

<https://doi.org/10.3176/tr.1999.1/2.03>

INFORMATIONAL ATOMISM VERSUS INFERENTIAL ROLE SEMANTICS IN JERRY FODOR'S

Concepts: Where Cognitive Science Went Wrong

Bruno Mölder

University of Tartu

Abstract. This review essay on Jerry Fodor's "Concepts: Where Cognitive Science Went Wrong" concentrates on Fodor's informational atomism with respect of concepts and his rejection of inferential role theories of concepts (definition theory, feature theory and theory-theory). It is argued that Fodor's arguments against both inferential role semantics and these theories of concepts are not conclusive and that his own account leaves out the cognitive significance of concepts.

Jerry Fodor's recent book "Concepts: Where Cognitive Science Went Wrong" that is based on his 1996 John Locke Lectures at Oxford, puts forth a theory of concepts that is both atomist and in concord with the informational semantics that he has developed over past ten years. As a part of arguing for his own account, he criticises those types of theories about concepts that are prevalent in the cognitive science community. If his arguments are successful, then it would perhaps bring about re-evaluation of the very foundations of cognitive psychology. Therefore this book deserves careful study both by philosophers and psychologists. In what follows, I shall have a look on Fodor's arguments against the competing theories of concepts as well as trying to analyse his own account. I shall come to the conclusion that Fodor's arguments are inconclusive and that his own theory as a whole is not more successful than the rival theories he criticises. However, I do not attempt to cover the whole book – for example, I will say almost nothing about Fodor's treatment of the innateness issue in the final chapters of his book.

1. The representational theory of mind and constraints to the theory of concepts

Fodor's views on concepts should be best understood in the light of the so-called representational theory of mind. The representational theory is not a proper theory in the first place, it is rather a kind of a picture or image of mind, which is supposedly so widespread in the cognitive science community that Fodor (1998:23)¹ describes it as "the only game in the town". The central idea of this representational picture is that mental representations are causally related to each other in virtue of what they represent, i.e., their intentional contents. Explanation in psychology makes use of nomic generalizations that quantify over contents of mental representations. Mental representations are particular physical symbols that have semantic properties. Mental processes could be modelled as computations that are understood as causal transformations of symbols (mental representations) that preserve their semantic values. What is it to have a symbol should not be explained in computational terms according to Fodor (p. 11), otherwise it would generate circularity problems. Therefore, Fodor's account of the content of mental representation is not computational; and as it will become evident in the last section of this essay, it is informational.

What is the relevance of this representational picture of mind for the theory of concepts? The relevance is actually mutual for Fodor. The representational theory needs the theory of concepts and the theory of concepts is in turn constrained by the representational theory. Concerning the first side of dependence it could be said that the important part of the representational theory of mind is the acknowledgement of primitive and complex mental representations. According to Fodor, the intentional properties and causal powers of complex mental representations derive from the contents and causal powers of their primitive constituents, namely concepts. The validity of representational theory of mind is therefore dependent on there being an adequate theory of concepts.

On the other side, the representational theory of mind itself constrains a range of conditions that an adequate theory of concepts has to answer. It is reasonable to outline them here, since they are important in assessing the acceptability of a theory of concepts. They are the following (pp. 23–28):

1). Concepts have to be capable of being in causal interaction with mental processes (which are "computations on structured mental representations" for Fodor (p. 22). Although it is arguable whether this commitment to the theory of concepts is too heavy, it follows from the theoretical presumptions of the representational theory of mind. So this condition has to stand if one aims to find a theory of concepts suitable for the representational theory of mind.

2). Concepts are applied to the objects and events in the world and this application has conditions of semantic evaluation. Fodor (p. 24) draws an

¹ In what follows, I will refer to Fodor (1998) only by page numbers. If the reference is made to some other Fodor's publication, then the year is also used.

important distinction between the conditions of semantic evaluation and the confirmation conditions. Being a realist about meaning, he rejects all kinds of verificationism: a concept is satisfied by its extension and this is its essential property. If a concept of *x* does not apply to the instances of *x*, ipso facto it is not the concept of *x*. Confirmation conditions, on the other hand, are epistemic matter and the misapplication of a concept does not have any implications to its meaning.

