



**KRYSTYNA BIELECKA**  
UNIVERSITY OF WARSAW

## **SEMANTIC INTERNALISM IS A MISTAKE**

In this paper, I introduce the concept of *narrow content* (Section 2.1) to discuss an account of narrow content by analyzing Fodor's methodological solipsism (2.2). I point out that Fodor's formalism, that is, the position according to which the content is reduced to formal properties of mental representation, eliminates (at least - as I show in Section 2.2.4 - in Stich's interpretation) semantic properties in favor of the syntactic ones. In addition, it leads to the conceptual problems indicated by J. Searle, S. Harnad (Section 2.3), and T. Burge (Section 2.4). In a nutshell, semantic internalism, as reviewed in this paper, does not offer an account of content that would be properly contentful, because it provides no grounds to ascribe truth or other semantic properties to representations. In particular, it is either unsatisfactory, because it reduces content to formal properties or inconsistent, because it appeals to innate contents that itself has not been properly explicated; moreover, innate factors, as I argue, are not merely individual. Consequently, I reject semantic internalism in favor of externalism.

The purpose of this paper is to argue against the usefulness of narrow content in the account of mental representation. By reviewing the classical arguments in favor of the narrow content, I show that the notion is inevitably wrong-headed. This is probably the reason why even one of the most radical proponents of narrow content, J. Fodor, changed his mind and rejected the narrow content in favor of wide content (Fodor 2008). Any future effort of defending the notion of narrow content will have to face the challenge of demonstrating that the narrow content has semantic properties.

## 2.1. The philosophical notion of intension and extension

It is generally assumed that there is an analogy between linguistic meaning and content of mental representation (Pitt 2013). Traditionally, intension or connotation (meaning for linguistic expressions, content for mental representation) and extension (mental representation can be about something, true or false about an object, or true or false *simpliciter*) are attributed to mental representations and linguistic expressions.<sup>1</sup> The correspondence between language and thought with reality can justify this analogy; both linguistic expressions and mental representations refer to reality and describe it.

A complex expression is extensional if its denotation is a function of denotations of its constituent expressions (Jadacki 2001). In extensional contexts, substituting one of the constituents of an extensional expression with a constituent with the same denotation does not change the logical (or semantical) value of the whole expression. Knowing the denotation of constituents of an extensional expression is sufficient to determine the logical (or semantical) value of the whole expression. I'll illustrate this by two sentences:

- a) George W. Bush voted for Barack Obama in 2012.
- b) The last but one ex-president of the US voted for Barack Obama in 2012.

Sentence b) was formed as a result of substituting one of the constituents of an extensional expression, "George W. Bush", with a constituent with the same denotation, "the last but one ex-president of

---

<sup>1</sup> Connotation is the minimal set of properties related to a language expression X that anyone speaking the language to which X belongs can use to recognize the referent of X. The notion of connotation is traditionally used interchangeably with a concept of intension (Copi and Cohen 2002). However, these concepts are sometimes differentiated. This happens when one assumes a slightly different sense of intension. In the Carnapian tradition, intension is the function of language expressions onto noncontradictory sets of propositions (Carnap 1947), and in tradition of two-dimensional semantics it is a function onto possible worlds (Chalmers 2004). This function assigns the extension to a term (in a given possible world). For example, in our possible world, the terms "Evening Star" and "Morning Star" have the same intension across contexts, but different connotations.

the US", but the logical value of the sentence did not change. Both expressions refer to a specific person, that is the former US president, George W. Bush.

An expression is intensional (referentially opaque) if and only if it is not extensional. In expressions occurring in intensional contexts, substitution of a selected constituent of a sentence with another constituent of the same logical (or semantical) value may change the logical value of the whole sentence. Intensional expressions include such sentences as "x thinks that p", "x knows that p", "x wants p", "x believes that p". Suppose that George W. Bush voted for Barack Obama, but Johnny does not know that George W. Bush is the last but one ex-president of the US. I'll illustrate this by the following example:

- a) George W. Bush voted for Barack Obama in 2012.
- b) The last but one ex-president of the US voted for Barack Obama in 2012.

Both constituent parts, namely "George W. Bush voted for Barack Obama in 2012" (in the sentence a\*) and "The last US expedition voted for Barack Obama in 2012" (in the sentence b\*), have the same logical value, but sentence a\* may have a different logical value than b\* if Johnny thinks that the last but one ex-president of the US and George W. Bush are two different people.

## **2.2. Naturalized conceptual role semantics**

A naturalistic account of content of mental representation that explicates the content in terms of their functional roles in a cognitive system is a promising attempt to provide semantics for psychology. By a functional role, it is generally meant:

the role of that representation in the cognitive life of the agent, e.g. in perception, thought and decision-making (Block, 1998).

This definition of a conceptual role is, however, very general. It can be accepted both by internalists, who adopt the notion of narrow content, as well as externalists, who embrace wide content. Narrow content is limited to the functional role within the cognitive system, while wide content also includes the context and environmental circumstances in which the cognitive system is situated.

