have then been proposed, with a very promising one by Steels claiming that none of these really solved Harnad's problem. Taddeo and Floridi introduced the Z condition—concretizing the SGP. Finally, Müller and Fields showed that it is unsolvable, and that it can be delegated to the hard problem of consciousness. [25] (p. 260)

Christophe Menant has proposed an approach called Meaning Generator System to describe how elementary life generates meanings [30]. In this model, a system can generate meaning depending on the interaction between the environment and the internal constraints of the system. For example, a paramecium can move away from acid water based on the interaction between the hostile environment and the internal constraint "stay alive" and generate the meaning "presence of acid not compatible with the 'stay alive' constraint" [30]. However, this pattern cannot be directly applied to AI systems, for the nature of human mind as the root of constraints is not fully understood so far. Therefore the SGP is still unsolvable. Though Menant didn't refer to the consciousness problem explicitly, I believe the "unknown nature of human mind" he mentioned also contains the unknown nature of consciousness.

We can find similar arguments in Bringsjord and Rodríguez's articles; though these comments differ in details, the main idea remains the same.

These views would have adverse effect on the research of the SGP. The Consciousness problem is a frame everything can be put in. Claiming the SGP has a functional part and an experiential part is to say nothing of it, but if we ask how can an AI system have intentionality, we still have a lot of work to do: the first step is to find an appropriate theory which can be applied to AI's intentionality because most theories of mental content

focus on people or organism's mental states; the second step is to explore what kind of AI research can be coherent with the theory.

4. The Denial of Intrinsic Intentionality and the New Direction of SGP

Echoing the division of hard/easy problem of the SGP, people who argued that their strategy could solve the SGP is essentially solving the functional part of it and people who thought that the SGP is to remain unsolved are considering the consciousness part of it. Vogt and Steels are confident about their strategy from the functional point of view, for it is a very well-organized realization of a hybrid system, therefore they have the slogan "The SGP has been solved . . . " [13]. Bringsjord did not explicitly refer to consciousness but argued that the "Chinese Room Argument" was so sound and the strategies of the SGP are just "physicalized symbol systems" [10], that it is unsolvable and titled their paper *the Symbol Grounding Problem . . . Remains Unsolved*. Similarly, Bielecka has discussed the difference between the SGP which could be solved by using already grounded symbols and the real SGP she called the non-derivative grounding, by which the symbol system can have intrinsic meaning [31].

Now the paradox is clear: If the SGP is just a functional problem, it should be researched by Scientists. If it is related to consciousness, it is an unsolvable philosophical problem. Those people who proposed strategies and argued that through these strategies they can make their AI obtain intrinsic meaning were essentially arguing that a functional strategy can solve the consciousness problem, which is contradictory and needs a lot of explanation. As for people who supported the division of intentionality, they just make the SGP trivial: the functional part is easy to solve and the conscious part is unsolvable.

As the division is the root of the paradox, breaking out of such a situation would mean rejecting the division, and some people have already been doing this. As early as 1987, Daniel Dennett denied the division between intrinsic intentionality and derivative intentionality, arguing that all intentional endowments are instrumental and that their usefulness lies in predicting the behavior of people or animals [32]. Thus, all intentionality is derivative intentionality, and intentional endowments do not have an intrinsic nature. Dennett also argued that Searle confused intentionality with the consciousness of intentionality: "Searle has apparently confused a claim about the underivability of semantics from syntax with a claim about the underivability of the consciousness of semantics from syntax. For Searle, the idea of genuine understanding, genuine "semanticity" as he often calls it, is inextricable from the idea of consciousness [32] (p. 336)". Paying attention to consciousness forces us to think from a first-person perspective.

And the way of avoiding getting into the trap of the division of intentionality is to introduce a general problem of intentionality, that is, how can an AI system obtain intentionality? As we have a lot of philosophical theories such as causal theories and teleological theories at hand, and if these can be used to solve the SGP, there is no special need to propose a consciousness theory.