3). Concepts compose mental representations and according to the principle of compositionality, “mental representations inherit their contents from the contents of their constituents” (p. 25). This property of mental representations, if true, turns out to be an extremely powerful weapon in Fodor’s hands, since it explains productivity and systematicity of propositional attitudes and makes it possible for him to argue against inferential role theories, since allegedly those do not allow this feature. The productivity of propositional attitudes is the fact that one could have infinitely many beliefs and desires with different contents. The systematicity is the ability to think thoughts with combined conceptual contents, once one has an ability to entertain its constituent concepts. The representational theory of mind explains systematicity and productivity by appeal to there being infinitely many possibilities of combining finite number of primitive concepts.

4). This condition concerns the issue about the acquisition of concepts and Fodor (p. 27) claims that many concepts are not innate but learned. Fodor tackles the difficult problem of reconciling his atomist account of concepts that are not innate with the usual claim that atomist concepts have to be innate. I do not discuss the issue of innate concepts in this essay, but I have to record my puzzlement of why Fodor takes the condition that many concepts be not innate as the condition that originates from the representational mind. The representational picture is fully compatible with the idea that primitive representations are innate – in fact, even Fodor himself has once held this view.

5). Concepts have to be conceived as public entities in the sense that tokens of the same type of concept could be possessed by different people. This requirement could be read in various ways, for example, to include people only from the same historical-cultural period or from the same linguistic environment, but Fodor (pp. 28–29) favours the strongest version according to which people with very different knowledge and experience and cultural background could nevertheless possess concepts of the same type. This antirealist condition to the theory of concepts is needed according to Fodor, since it makes possible to apply psychological intentional generalizations to different people. Conceptual relativism, on the other hand, has its consequence that the psychological generalizations of our time do not apply to, e.g., Socrates. One should notice, however, that since Fodor’s theory of concepts is informational, all that is needed for the essential change in concepts is the change in the environment, the change in those properties our concepts are locked² onto. It seems to follow that Fodor has to concede that the Inuit, for

² The metaphor of “locking” originates from Loewer and Rey (1991) and captures the idea of the nomological and causal dependence of representational content from environmental properties.

example, should have fundamentally different concepts from ours, unless Arctic is environmentally similar to New Jersey, Oxford or Tartu which it is not.

These are the conditions that Fodor takes to be non-negotiable in the sense that disputing these conditions means disputing the very representational theory of mind itself and his claim is that some existing theories of concepts (the feature theory and the theory-theory, but not the definition theory) fail to satisfy them. And here is the reason for Fodor to reject those theories. This incompatibility with conditions is not, however, the sole reason why Fodor thinks that those theories are mistaken. According to him, they all presuppose the same metaphysical account of concepts that is not true. In the following section, I concentrate on the arguments against this view.

2. The rejection of inferential role semantics by Fodor and Lepore

The account of concepts that Fodor rejects is inferential role semantics. Versions of this idea are widespread in philosophy and cognitive science (see pp. 35–36). In the theory of meaning, it is in the form that the meaning of an expression is given by its role in a language. Originally proposed as an account of the meaning of logical constants, the idea was extended to non-logical terms. For present purposes, it is the thesis that the content of primitive concepts is constituted by those inferential relations the concepts enter into. This sets its requirement also to the form of concepts' possession conditions. For example, the content of the concept CAT³ is constituted by the inference from CAT to ANIMAL and possibly to NOT A TIGER and to the complex concept IS USUALLY DOMESTIC AND HAS WHISKERS. Accordingly, possessing the concept CAT entails the possession of lot of other concepts, e.g., ANIMAL, FURRY, etc., and having the ability to draw meaning-constitutive inferences. The central claim in Fodor's book is that this view is false. One could find the general arguments in support of this claim in Fodor and Lepore (1991). Since Fodor's and Lepore's opposition to inferential role semantics has been a matter of the great debate in recent years and is already a topic of its own, I do not attempt to cover this issue entirely in the present essay, but try instead to delineate the main arguments against inferential role semantics and assess their success.

Fodor's and Lepore's (1991) critique of inferential role semantics is centred around two issues, namely, that it leads to total holism concerning meaning and that it falls short of accounting the compositionality.