### **2.2.1. In search of narrow content. Cognitive nature of content**

According to internalists, the concept of narrow content suffices to describe, explain, and predict the intentional states of all cognitive systems. They consider the concept of wide content to be defective and useless in psychology, as the concept seems to be too dependent on the circumstances of its occurrence and insufficiently dependent on the structure of the cognitive system itself. In order to evaluate these theses, I will closely examine the concept of narrow content, the arguments evoked for its defense, and intuitions underlying internalist views.

Internalists point out to the mind's ability to think about things that are not the case. I will illustrate such intuitions by an example of Johnny standing at the bus stop waiting for the bus. If he has an incorrect timetable, he may still think of the bus, even though a vehicle is not arriving and will not arrive. Moreover, the same thought about the bus can reappear in the boy's head at various times and at different places: both at the bus stop and at home. So, it would seem that thoughts preserve their content regardless of their context (place or time). Johnny can also think at any time of a mountain of gold, even though he knows perfectly well that it does not exist. Thoughts also retain their content when they do not refer to anything real.

The concept of narrow content preserves the independence of content of thought with regard to such factors as reference and satisfaction conditions. Thus, an internalist G. Segal describes it as *cognitive* content in contrast to referential/truth-conditional content (Segal 2000).

Internalism opposed externalism, including Wisconsin style informational-causal theory of reference, which rejects the autonomy of content from reference. The idea of extreme externalism was spelled out by H. Putnam in his Twin-Earth thought experiment (Putnam 1975). Putnam shows that intension does not determine extension. Imagine that on a distant Twin Earth, a liquid called "water" does not refer to a chemical substance H<sub>2</sub>O, but to XYZ. Water from Earth is indistinguishable from water on Twin Earth: it looks the same and tastes the same. Imagine that before the discovery of the chemical structure of water, before 1750, on Earth, there was Oscar<sub>1</sub>, who had a twin brother on the Twin Earth, Oscar<sub>2</sub>. Oscars did not distinguish XYZ from H<sub>2</sub>O and for both the intension of the expression "water" was the same: a colorless, potable liquid. Putnam claims that the meaning of the

word "water" as used by Oscar<sub>1</sub> was not the same as the meaning of a word "water" uttered on the Twin Earth, because these expressions differed in their denotation. It is not the case that intension fully determines extension, because the environment is also critical for meaning.

Internalists claim that there was a common meaning of the word "water" on both planets, which is not reflected in claims of externalists, according to which the content of the same term was different on both planets. Their claim overlooks, according to internalists, an essential aspect of mental content thanks to which one can predicate the same attributes about the same object independently of external factors.

### **2.2.2. Determination of narrow content in terms of supervenience**

According to Block's definition, internalists propose to characterize content only in terms of its causal-inferential roles within an individual cognitive system. Narrow content is therefore a part of the internal structure of an individual cognitive system; it participates in its cognitive life – in its inferences, decision-making, and so on (Block 1987). The fact that it is all about the internal structure of an individual cognitive system is emphasized by Segal (who defines narrow content in terms of local supervenience), according to which narrow content is a property of mental representation completely independent of any external factors. Internalism assumes that narrow content can only be determined by the internal structure of an individual cognitive system.

The content of mental representation is entirely determined by intrinsic properties of an agent or a cognitive system. An intrinsic property is a property that the object has (or not) regardless of what is the case beyond that object (Yablo 1999). An example of an intrinsic property is a square's property of having-four-equal-sides: a square always has four equal sides, regardless of how things are outside it. On the other hand, being a living organism is not an intrinsic property of an organism, because the organism would not have this property if it didn't, for example, breathe oxygen. And so, intrinsic properties of a cognitive system are those properties of the internal structure of the cognitive system that remain independent of any external factors of the system. To talk of such properties, I will use a term *microstructural properties*. Microstructural properties are properties of an internal structure of a cognitive system and its parts (and relationships between

them) - unlike macrostructural properties which include relationships with the environment and other agents.<sup>2</sup>

The claim about the determination of narrow content expressed in terms of local supervenience states that contentful properties depend only asymmetrically on microstructural properties of a cognitive system. Segal characterizes narrow content in terms of local supervenience on the microstructure of a cognitive system. He argues that microstructural properties are sufficient to determine the neural and computational properties of a system, i.e., narrow content:

Fix an object's microstructure and you fix its atomic and molecular structure, its neurological and computational properties, and so on. (Segal 2000, 14).

The definition of narrow content in terms of local supervenience allows us to explicate more precisely how content is determined according to internalism, and at the same time, to indicate a problem related to the determination understood this way. Local supervenience does not allow us to precisely define the character of narrow content, since it does not provide a way to define properties independent from the external environment to the system, and narrow content depends only on these properties. On the contrary, it excludes only certain groups of (externalistic) accounts. Moreover, the consequence of a local supervenience claim is that all cognitive interactions with environment are irrelevant for content, which is fully reducible to the microstructure of the system.