4.1. The Theories of Naturalizing Intentionality

Intentionality is a feature that mind displays to be about or represent its object. The contemporary discussions of intentionality were launched by Brentano and his famous, perhaps also the most controversial thesis is that "intentional inexistence is characteristic exclusively of mental phenomena [33] (p. 68)". So, if we accept this conclusion, then physical objects like an AI system can never have intentionality and the intentionality as mental phenomenon cannot be explained by physicalism. This kind of explanation of intentionality is actually the same as Searle's intrinsic intentionality, so in fact the difficulty in the SGP is that the intentionality of AI is constrained in mental phenomena.

However, some argued that intentionality can be exhibited by non-mental objects and believed that there is a naturalized way to give the explanation, which could be called "the naturalization of intentionality". The theories of naturalizing intentionality should be considered as the main resources of solving the SGP since it does not presuppose that there

should be a division of intrinsic/derived intentionality. The two main strategies are causal theories and teleological theories and it should be answered how these theories could be applied to AI.

The basic idea of causal theory can be illustrated as what Fodor called "the crude causal theory": It is metaphysically necessary that if tokens of F are caused by and only by instances of the property G, then F refers to G [5]. But the obvious problem is that F can be caused not only by G, but also by H, J, . . . so the content could be disjunctive and F could refer to "G or H or J . . . .", which is called "the disjunction problem". And in most situations, when a belief is true, there is a correctly corresponding relation between F and one of these disjunctions, with others being false, so the possibility of making mistakes is often referred as the misrepresentation problem, or the problem of error.

Dretske has developed an informational version of causal theory that a representation R carries information of the property G so that R refers to G [6]. The informational relation is a kind of nomological relation which means the cause is counterfactually necessary for the effect. For example, to say that the fire caused the smoke is to say that there cannot be any smoke unless there is a fire. But the disjunction problem still persists, for if G also carries the information H, and H carries the information J, then R would still refer to G or H or J.

Fodor and Dretske both proposed theories to tackle the problem, however not successfully. This is a fatal weakness, for in our daily lives nearly everyone makes mistakes. An appropriate theory of mental content should allow for errors and explain the normativity of mental states.

In contrast, the main advantage of teleological theories is that they address the problem to a large extent. Dretske, Ruth Millikan [34], and Karen Neander [35] have all proposed their own theories. Although these theories are different in many aspects, the basic and common idea is that the normativity of semantics derives from biological functions. The term function in teleological sense is not the causal role it plays in a system as a part of it, but what it was selected to do. For example, the heart is selected to pump the blood, but not to generate the sound of heartbeat, which is also the causal effect of the organ. So just as the biological function can disentangle the "correct" relations from the multifarious causes and effects, the correct relations between representations and content can also be selected by the history of evolution. For instance, the waggle dance of honeybees can correctly refer to the direction and distance of flowers because the relations which can promote the fitness of the honeybee population are selected.

Nevertheless, these theories mostly focus on people or other organisms with a history of evolution. To what extent they are applicable to the SGP needs to be illustrated. As AI systems are essentially causal systems and nearly every part of the system can have a functional explanation in a causal sense, it is reasonable to apply causal theories. But just as Bielecka pointed out [16], AI systems also faced the disjunction problem and this could be found in modern AI technology: face recognition systems always have a very small possibility to misrecognize someone's face.

To address the problem, it seems the teleological theories are the only hope. But there are some difficulties we need to solve. The first is to choose an appropriate theory. And the second, perhaps the more important, is to equip AI with a history of evolution. The good news is that we already have a research area of AI called "evolutionary robotics", which may make it possible for AI to have an evolutionary history under laboratory conditions. We will first investigate this research field and then explore which teleological theory can be applied to it.

4.2. The Evolutionary Robotics Research

The core theory in evolutionary robotics is evolutionary algorithms which existed since the research of John Holland in the 1960s [36]. The basic idea of evolutionary algorithms is simple: first randomly generate the population of individuals, then evaluate the performance of each one in the population and select the individuals with the best performance,

and then use the crossover and mutation operations of their "genes" or "chromosomes" to generate a new population. The last step is repeating these steps until the performance stops getting better, in other words, finding the optimal solution of the problem.