Let us take the holism issue first. If the inferential role theorist claims that the meaning or content of a concept is constituted by the concept's inferential relations, he has to specify which inferential relations from the total set of concept's inferential relations are constitutive. One possible answer could be that all

³ I follow Fodor's usage in that the names of concepts are in capitals, and this is not the structural description of concepts.

inferences containing a given concept form a part of its conceptual role and hence constitute meaning. But such a move would carry damaging holistic consequences for the inferential role theory (Fodor and Lepore 1991/1994:145). If one infers from "Cat mews" that "Animal is hungry" and as stipulated, all inferences are constitutive to meaning, then this inference is part of the meaning of concepts TO MEW and CAT. Now, if there are people who do not draw that inference, then their respective concepts CAT and TO MEW are different. Proneness to draw certain inferences depends on one's beliefs. Since it is uncontroversial that people's beliefs differ and given the stipulation about concepts it follows that people cannot share concepts. It's even worse, since one's own beliefs tend to change over a period of time, then the content of one's concepts has to undergo corresponding changes if all concepts inferential relations are constitutive to its meaning (cf. Block 1993:2).

Another problem for inferential role theory arises from compositionality (Fodor and Lepore 1991/1994:147). As stressed above, the compositionality of mental representations and language provides an explanation to the productivity and systematicity of thought. The trouble for the inferential role theory is exemplified in the following *modus tollens*. If the inferential role is to constitute meaning, then inferential role should be compositional. Inferential role is not compositional, therefore meaning is not constituted by the inferential role. An example: the inferential role of the MEWING CAT does not depend only on the inferential roles of its constituent concepts, but also on all the inferences that one is prepared to draw from the fact that there is a cat that mews. If one believes that mewing cats are hungry then this is a part of the inferential role of the MEWING CAT. According to the compositionality requirement, it should follow from the inferential roles of TO MEW and CAT that mewing cats are hungry, but it obviously does not. In other words, all what one could ever believe about the mewing cat should follow from the meanings (i.e., inferential roles) of the cat and to mew, the conclusion that seems unacceptable. Fodor and Lepore argue that an inferential role theorist could modify one's theory by stipulating that only analytic inferences are part of the concept's meaning but not all its inferential roles. Inference from the mewing cat to the mewing animal is analytic, since it is based on the (putatively) analytic truth that cats are animals. This inference is compositional, because the mewing animal comes from two inferences: the mewing cat (mewing and cat (animal). Does this move represent a solution for the inferential role theory that is both able to explain compositionality as well as to avoid holism? Fodor and Lepore do not think so. They argue that the inferential role semantics has to allow "lexical analyticity" that comes from the meanings of terms (viz., the inference: the mewing cat (animal), not "structural analyticity" that comes trivially from the structure of an expression (viz., the mewing cat (cat). To allow lexical analyticity is to show that the intuitive difference between inferences like the mewing cat (animal and the mewing cat (hungry is a principled difference. One way to do that is to claim that the former inference is analytic because it is part of the meaning of CAT that cats are animals and that it is not possible to possess the concept CAT without also possessing the

concept ANIMAL. The second inference, on the other hand, is synthetic, since HUNGRY is not a part of the meaning of the MEWING CAT. If the inferential role theorist stipulates in order to save compositionality that the meaning of an expression is constituted by expressions inferential relations that are lexically analytic, then he commits oneself to the analytic/synthetic distinction. Drawing on Quine's arguments, Fodor and Lepore claim that this move is unsuccessful, since there are reasons to believe that the distinction between analytic and synthetic inferences is unprincipled. So is the inferential role theorist back in the claim that all expression's inferential relations are meaning-constitutive. This claim leads directly to holistic consequences and does not allow to account for compositionality.

Although it is arguable whether Quine showed that the analytic/synthetic distinction is unprincipled (see Boghossian, 1994a, 1994b, 1997), I do not touch this exegetic question here, because my aim is to delineate Fodor's and Lepore's views about the inferential role semantics and to observe how the rejection of inferential role account is connected to Fodor's rejection of the prevalent theories of concepts, which I shall examine in the next sections. I would only like to point out that in a more extended treatment, this would be a crucial issue. Fodor's and Lepore's belief that Quine was right could be nothing more than faith, since there are arguments in the philosophical literature that Quine was not right after all or that his arguments are not relevant in the present context (see Devitt 1996, Chap. 1), but as far as I know Fodor and Lepore do not address those arguments. Reliance on the distinction between analytic and synthetic inferences would enable for the inferential role theorist to escape from Fodorian and Leporian traps and to maintain that some of concept's inferential relations indeed constitute its meaning without falling into holism.