Here, an ontological reduction is at stake, that is, a relation between elements of the real world, such as objects, events or properties. It occurs if relationships, such as elimination, identity, superposition, realization, or supervenience occur (van Gulick 2001). Because in internalism supervenience between properties is assumed, it is an example of such an ontological reduction.

---

<sup>2</sup> The terms *microstructural* and *macrostructural* have been proposed by R. Poczobut, who formulated the supervenience claim in terms of micro and macrostructure (Poczobut 2007).

### **2.3. Fodorian methodological solipsism**

I will now concentrate on Fodor's defense of the concept of narrow content in his methodological solipsism (Fodor 1980; Stich 1980). Narrow content in Fodor's view arises from the reduction of semantic properties to syntactic ones and to innate semantic properties as well. This reduction, is not, however, a full naturalization, as it is unclear how innate content is determined. Thus, only a total reduction of the content to syntactic properties is fully consistent and naturalistic, but such an account on narrow content deprives it, alas, of its content. Consequently, the syntactic understanding of content is - contrary to what Fodor claims - inadequate for psychology and cognitive sciences. Thus, conceptual role semantics that would accept only narrowly understood content would not be a semantics for psychology.

#### **2.3.1. Narrow content in methodological solipsism**

Fodorian methodological solipsism plays a key role in developing the concept of narrow content in psychology and philosophy of psychology. Within this framework, Fodor attempts to defend methodologically individualist psychology as the only proper approach to psychological research. He responds to Putnam's counterarguments against the classical claim that intension determines extension<sup>3</sup>

Narrow content in methodological solipsism is characterized by inferential roles that are syntactic and computational.<sup>4</sup> Fodor argues that knowledge of intrinsic properties, especially formal representational properties suffices to describe the content of a

---

<sup>3</sup>Later Fodor did not link individualism in psychology and methodological solipsism so closely. Moreover, he distinguished between methodological solipsism and methodological individualism (Fodor 1987). According to methodological solipsism, mental states are individuated without semantic valuation; an (externally) relational taxonomy of mental states is methodologically unacceptable. Methodological individualism, on the contrary, allows for relational individuation of mental states provided that a mental state property is only included in the mental states taxonomy if it is causally relevant (Heath 2015).

<sup>4</sup>Identification of inferential roles with causal ones stems from the classic computational account of functionalism. Inferential roles characterized syntactically are roles in a computational architecture of mind (Field 1978; Fodor 1975).

representation. He claims that it suffices to express all content relevant to cognitive psychology.

According to Fodor, a representational relation has two related members. It consists of a relation to a sentential object, described formally (syntactically), and of a relation to this object, described semantically in terms of truth and reference. However, as Fodor adduces, formal properties of narrow content are sufficient to distinguish semantically different representations. Fodor is, therefore, a proponent of a formalist account of a theory of content, whereby formal properties of signs (e.g., their shapes or structure) and syntactic rules are sufficient to characterize content.

Fodor describes the mind as a so-called oracle machine, where "oracle" is understood, after Turing, as a procedure that settles a question in a non-computational (non-algorithmic) manner. According to Fodor, the role of oracle is played by the perceptual states of an environment:

The point is that, so long as we are thinking of mental processes as purely computational, the bearing of environmental information upon such processes is exhausted by the formal character of whatever the oracles write on the tape. In particular, it doesn't matter to such processes whether what the oracles write is true; whether, for example, they really are transducers faithfully mirroring the state of the environment, or maybe the output end of a typewriter manipulated by a Cartesian demon bent on deceiving a machine (Fodor 1980, 65).

Fodor claims that a formal difference makes a functional difference, which in turn makes a causal difference:

The form of explanation goes: it's because different content implies formally different internal representations (via the formality condition) and formally distinct mental representations can be functionally different; can differ in their causal role. Whereas, to put it mildly, it is hard to see how internal representations could differ in causal role unless they differ in form (Fodor 1980, 68).

For Fodor, this is a pragmatic argument for substituting formal properties with semantic ones in explanations.

Fodor, like Davidson, defends folk psychology, and therefore points out that statements about beliefs, thus referentially opaque



contexts, are crucial in folk psychology. He even claims that folk psychology does not need anything more than an explanation of representation in opaque contexts, in which generalizations are about what people mean by propositions to which they express propositional attitudes.

Propositions that occur in opaque contexts differ not only in content but also in their form. That is why such sentences do not undermine the Fodor's assumption that only intrinsic structural properties have causal powers. Formal properties suffice, according to Fodor, to adequately describe the content, such as beliefs, even in opaque contexts. Later in this paper, I will argue against Fodorian concept of narrow content, according to which a formal difference is sufficient to explain the difference in intension.