The theory can be applied to a wide range of scenarios such as solving mathematical problems, optimizing the parameters of neural networks, generating computer programs, and so on. Evolutionary robotics can be regarded as the application of evolutionary algorithms to robotic research. We focus on evolutionary robotics because for the SGP, symbols must be connected to something other than just other symbols. So, the ability of interacting with the world is necessary. Even though neural networks or computer programs can have an evolutionary history, they are still meaningless because the symbols do not refer to anything else.

The typical example suitable for the discussion of the SGP is the experiment Dario Floreano and Francesco Mondada did in 1996 [37]. The testing ground for their robots is an open platform with a black fan area as the charging zone at the top left corner and a tiny light on the same position to illuminate the area. The robots, without any knowledge about the platform, only have 60 s to live so they need to learn to stay at the charging zone for a while to obtain their time. The evolutionary mission is to move as much as possible, so they have evolved the basic strategy to move around the ground until they only have around two seconds life and then go back to the charging area.

Now there is an important relation between the representation R and the content C, "the charging area". The vehicle of the representation could be the symbols in the robots' symbol system which also control the behaviors of the robots, so when it's running low in power, the representation must be activated to guide the correct actions. The problem is what is the actual content of R, as the tiny light is on the same position as the charging area, so it can refer to the light, or the charging area, and the disjunction problem reappears.

If we follow traditional explanations such as Millikan's theory, the content should be the charging area since it is what really promotes the fitness of the population. But they also did some successive experiments which conflict with our intuitions. In order to understand how the robots can recognize the charging area, they changed the position of the light to the top right corner, and it turned out that the robots just follow the light and go to the new position rather than the top left area. The result showed that the robots relied on the light to learn the position of the charging area.

To address the conflicts, we could invoke Neander's teleological theory that she called "informational teleosemantics", which is fully presented in her book *A Mark of the Mental: In Defense of Informational Teleosemantics* [35]. The core idea is that the natural selection is not only applied to the effects of the system (referring to charging area make they survive), but also to the causes (the tiny light) of representation, and the latter is the appropriate content. She used the term "response function" to explain how the sensory-perceptual systems are selected to respond to something and generate the corresponding states. In the example here, we can say that the robots are selected to respond to the tiny light so that they can survive. And as the explanation is coherent with mainstream cognitive science and the causal and informational explanations of mental content, it is better to argue that the content is "tiny light" rather than "the charging area", therefore the disjunction problem could be solved.

Now we get back to the SGP. As mentioned earlier, we treated the SGP as a general problem of intentionality: how can an AI system have intentionality? We first reviewed the two main families of theories of naturalizing intentionality, and the conclusion is that we need an AI system with a history of evolution to apply the teleological theories. Then we introduced the evolutionary robotics to demonstrate how it works. Though many theories are available, perhaps Neander's informational teleosemantics is the most compatible theory with evolutionary robotics. So, the solution of the SGP could be as follows: through interacting with the world, the meanings could be transferred to the symbol system by its sensory and motor system, and through the evolution of the embodied AI system, the meanings could be fixed with the related symbols.

Naturally, there are a lot of details and new problems that need to be further discussed. The first is still the disjunction problem. Though Neander's theory has eliminated many possible disjunctive contents, still some are hard to distinguish. For example, being triangular and being trilateral are co-instantiated properties of a triangle and may cause the same representation, but they are obviously distinct. How to resolve such a problem is still inconclusive. Perhaps a more difficult problem for AI is to explain the content of the abstract concepts that cannot have impact on fitness such as democracy, quarks, and justice. But this is not a special problem of teleological theories: no theories of mental content can have a prefect explanation of it. The third problem may be the evolution of robots under laboratory conditions, in which the intentionality of the robots is not obtained by nature, but derived from human-beings. One reply to this is that the laboratory condition is temporary, theoretically we could design robots which can live in the real world and evolve. For example, we could construct a robot recharged by solar energy and the primary task for it is to evolve the function towards the sun.

