New Ideas in Psychology 40 (2016) 77–85

Contents lists available at ScienceDirect

New Ideas in Psychology

journal homepage: www.elsevier.com/locate/newideapsych

Symbol Grounding Problem and causal theory of reference

Krystyna Bielecka

University of Warsaw, Institute of Philosophy, ul. Krakowskie Przedmieście 3, 00-927 Warszawa, Poland

ARTICLE INFO

Article history:
Available online 7 February 2015

Keywords:
Symbol grounding
Representation
Function
Disjunction problem
Misrepresentation

ABSTRACT

In this paper, I juxtapose the Symbol Grounding Problem and causal theories of reference. In the first part of the paper, I show some basic assumptions they share in order to show, in the second part, some difficulties implied by these assumptions. These difficulties are: the meaning determination problem, the easy and hard disjunction problem, and the trivialization problem. My diagnosis is that both the easy and hard disjunction problem result from a more general difficulty with causal theories and the SGP solution, which is the possibility of misrepresenting, and in particular of accounting for system-detectable error. I emphasize some implications they have for the notion of representation. Finally, I enumerate some theoretical desiderata for a satisfactory account of naturalized semantics (and solutions to SGP) that would be free of the problems mentioned above.

Most attempts to solve the Symbol Grounding Problem (SGP) and causal theories of reference confront the question of the origins of meaning. As a consequence, some of their assumptions and arguments are similar. Each is therefore vulnerable to the objections that (1) representation, as these accounts define it, is trivially ubiquitous, and (2) they cannot account for misrepresentation or representational error. They also run into trouble when specifying representational content—relationships that are supposed to determine content leave it underspecified.

The first objection is related to the trivialization problem (Ramsey, 2007) and the second to the need for system-detectable error (Bickhard, 1993). The trivialization problem appears when any entity whatsoever can be treated as a representation. As a result, there is no reason to use the notion of representation because it does not play any specific explanatory role.

A causal relation, as presupposed by causal theories of content or by attempts to solve the SGP, gives rise to the problem of content determination. Such a relation is too weak to determine content because anything can be causally connected to anything. Something should be added, in order to handle the trivialization problem.

Any theory that is able to account for system-detectable error is thereby a solution to the trivialization problem. The idea is that only cognitive systems able to detect their own mistakes (system-detectable errors) qualify as systems that truly represent, because only in these cases does the notion of representation demand that a system distinguish an adequate representation from an inadequate one. There are persuasive arguments that, without a proper account of misrepresentation, meaning cannot be naturalized (Bickhard, 1993; Dretske, 1986).

In this paper, I briefly introduce the SGP, illustrated by a recent proposal to solve it in a 'praxical' way by Taddeo and Floridi. Then I sketch the structure of a causal theory of reference, and distinguish its two main variants. I indicate common motivations and assumptions in solutions to SGP and in causal theories of reference. Last—but not least—, I accentuate the difficulties that such theories face, and enumerate the number of desiderata that they should satisfy.

E-mail address: kikab@gazeta.pl.

http://dx.doi.org/10.1016/j.newideapsych.2015.01.006
0732-118X/

1. The Symbol Grounding Problem

The SGP can be traced back to the Chinese Room thought experiment (Searle, 1980). The Chinese Room may shed light on various philosophical issues, such as consciousness, the nature of understanding, or the status of artificial cognitive systems. In my view, what is most important is the fact that it started a debate over representation in AI by denying that an artificial system may possess intentionality in a non-derived way. (For my purposes, it is irrelevant whether Searle is right. There are arguments that the distinction between derived and non-derived intentionality is wrong (Dennett, 1987), but my aim here is merely to describe some solutions to the SGP no matter whether the original problem is genuine or worth solving.) Searle claims that only biological cognitive systems have original (non-derived) intentionality, and they are the only systems that ascribe meaning to entities with derived intentionality. All artificial systems (books, computers, and robots) have only derived intentionality. So it was natural for philosophers of AI to start reflecting upon the possibility of grounding symbols in non-biological cognitive systems. Would grounding make them similar to biological systems and warrant the view that they have original intentionality, too?