### **2.3.3 Concept innateness**

Fodor defends his own account of concepts, which is a philosophical interpretation of a classical theory of concepts, enriched by an innateness hypothesis. In a classical theory, the concept is:

a class representation, covering all relevant properties of such class. A criterion of relevance is repetition - an essential feature is the feature that characterizes all objects belonging to this class, i.e., - in other words - the common feature of all objects belonging to this class (Maruszewski 1983).

Maruszewski's definition differs just a little from the classical definition of connotation, according to which connotation of *z* is a property which applies to all *z*-s and only *z*-s (see also Jadacki 2001, 107).

The innateness hypothesis states that our basic conceptual apparatus is innate.<sup>5</sup> We have a language of thought, that is, an internal

---

<sup>5</sup>The concept of innateness is unclear and can be understood in many different ways, especially since in contemporary psychology one does not disregard the biological basis of cognition. In biology, it is by no means clear not only what is innate but also what innateness is; philosophers of biology challenge the utility of such a concept (Samuels 2002, 2004; Griffiths 2002). One of the explications of the concept of innateness on the biological ground defines innateness as a disposition to behave under normal conditions. This explication, however, must go beyond narrow content, as normal conditions always appear in an environment.

code, to which all other concepts can be reduced, and we can perform proper combinatorial operations (Fodor 1975). For example, a non-basic concept of BORING BOOK consists of basic concepts BOOK and BORING. One of Fodor's most important arguments for the language of thought hypothesis in psychology is the productivity of thought. Language is one such productive system. Fodor has to justify why it is language and not another productive representation system that is innate. The key argument for language innateness serves this purpose.

Fodor (1975), analyzing psychological theories of concept learning, notes that they all regard learning concepts as a process of hypothesizing. Hypotheses can only be posed in language, and, therefore, in order for a child to pose them she must have an innate language of thought. Before she can learn concepts she must be able to formulate such hypotheses. That is why language, and not a different productive representation system, is innate: in a different system such hypotheses cannot be made.

Let's go back to productivity. It is obvious that we can also think an infinite number of different thoughts; similarly, we can utter an infinite number of sentences, for example "Giraffes do not play poker". According to Fodor, it is impossible to explain the productivity of language and thought without assuming the compositionality of language. It is language that is so rich in structure that makes it productive. Therefore, as he claims, thinking occurs in a linguistic medium.

A special case of concepts in the Fodorian account are concepts that can no longer be broken into constituent parts. These include concepts for simple sensory qualities, i.e., shapes and colors, and the simplest colloquial concepts. The structure of a concept also plays a role in deciding whether a concept belongs to a basic category: it must be a concept without which other concepts cannot be created in virtue of a compositional principle. For example, a concept BORING BOOK is a non-basic complex concept consisting of a basic concept BOOK and a basic concept BORING - concepts BOOK and BORING cannot be simplified further and these are concepts without which creating a concept BORING BOOK would be difficult.

By design, methodological solipsism forbids citing environmental properties in explaining the determination of the content of such basic concepts. Thus, they remain unexplained, and the

Fodorian account can never really explain how their content was determined. The content of basic concepts remains an aporia of methodological solipsism. Within methodological solipsism, determining the content of basic concepts is logically impossible: if they are innate, they depend not only on the individual cognitive agent but also on biological inheritance, which goes beyond the agent.

The hypothesis of an innate conceptual apparatus could remain consistent with the local supervenience of such narrow content on the formal properties of a conceptual apparatus, if only the formal properties of this apparatus constituted the basis of conceptual content. Fodor, however, is opposed to the total reduction of semantics to formal properties. One way to solve the problem of the determination of content of basic concepts, which would allow him to preserve methodological solipsism, is to treat it precisely as a formalistic account, that is, purely syntactic. As a result, this could strengthen and radicalize Fodor's account.

Fodor does not accept the strong claim that all semantic properties can be reduced to syntactic ones, but he claims something weaker: formal properties are the most satisfactory indicator of content. I do not intend to argue with that: indeed, the form is the simplest indicator of difference in content and, in addition to this, it works well in many situations.

#### **2.3.4. Methodological solipsism and a formalist account**

In this section, I will examine Stich's more radical account which argues for a complete reduction of semantics to syntax and, consequently, for the elimination of the notion of representation from folk psychology. Stich indicates to what exactly the reduction of content to formal property leads. In essence, Stich shows that an internalist concept of content is not a concept of content. This means that internalism cannot naturalize intentionality.

Stephen Stich starts from a formalist interpretation of methodological solipsism, criticizing Fodor's argument for this position. Stich's counterargument can be understood as being directed against methodological solipsism as well as against psychology that uses the notion of representation. Stich's objection to Fodor's notion of narrow content shows that this notion is divergent from its folk counterpart and, consequently, cannot be used to defend folk psychology. Stich thus

argues for rejecting any concept of content. In addition, he is against representationalism because he thinks - as Fodor does - that computational psychology does not adhere to the principle of charity and thus, does not take the semantic properties of representations, such as truth, into account.