As concerns the accusation that the inferential role theory is unable to explain compositionality, then I know at least one theory of concepts that is a kind of inferential (or conceptual) role semantics – Peacocke's possession condition theory – that explicitly explains compositionality. Peacocke's theory (Peacocke 1992) could be regarded as a kind of non-holistic inferential role theory, since it individuates concept by its possession condition which mentions the canonical role of concept in judgements. One possesses a concept if one "knows what it is for something to be its semantic value" according to Peacocke (1992:23), and this identification makes it available for him to provide referential explanation to Evans' Generality Constraint, that is, it provides the means for him to explain the recombability of concepts, the systematicity and productivity of thought. Peacocke's theory also respects Fodor's compositionality principle, since concepts are the constituents of complex contents for Peacocke and the individuation conditions for complex contents are derived from the individuation conditions of its constituents. If we have this in mind, we see that it makes one of Fodor's central claims in his book inconclusive. The claim is that inferential role semantics is unable to explain compositionality, but we have a good example of inferential role theory that refutes this claim.

3. Fodor's critique of the definition theory of concepts

The definition theory is the classical view. According to it, concepts are what their definitions say they are, and by exercising philosophical analysis one uncovers those definitions. It could also be construed as a kind of non-holistic inferential role theory: the relation between a concept and its defining concepts is inferential. Since, according to the definition theory, those inferential connections between concept and its constituents are constitutive, they are knowable a priori (pp. 69–70). The definition theory consists of an epistemic claim about concept possession, i.e., to have a concept is to know its definition, that is connected to the modal claim about the a priori status of such knowledge. The connection between a concept and its defining concepts is necessary and constitutive. For example, BACHELOR is defined as “unmarried male person”. According to the definition theory, it is necessary that bachelors are unmarried and male and it is constitutive to the possession of the concept BACHELOR that it entails the possession of its constituents: UNMARRIED and MALE PERSON.

As a version of the inferential role account, the definition theory owes a story about the principle that distinguishes the defining conceptual connections from the non-constitutive ones and one could argue that this principle is pretty difficult to find. Fodor, however, is not going to run this line of argument, but centres instead on a weaker claim that the proponent of the definition theory could use in arguing for the existence of constitutive conceptual connections: it is intuitively plausible that some true sentences are analytic and the source of our analyticity intuitions is the fact that there are intrinsic connections among concepts (see p. 71).

It is important to notice that apart from his general antipathy towards and arguments against inferential role semantics, Fodor does not offer any arguments directed against the definition theory of concepts as such. He presumes that Quine's critique of the analytic-synthetic distinction was decisive and declares that it is wrong to assume that there are constitutive connections between concepts. He does not argue against the way definition theory connects concept individuation and concept possession, he simply presents an alternative, informational account where these two are separated and gives us a story about why we mistakenly believe that there are analytically true sentences. His claim is that philosophers are not interested in definitions as such, they just use them to explain analytic sentences and a priori truths. If it could be possible to give an alternative explanation to analyticity, there would be no need to invoke definitions. Fodor's strategy is therefore to show that analyticity is an illusion, and since the illusion could be explained away without invoking conceptual connections, there arises no need for definitions, “and, if there are no definitions, then there are no definitions for concepts to be.” (p. 87). In other words, he attempts to uncover a source of our analyticity intuitions that does not imply the existence of intrinsic conceptual connections. If there is such a source, then the argument from analyticity intuitions to constitutive conceptual connections is undercut.

Before going to Fodor's alternative story I would like to point out two problems with his tactics. First, it is clear that if this particular argument fails, then the definition theory is not yet proved to be wrong. There may be other ways of arguing for constitutive conceptual connections than by drawing on analyticity intuitions. Second, the claim that philosophers have not felt any intrinsic interest towards definitions, even if true, does its job in devaluating definitions but it does not show that there are no definitions or that there could not be any definitions. Let us consider an analogy. From the fact that we do not need the hypothesis of God's existence to explain the movement of planets, it does not follow that God does not exist. Even the contrary could be true. From the fact that at some point in the history we have felt the need to invoke God's existence to explain the movement of planets, it does not follow that God exists. There is a kind of realist intuition behind these examples that separates metaphysical issues from our particular theoretical interests, but I believe that this is justified in present context insofar Fodor is committed to (that kind of) realist views and that he actually is should be evident from his informational account of content.