One version of this question can be found in the Symbol Grounding Problem (SGP), exposed first by Stevan Harnad (Harnad, 1990). Harnad has developed the Searlian experiment further in two versions: in the first version, Harnad describes a process of learning Chinese as a first language from a Chinese–Chinese dictionary, while in the second version, he focuses on the process of learning Chinese as a second language (this, however, he considers much easier to solve). In his exposure of the problem, symbols have to be grounded in something other than meaningless symbols. Other symbols do not constitute meaning, which is to be related to the outside world. The problem then is not about the possibility of translating the form of language L1 to the form of language L2; it is rather about the very possibility of the correspondence of a linguistic form to reality. Thus, the real Symbol Grounding Problem is indeed similar to the question of how symbols refer.

The SGP is about the theoretical and practical possibility of building an artificial cognitive system (or artificial agent) that has the autonomous semantic ability to associate symbols with their referents in the world. Unfortunately, the way Harnad formulated the SGP was not stated precisely enough to cut much confusion from the core of the problem, which concerns the semantic autonomy of an artificial cognitive system. For example, there are philosophers and engineers satisfied with semi-autonomous systems that are able to ground symbols with the indispensable help of other artificial systems or humans (which is called Symbolic Theft) (Cangelosi & Harnad, 2002). Their interpretation of autonomy allows them to use the already-grounded symbols, and the problem they solve is similar to the SGP (but not *exactly* the same). However, in Harnad's exposure of the problem, symbols have to be grounded in something other than meaningless symbols. Other symbols do not constitute meaning, which is to be related to the outside world. One formulation of the so called non-derivative SGP was proposed by Taddeo and Floridi:

> 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?
>
> Taddeo & Floridi, 2007, p. 370.

It is important to note what SGP is not. The problem can be easily confused with the question of how human beings learn language. Notice however that to learn a language, children obviously rely on the instruction of their parents and the functioning of a linguistic community. SGP is rather a question of how the meaning of linguistic tokens is established by fundamental, non-semantic facts, or how representation emerges from non-representational facts. This is the same question that theories of naturalized semantics face, but not necessarily the one that is answered by theories of evolution of natural language.

Taddeo and Floridi also spelled out a basic principle, called the Zero Semantic Commitment Condition (ZSCC). The principle constrains possible solutions of the SGP by excluding assumptions that any satisfactory theory of non-derivative symbol grounding should avoid. They state it as follows:

(ZSCC) No valid solution of the SGP can rely on forms of

(a) innatism, since no semantic resources (some *virtus semantica*) should be presupposed as already pre-installed in the AA [=artificial agent]; and

(b) externalism, since no semantic resources should be uploaded from the "outside" by some *deus ex machina* that is already semantically-proficient. (Taddeo & Floridi, 2007, p. 370).

Solving the non-derivative SGP in accordance with the ZSCC is difficult because the ZSCC disallows any semantic instruction from the outside (this is how Taddeo and Floridi understand 'externalism' in the quotation above; it is not to be confused with externalism in naturalized semantics, such as Millikan's teleosemantics), and excludes nativism or innatism (at least on the first glance). As a consequence, any source of symbolic interpretation other than the cognitive system itself should not intervene in the process of grounding. Other resources, such as syntax, computational abilities, perceptual, procedural abilities, are allowed, but only if they are not semantic.

The ZSCC can be applied to any cognitive system that has an ability to communicate, even if it is a distributed system. Any naturalistic semantics should explain how symbols gain their meanings that are not derived from other meanings. It is worth noticing that the question of the meaning of arbitrary symbols does not appear in the case of communication in a population of simple animals such as bees. In such a case, there is only a question of meaning of natural signs. For a naturalized semantics, the conventional meaning is often described as being derived from natural meaning and conventions are treated as effects of the learning process and evolution (Skyrms, 2010). However, such a naturalistic theory should not be seen as violating