Stich argues against Fodor that his concept of narrow content leads to undesirable consequences:

First, most computational (or formal) mental states will have tokens (either actual or possible) whose contents are radically different from one another, as judged by our "aboriginal, uncorrupted, pretheoretic intuition". Second, there will be some computational mental state types whose tokens can be assigned no content at all by our aboriginal intuitions, though these "contentless" computational states will serve the purposes of the computational theory of mind fully as well as their contentful cousins (Stich 1980, 97).

First, Stich observes the concept of narrow content leads to ascribing the same content in intuitively different cases. Second, he stresses that an account of narrow content typical for methodological solipsism leads to one more undesirable consequence: it does not forbid attributing content to beliefs radically diverging from our own beliefs, even if they violate common intuitions about their content.

The cases of the first kind are analogous to Putnam's example of Twin Earth. I will mention one of them. In Stich example, Fodor from Yon (Putnam's Twin Earth analogue) appears and utters, analogically to Fodor from Earth: "Jimmy Carter is from Georgia." On the Twin Earth far apart from our Earth, even further than Yon, there could be a Twin Fodor, uttering the same sentence "Jimmy Carter is from Georgia". But then, according to the Fodorian account of narrow content, the same content should be attributed to this sentence on Twin Earth. Stich notices that this is completely inconsistent with a common intuition that different Fodors, depending on where they are located, speak of a different Jimmy Carter, depending on where they come from (assuming that on Yon and on Twin Earth there is also Georgia). I agree with Stich that the context of utterance should be taken into account while assigning content to a belief. Indeed, defending an account of reference of proper names requires one to accept a causal account of content that is incompatible with methodological solipsism. Dependence of content

on context is one of the reasons for adopting the wide account of content.

The notion of narrow content should also be attributed to such things or creatures to which the principle of charity would forbid attributing any content. Stich illustrates this with examples of robots whose beliefs are so distant from ours that one can never agree that, according to the principle of charity, their representational states have any content. If there was a robot simulating a human being with beliefs vastly contradicting one other, the principle of charity would not allow us to attribute content to it, because, in such a case, neither truth nor accuracy of its "beliefs" could be treated seriously. Fodor has no way to deny that these "beliefs" are meaningful by his own lights.

In conclusion, Stich shows that a formalist account does not lead to a defense of folk psychology but rather to its rejection. Although he agrees with Fodor that a formalist account suffices to describe content, he goes a step further, claiming that a notion of representation should be rejected totally from folk psychology. At the same time, according to him, we must reject a notion of narrow content and substitute it with a notion of form. I agree with Stich's argument against a notion of narrow content if narrow content is reduced to formal properties. Such a reduction does not properly describe content in contexts in which expressions differ in content but not in form. I propose, however, to treat Stich's argument - contrary to his intentions - as a warning against an excessively hasty reduction of content to form and against the elimination of the concept of mental representation.

Narrow content in methodological solipsism will not allow us to distinguish between representations whose form does not decide their meaning. This group includes homonyms (such as "bank") and representations whose meaning depends on the environment (Putnam's example of water on Twin Earth). Their form is the same, but the content is different because: (A) both intension and extension are different (in the case of homonyms); or (b) extension (of representations whose content depends on the environment) is different. A formalist account could deal with homonyms at the level of expressions, denying *de facto* their existence: by separating those sentences where the word "bank" in the meaning of "a bench of the river" is different from sentences in which there is a "bank" in the meaning of "a building in which you put your money" because of

inferential roles that these words play in sentences, and the roles determine the meaning of the word "bank". For formalists more difficult are homonyms at the level of the sentence, where a pragmatic context plays a decisive role: the sentence "You have huge feet!" uttered in a shoe store expresses the fact that someone has feet of large size, but uttered in the presence of someone with small feet is ironic. One should also remember the role played by the intonation: the same sentence said in a dismissive tone can be offensive (Dennett 1991). In the case of homonyms at the level of the whole sentence, resorting to inferential roles is practically infeasible - it is impossible to distinguish sentences that are so strongly contextually dependent, purely on a syntactic level, thus invoking only their formal properties.

The formalist account may, however, show the difference in meaning between the sentences "George W. Bush voted for Barack Obama in 2012" and "The last but one ex-president of the US voted for Barack Obama in 2012," because the term "George W. Bush" has a different form than the term "the last but one ex-president of the US".

The formalist account is valid in many interesting cases, in which the form of a vehicle corresponds to representational content. It would be a mistake to ignore formal properties in an account of content. However, in order for a formal account to fully replace semantics, it should be able to explain the cumbersome cases described above.

Thus, the adequate account of content should not be a purely formal conception of content, since such an account is powerless in those contexts, in which the reference clearly decides about content. Twin Oscar's statement about water is therefore considered to be different from Earthly Oscar's statement in which "water" refers to a different chemical structure on Earth than the one to which it would refer on Twin Earth. It is therefore reasonable to assume that an adequate account of content should not be merely formal.