Let us go back to the analyticity intuition argument for the definition theory. What is Fodor's alternative explanation to analyticity? In explaining analyticity intuitions, he makes use of the Putnam's idea of one-criterion concept (see Putnam 1983). This is the concept for which there is only one way of deciding whether the concept is applied correctly. Fodor modified use of this idea is the following:

suppose you think that the only epistemic route from the concept C to the property that it expresses depends on drawing inferences that involve the concept C. Then you will find it intuitively plausible that the relation between C and C* is conceptual; specifically that you can't have C unless you also have C*. And the more you think that it is counterfactual supporting that the only epistemic route from C to the property it expresses depends on drawing inferences that involve the concept C*, the stronger your intuition that C and C* are conceptually connected will be. (p. 83.)*

That is to say that we have analyticity intuitions in case of so-called one-criterion concepts, in which case we have only one method of deciding whether the concept is correctly applied. This helps to explain why we have so few analyticity intuitions. In most cases there are various mechanisms for gaining "semantic access" and not all of them involve inferences containing other concepts. To borrow the example from Fodor (p. 82): although "water is H₂O" is necessary, it is not generally believed to be analytic. The reason is that while there are many ways to determine that something is water, these ways generally do not consist in ascertaining first that it is H₂O and then inferring from H₂O to water. The situation is different in case of concepts like BACHELOR. The only way to determine that somebody is bachelor is to find out whether he is unmarried and male person. The concept BACHELOR is connected to bachelorhood only by inferences that involve concepts UNMARRIED and MALE PERSON.

Somebody might ask that why should we take this explanation as an explanation to our putatively mistaken intuitions of analyticity and not to the mechanism of analyticity itself. It seems that Fodor's reason for not taking this as an explanation to analyticity phenomenon is the fear of circularity. There has to be a way for individuating "epistemic routes" or criteria, but this would make use of the notion of analyticity again. The explanation of intuitions, on the other hand, does not have to presuppose a story about the individuation of "epistemic routes" (p. 82).

So, Fodor's point is that analyticity is an epistemic property: we believe that there are analytical connections, if we have such concepts that we have only one way to decide that they apply. Analyticity is not a semantic property, it does not arise from the conceptual connections between concepts and their definitions. This is a crucial point: if Fodor is right and the source of analyticity intuitions for inferences containing certain concepts is the lack of different ways of the application of those concepts, then we have alternative explanation to our intuitions. It means that the definition theorist cannot justify the introduction of intrinsic conceptual connections by claiming that it is the only explanation for analyticity intuitions. In the lack of separate justifications and given the critique of intrinsic connections among concepts, the definition theory seems to rest on a rather shaky ground.

Nevertheless, because of the above-mentioned problems with Fodor's strategy, I am not convinced that Fodor has succeeded to show conclusively that the definition theory is mistaken and also that there are no definitions. He argues that abandoning definitions should be easy, since there are practically no examples of definitions, except perhaps BACHELOR (p. 45). It appears, however, that there are whole classes of concepts that are best explained in terms of the definition theory. There are, for example, some terms in logic, mathematics and physics that are introduced by giving their definitions. These are concepts which content is strictly determined by their definition, but they could be such that the only mechanism of semantic access we have is theoretical inference. Such terms could be an example of concepts that are individuated by inferences to their definitions. The conditions of the possession for these concepts should also mention their conceptual connections given by the definition, so that one could not master these theoretical concepts unless one has grasped the definitions by which they are introduced and individuated. Even if the definition theory is not correct as a theory of all the concepts, it could be correct with respect to some subclass of concepts. It is a separate question of why we should expect to have one single theory for all kinds of concepts without respective qualifications if the class of concepts is not uniform and this question could be quite unpleasant for Fodor, since his informational atomism that is based on the uniform locking relation between concept and its object does not allow necessary qualifications. I will return to this issue in the final section.

4. Fodor's critique of the feature theory of concepts

Fodor credits the feature theory with the status of being the standard view on concepts among cognitive scientists. Maybe this is the reason why his presentation of the feature theory remains sketchy and fails to distinguish between different type of feature⁴ theories (For an overview of different types, see Eysenck and Keane 1990). In what follows, I attempt to give a short summary of this theory and consider Fodor's arguments against it.