The purpose of this paper is to give a new direction in solving the SGP, the strategy proposed here is a frame to show how it is possible to solve the problem of how can an AI system have intentionality and the theories of naturalizing intentionality present a promising answer. But if we continue to follow the division of intentionality from Searle, it is hard to see how to continue the research of the SGP.

5. Conclusions

This paper analyzes leading solutions to the SGP and the philosophical debates concerning the problem of consciousness triggered by Floridi and Taddeo. The problem of consciousness implied in the SGP can be traced back to the division between intrinsic intentionality and derivative intentionality in the "Chinese Room Argument". Since intrinsic intentionality depends on a subjective conscious experience, as long as the division is maintained, the problem of consciousness in the symbol grounding problem is inevitable. The solution proposed here is to deny intrinsic intentionality and to treat the SGP as a general problem of AI's intentionality. We combined Neander's informational teleosemantics with evolutionary robotics to open a new direction of answering the question, but it is still worthwhile to investigate further on how to make better use of naturalizing intentionality theory.

References

1. Searle, J. Minds, Brains, and Programs. *Behav. Brain Sci.* **1980**, *3*, 417–424. [CrossRef]
2. Searle, J. Why Dualism (and Materialism) Fail to Account for Consciousness. In *Questioning Nineteenth Century Assumptions about Knowledge*; Lee, R., Ed.; SUNY Press: New York, NY, USA, 2010; Volume III, pp. 5–48.
3. Searle, J. Minds and Brains without Programs. In *Mindwaves*; Basil Blackwell: Oxford, UK, 1987; pp. 209–223.
4. Harnad, S. The Symbol Grounding Problem. *Phys. D* **1990**, *42*, 335–346. [CrossRef]
5. Fodor, J. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*; MIT Press: Cambridge, MA, USA, 1987.

6.   Dretske, F. *Knowledge and the Flow of Information*; MIT Press: Cambridge, MA, USA, 1981.

7.   Harnad, S. Minds, Machines and Searle. *J. Exp. Theor. Artif. Intell.* **1989**, *1*, 5–25. [CrossRef]

8.   Taddeo, M.; Floridi, L. Solving the Symbol Grounding Problem: A Critical Review of Fifteen Years of Research. *J. Exp. Theor. Artif. Intell.* **2005**, *17*, 419–445. [CrossRef]

9.   Steels, L. The Symbol Grounding Problem Has Been Solved, so What's Next? In *Symbols and Embodiment Debates on Meaning and Cognition*; Oxford University Press: New York, NY, USA, 2008; pp. 223–244. [CrossRef]

10.   Bringsjord, S. The Symbol Grounding Problem Remains Unsolved. *J. Exp. Theor. Artif. Intell.* **2015**, *27*, 63–72. [CrossRef]

11.   Chalmers, D. Facing up to the Problem of Consciousness. *J. Conscious. Stud.* **1995**, *2*, 200–219.

12.   Harnad, S. Doing, Feeling, Meaning and Explaining. *On the Human*. 2011. Available online: https://eprints.soton.ac.uk/272243/ (accessed on 24 September 2022).

13.   Vogt, P. The Physical Symbol Grounding Problem. *Cogn. Syst. Res.* **2002**, *3*, 429–457. [CrossRef]

14.   Brooks, R.A. Elephants Don't Play Chess. *Robot. Auton. Syst.* **1990**, *6*, 3–15. [CrossRef]

15.   Chandler, D. *Semiotics: The Basics*, 3rd ed.; Routledge: New York, NY, USA, 2017.

16.   Bielecka, K. Symbol Grounding Problem and Causal Theory of Reference. *New Ideas Psychol.* **2016**, *40*, 77–85. [CrossRef]