## **2.4. Arguments against the formalist account**

### **2.4.1. Chinese Room thought experiment**

The most well-known polemic against supporters of a formal account of representation, in this case symbolic representation, can be found in J. Searle's Chinese Room thought experiment (Searle 1980). Searle who doesn't know any Chinese is enclosed in a room with a text file in

Chinese ("script") with some additional files: a set of rules (equivalent to a program) correlating the second file with the first one (called by Searle "a story") and a set of instructions given in English (questions), allowing to correlate elements from the third file with the first two. These instructions specify how to send certain Chinese symbols of certain shapes, responding to those shapes in a third file. Both the program and answers to such questions are for Searle purely syntactic transformations of symbols. Searle, in his room, is supposed to answer the questions given sometimes in English, and sometimes in Chinese. As it turns out, his answers to the questions in Chinese do not differ from those of a Chinese man who natively speaks Chinese. Additionally, the answers to questions posed in Chinese are as good when seen from the outside as the answers to questions in English. In the first case, Searle's answers are only non-interpreted symbols.

Searle's thought experiment is supposed to deal with many philosophical issues, including consciousness, artificial intelligence, and mental representation. I will focus on the last issue, and within it, on the subject of my interest: the possibility to characterize content solely formally. For this purpose, I will appeal to one of many formulations of Searle's argument, and within it, to the interdependence of form and content (Hauser 1997). The argument has the following form:

1. Programs are purely formal.
2. Minds (or at least human minds) have semantics, mental content.
3. The syntax itself does not constitute content nor is a sufficient condition for content.

Programs themselves are neither constitutive nor sufficient for the functioning of mind. (Preston and Bishop 2002).

Many philosophers question the validity of Searle's argument (Dennett 1987; Chalmers 1996). I think, however, that the core of Searle's argument, that is the claim that syntax is neither identical to content, nor sufficient to describe semantic content, points out a problem that has not been solved by a formalist account of mental representation. It is the case independently of Searle's own account of intentionality (Searle 1983) that is based on his intuition about the role of consciousness.

#### **2.4.2. The systems reply**

The systems reply to Searle's experiment comes down to a claim that although Searle, as a person confined in the Chinese room, does not understand Chinese, the whole system does (Searle 1980). Supporters of the systems reply point out that the fact that a person in a room does not understand Chinese does not imply that the system does not. And Searle has never shown anywhere that a whole system does not understand it. According to his opponents, Searle makes a mistake of identifying the part of a system with the system as a whole. Searle would be right only if understanding could be divided like mass. We may cut an apple into pieces: the mass applies both to an apple and to its component parts. Searle must in fact assume that every part of a systems thinks. So, if a person himself understands, a stomach or a liver, for example, understands too; if a stomach would not understand, then a person would not understand too (Copeland 1993).

In response to this objection, Searle argues that, based on his opponents' arguments, a system that has memorized incomprehensible rules constitutes, together with a sheet of paper, a thinking system, which would be absurd. Searle states that there must be a difference between "genuinely mental" systems and those that are not genuine, and that the system itself must be able to detect the difference. Such a system displays - according to him - biologically "hardwired" intentionality.

The problem posed by Searle is deep but his solution unsatisfactory. This is because Searle *a priori* settles the intriguing problem of demarcation between thinking systems and other systems, without showing what the differences actually are. The claim that systems are different definitely does not suffice as a solution. And why exactly are consciousness and biological brains important according to him remains unexplained.

#### **2.4.3. Chinese Room thought experiment reloaded**

In this section, I will briefly present Searle's experiment in a version slightly modified by S. Harnad; here I also separate the problem of intentionality from the problem of consciousness. Harnad helps us better describe the problem of the relation between a formal symbol and reference of mental representations, which, thanks to his paper, in



artificial intelligence has been called "the symbol grounding problem" (Harnad 1990).

Harnad's experiment has two versions. In the first version, he describes learning Chinese as the first language using only a Chinese-Chinese dictionary, while in the second one, he describes learning Chinese as a second language. The latter one he considers to be feasible, though difficult. However, theoretically crucial is the first one.

In Harnad's variant, the Chinese Room experiment is about grounding symbols in something other than other meaningless symbols. It is not other symbols, according to Harnad, that constitute meaning, but relations of the cognitive system with the world. The problem is not whether it is possible to translate some specific language of a given linguistic form into another language of another linguistic form, but rather how to relate a linguistic form to the world.

Searle's argument, which is more evident in Harnad's version of it, is that syntactic properties are not sufficient to capture semantics, because for a proper description of semantics one also needs reference and logical value. The argument is thus directed against the internalism of a formalist vein. The formalist account does not allow us to account for the reference. As Fodor shows in the case of sentences with intensional contexts, formal properties of a vehicle make it possible to infer much about truth or falsity of representation. A formalist account does not, however, provide any explanation as to why a representation can be a vehicle of truth. Even if you accept a formalist assumption that mental representations have syntax, which is causally efficacious, it is not clear at all that representations have a property of being true or false. Moreover, under formalistic assumptions, it is by no means clear what physical structure could be considered syntactic and why certain syntactic constructs would correspond to falsehoods and others to truths.

### **2.5. Internalism and intension determination via learning**

Internalists have difficulties in explaining action, which is related to their psychologically implausible approach to learning concepts. Fodor assumed that there were necessary and sufficient conditions for having concepts and that an account of narrow content should serve as a satisfactory psychological theory of learning (Fodor 1980). Such an

approach to learning does not allow, however, to take into account determination of intension via learning.

Burge's argument (Burge 1979) is aimed against the account of narrow content and is based on the human capacity to learn concepts. In his example, we are dealing with a thought experiment built analogously to Putnam's experiment, but Burge argues for a role of social context in content determination rather than for the determination by some physical facts. In the experiment, we compare a person whose physical states from birth until now are the same, but which occur in two situations that differ only in the linguistic community or in the social environment. As a consequence, such a person in these two situations uses a term *arthritis* differently: in the first situation the person knows well the extension of the term, and in the other, he or she uses the term *arthritis* to designate a disease that can occur in both muscles and joints. According to Burge, the extension of the term depends on the social context in which the person is raised and in which such a term is used. A defender of narrow content could answer that in the second situation the person has only an inkling about arthritis. He or she knows only that it is a disease but he or she is mistaken about what kind of disease it is. However, the defenders of narrow content go too far. They claim that assigning to a person any knowledge of the term's extension is unjustified in the second case, since it is not known what it describes.

Usually, learning concepts is time-consuming and gradual. However, according to the classical account, you have the same concept only if you mastered it completely; so the concept of arthritis that is not fully mastered is not yet a concept of arthritis. If that is the case, then we would have to assume that when we do not know necessary and sufficient conditions of concept application (and Fodor himself argues for the claim that, in general, we do *not* know them; cf (Fodor et al. 1980)), we do not know the same concepts. Without the assumption of innateness, this leads to a very peculiar consequence. It is not easy to indicate necessary and sufficient conditions for the use of terms such as "game", "chair", or "animal", and if they were not innate, then according to the classical theory of concepts, we should say that we do not know them at all. However, if they are innate, then their content is not determined individually. Here again we come across the fundamental aporia of the Fodorian account: his nativism excludes methodological

solipsism, since innate concepts must have content determined by factors that do not supervene locally, i.e., have content that goes beyond narrow content.

What an internalist, such as Fodor, argues against Burge, exposes the weaknesses of Fodor's internalism. Of course, an internalist could give up also the classic theory of concepts and nativism, but then Burge's argument would strike him. So he would have to agree that some determinants of content are social.

## **2.6. Summary**

In this paper, I demonstrated that the argument of one-factor internalist account for the sufficiency of narrow content in the theory of representation is inadequate. To summarize, the characterization of narrow content leads either to ambiguity or to depriving the resulting concept of content of semantic properties. If by "narrow content" we mean – like Segal – the property of representational content that is completely independent of external factors to a cognitive system, the concept of content remains elusive and nobody knows what it could be. Although understanding content as partially independent from contextual factors allows us to hold content properties invariant in various situations, it seems that understanding content in total abstraction from the external factors of such properties does fit the bill. On the other hand, Fodor's formalistic account, in particular in Stich's radical interpretation, eliminates the properties of content to replace them with syntactic ones.

Reasons quoted by defenders of narrow content, such as the ability to articulate thoughts independently of the context or thinking about non-existent objects, speak in favor of the concept of narrow content. Nevertheless, the concept of narrow content abstracts away from both reference and satisfaction conditions, without which it is impossible to understand how mental representations can be vehicles of content. The lack of connection to reference and satisfaction conditions makes it for the account of narrow content impossible to state anything about the adequacy of representation with regard to their targets or referents. Some of these representations apply to an environment, which, in the correct account of content, would explain adaptive behaviors of animals as based on adequate representations of an environment, such as orientation in an environment (e.g., through

cognitive mapping). Narrow accounts of content do not allow us to state that, for example, a predator made a mistake in hunting while looking for a victim. For this reason, semantic internalism is a mistake as a solution to the problem of intentionality.