17.   Nöth, W. *Handbook of Semiotics*; Indiana University Press: Bloomington, IN, USA, 1990. [CrossRef]

18.   Raczaszek-Leonardi, J.; Deacon, T. Ungrounding Symbols in Language Development: Implications for Modeling Emergent Symbolic Communication in Artificial Systems. In Proceedings of the 2018 Joint IEEE 8th International Conference on Development and Learning and Epigenetic Robotics, ICDL-EpiRob, Tokyo, Japan, 17–20 September 2018; pp. 232–237. [CrossRef]

19.   Deacon, T. *The Symbolic Species*; W.W. Norton: New York, NY, USA, 1997.

20.   Harnad, S. Alan Turing and the "Hard" and "Easy" Problem of Cognition: Doing and Feeling. *Turing100: Essays in Honour of Centenary Turing Year 2012*. Available online: https://arxiv.org/abs/1206.3658 (accessed on 24 September 2022).

21.   Taddeo, M.; Floridi, L. A Praxical Solution of the Symbol Grounding Problem. *Minds Mach.* **2007**, *17*, 369–389. [CrossRef]

22.   Müller, V. Which Symbol Grounding Problem Should We Try to Solve? *J. Exp. Theor. Artif. Intell.* **2015**, *27*, 73–78. [CrossRef]

23.   Müller, V. The Hard and Easy Grounding Problems. *Int. J. Signs Semiot. Syst.* **2011**, *1*, 70–73.

24.   Rodríguez, D.; Hermosillo, J.; Lara, B. Meaning in Artificial Agents: The Symbol Grounding Problem Revisited. *Minds Mach.* **2012**, *22*, 25–34. [CrossRef]

25.   Cubek, R.; Ertel, W.; Palm, G. A Critical Review on the Symbol Grounding Problem as an Issue of Autonomous Agents. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 256–263. [CrossRef]

26.   Searle, J. Consciousness, Explanatory Inversion, and Cognitive Science. *Behav. Brain Sci.* **1990**, *13*, 585–596. [CrossRef]

27.   Harnad, S. Symbol Grounding Problem. *Scholarpedia* **2007**, *2*, 2373. [CrossRef]

28.   Harnad, S. Symbol Grounding Is an Empirical Problem: Neural Nets Are just a Candidate Component. 1993. Available online: http://cogprints.org/1588/1/harnad93.cogsci.html (accessed on 24 September 2022).

29.   Davidson, P. Toward a General Solution to the Symbol Grounding Problem: Combining Machine Learning and Computer Vision. In Proceedings of the AAAI Fall Symposium Series, Machine Learning in Computer Vision: What, Why and How, Lund, Sweden, 22–24 October 1993; pp. 157–161.

30.   Menant, C. Turing Test, Chinese Room Argument, Symbol Grounding Problem. Meanings in Artificial Agents. *Am. Philos. Assoc. Newsl. Philos. Comput.* **2013**, *13*, 30–34.

31.   Bielecka, K. Why Taddeo and Floridi Did Not Solve the Symbol Grounding Problem. *J. Exp. Theor. Artif. Intell.* **2015**, *27*, 138. [CrossRef]

32.   Dennett, D. *The Intentional Stance*; MIT Press: Cambridge, MA, USA, 1987.

33.   Brentano, F. *Psychology from an Empirical Standpoint*; Routledge: London, UK, 2012.

34.   Millikan, R. *Varieties of Meaning*; MIT Press: Hong Kong, China, 2004. [CrossRef]

35.   Neander, K. *A Mark of the Mental: In Defense of Informational Teleosemantics*; The MIT Press: Cambridge, MA, USA, 2017. [CrossRef]

36.   Holland, J. Outline for a Logical Theory of Adaptive Systems. *J. ACM* **1962**, *9*, 297–314. [CrossRef]

37.   Floreano, D.; Mondada, F. Evolution of Homing Navigation in a Real Mobile Robot. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **1996**, *26*, 396–407. [CrossRef] [PubMed]