## REFERENCES

- Block, Ned. 1987. "Advertisement for a Semantics for Psychology." *Midwest Studies in Philosophy* 10 (1): 615–78. doi:10.1111/j.1475-4975.1987.tb00558.x.
- Burge, Tyler. 1979. "Individualism and the Mental." *Midwest Studies in Philosophy* 4 (1): 73–121. doi:10.1111/j.1475-4975.1979.tb00374.x.
- Carnap, R. 1947. *Meaning and Necessity*. Chicago: University of Chicago Press.
- Chalmers, David J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- . 2004. "Epistemic Two-Dimensional Semantics." *Philosophical Studies* 118 (1/2): 153–226. doi:10.1023/B:PHIL.0000019546.17135.e0.
- Copeland, B. Jack. 1993. *Artificial Intelligence: A Philosophical Introduction*. Oxford, UK; Cambridge, Mass.: Blackwell.
- Copi, Irving, and Carl Cohen. 2002. *Introduction to Logic*. 11th ed. Upper Saddle River N.J.: Prentice Hall.
- Dennett, Daniel C. 1987. *The Intentional Stance*. Cambridge, Mass.: MIT Press.
- . 1991. *Consciousness Explained*. New York: Back Bay Books / Little Brown and Company.
- Field, Hartry. 1978. "Mental Representation." *Erkenntnis* 13: 9–61.
- Fodor, Jerry A. 1975. *The Language of Thought*. 1st ed. New York: Thomas Y. Crowell Company.
- . 1980. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology." *Behavioral and Brain Sciences* 3 (1): 63–63. doi:10.1017/S0140525X00001771.
- . 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass.: MIT Press.
- . 2008. *LOT 2: The Language of Thought Revisited*. Oxford University Press.
- Fodor, Jerry A., M F Garrett, E C Walker, and C H Parkes. 1980. "Against Definitions." *Cognition* 8 (3): 263–367.
- Griffiths, Paul Edmund. 2002. "What Is Innateness?" *The Monist* 85 (1).

- Gulick, R. van. 2001. "Reduction, Emergence and Other Recent Options on the Mind/body Problem. A Philosophic Overview." *Journal of Consciousness Studies* 8 (9–10): 1–34.
- Harnad, Stevan. 1990. "The Symbol Grounding Problem." *Physica D* 42: 335–46.
- Hauser, Larry. 1997. "Searle's Chinese Box: Debunking the Chinese Room Argument." *Minds and Machines* 7 (2): 199–226.
- Heath, J. 2015. "Methodological Individualism." In *The Stanford Encyclopedia of Philosophy*.
- Jadacki, Jacek. 2001. *Spór O Granice Języka: Elementy Semiotyki Logicznej I Metodologii*. Warszawa: Wydawn. Naukowe Semper.
- Maruszewski, Tomasz. 1983. *Analiza Procesów Poznawczych Jednostki W Świetle Idealizacyjnej Teorii Nauki*. Poznań: Wydawn. Nauk. Uniwersytetu im. A. Mickiewicza.
- Pitt, David. 2013. "Mental Representation." Edited by Edward N Zalta. *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/fall2013/entries/mental-representation/>.
- Poczobut, Robert. 2007. "Eksternalizm Treści Umysłowej a Superweniencja." *Kognitywistyka I Media W Edukacji* 2 9 (1): 82–106.
- Preston, John, and Mark Bishop. 2002. *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford; New York: Clarendon Press.
- Putnam, Hilary. 1975. "The Meaning of Meaning." In *Philosophical Papers, Vol. II: Mind, Language, and Reality*. Cambridge: Cambridge University Press.
- Samuels, Richard. 2002. "Nativism in Cognitive Science." *Mind and Language* 17 (3): 233–65. doi:10.1111/1468-0017.00197.
- . 2004. "Innateness in Cognitive Science." *Trends in Cognitive Sciences* 8 (3): 136–41. doi:10.1016/j.tics.2004.01.010.
- Searle, John R. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3 (3): 1–19. doi:10.1017/S0140525X00005756.
- . 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge / New York: Cambridge University Press.
- Segal, Gabriel. 2000. *A Slim Book about Narrow Content*. Cambridge Mass.: MIT Press. [http://www.worldcat.org/title/slim-book-about-narrow-content/oclc/42771559&referer=brief\\_results](http://www.worldcat.org/title/slim-book-about-narrow-content/oclc/42771559&referer=brief_results).

- Stich, Stephen P. 1980. "Paying the Price for Methodological Solipsism."  
*Behavioral and Brain Sciences* 3 (1): 97.  
doi:10.1017/S0140525X00002016.
- Yablo, S. 1999. "Intrinsicness." *Philosophical Topics* 26 (1/2): 479–505.

## **ABSTRACT**

### **SEMANTIC INTERNALISM IS A MISTAKE**

The concept of narrow content is still under discussion in the debate over mental representation. In the paper, one-factor dimensional accounts of representation are analyzed, particularly the case of Fodor's methodological solipsism. In methodological solipsism, semantic properties of content are arguably eliminated in favor of syntactic ones. If "narrow content" means content properties independent of external factors to a system (as in Segal's view), the concept of content becomes elusive. Moreover, important conceptual problems with one-factor dimensional account are pointed out as a result of analysis arguments presented by J. Searle, S. Harnad and T. Burge. Furthermore, these problems are illustrated with psychological and ethological examples. Although understanding content as partially independent from contextual factors allows theorists to preserve content properties, it seems that understanding content in total abstraction from external factors of these properties is implausible. As a result, internalism is rejected in favor of externalism.

**KEYWORDS:** internalism; externalism; one-factor dimensional account of representation; mental representation; Fodor; methodological solipsism