According to the feature theory, concepts are bundles of properties and features that are related to each other. Those features form the basis of inference whereby it is possible to judge whether the given feature is individuating with respect to the given concept. The feature theory and the definition theory differ in what they consider important in concept-constitutive inferences. In the definition theory, the concept-constitutive inferences were necessary, whereas for the feature theory, the inferences involving given features are constitutive for a given concept if they are statistically reliable (p. 89). A given inference is reliable in Fodor's use of the term, if it assigns a feature to the concept which extension usually has this property, or in other words, that the probability of concept's extension having his property is high (p. 92). For example, the necessary condition of x's being snow is not x's being white. But usually snow is white, or that it is highly probable that if x is snow, then x is white. Therefore, "being white" is a feature of the concept snow and one could take the inference from "is snow" to "is white" as constitutive of the concept snow.

Concepts and their features could be organised into the hierarchical tree structure in which the subordinate concept has the features of its dominating or superordinate concepts (an example of such a structure is given at p. 90). The main idea of the feature theory is that one's having a concept amounts to one's having the mental representation of such conceptual hierarchy. When the hierarchy is mentally represented, there are two things that are operative in the mastery of concept – basic level concepts and prototypes. The concepts that belong to the basic level are those whose features are most differentiating with respect to the hierarchy tree. Let us imagine a branch of concepts from superordinate to subordinate: weapon—gun—rifle. To say of something that it is a gun is more informative than to say that it is a weapon, and less informative than to say that it is a rifle. This is because a basic level concept (GUN, in this example) inherits less features from its superordinate level than it transfers to its subordinate level. The prototypical representative is defined in each conceptual

⁴ I will call it the "feature theory", since it stresses the importance of features in the concept individuation. Fodor terms it both "statistical theory" and "prototype theory", but I believe that the first name is inconsistent with the way the definition theory was named—following Fodor's usage, one should have called it the modal theory of concepts. The second name is simply misleading, and this has consequences to Fodor's argument: contrary to appearances, the theory doesn't say that concepts are prototypes.

level by its relation to its family (or, as Fodor says, its sisterhood). A given concept is a prototype if its feature bundle has the most representative sample of features than any other concept in the given family in the given level. It may be said that prototypes have a central role to play in thinking and there is experimental evidence for the idea that concepts have this organised hierarchical structure and that there are prototypes. Prototypes are concepts that come to mind first when one thinks about some conceptual categories; prototypes are most frequently used concepts; out of the class of concepts, prototypes are among the concepts that are learned in the beginning of the conceptual development (p. 93). It seems to be a common view in the cognitive science community that if one possesses a concept one possesses its prototype as well.

Fodor is critical to this view and he claims that since concepts do not have an internal structure (according to his atomist theory), they cannot be prototypes. His argument is that the feature theory fails because it cannot satisfy Fodor's third adequacy condition for the theory of concepts. In a nutshell, the argument is the following. Concepts are systematic and productive. These properties of concepts are explained by compositionality, therefore concepts have to be compositional. Prototypes are not compositional, therefore concepts cannot be prototypes. Let us examine his argument more closely. Compositionality is the property of complex concepts that requires that the meaning or content of complex concepts is derived from the contents of their constituents. Fodor's two arguments are that

"indefinitely many complex concepts have no prototypes; a fortiori they do not inherit their prototypes from their constituents"

"there are indefinitely many complex concepts whose prototypes aren't related to the prototypes of their constituents in the ways that the compositional explanation of productivity and systematicity requires" (p.100.)

Fodor's example to the first argument is what he calls a "Boolean" concept such as NOT A CAT. It is a complex concept and does not have a prototype, although its components have prototypes. Anything that is not a cat, e.g. DOG, could serve as its prototype, but this cannot possibly be the case according to Fodor (p. 101), since it would follow that the more similar something is to a dog the less prototypical it is for a cat, and the more something differs from a cat, the more similar it is to a dog. The last conjunct is false, since if something is not a cat, it is not thereby similar to a dog.

Fodor's example to the second argument is PET FISH. The prototype for PET and the prototype for FISH do not compose together the prototype for PET FISH. At the same time, something could be a prototype for PET FISH without being the prototype for either of its constituents. If prototypes were constitutive to meaning, then it would follow according to Fodor (p. 102) that "you could know what "pet" means and know what "fish" means and still not know what "pet fish" means." He concludes that "if meanings were prototypes, then the meaning of "pet fish" wouldn't be compositional" (p. 102).

Apart from those arguments against the view that concepts are prototypes that draw on particular examples, Fodor has also an argument against feature theories that is more general. He claims that as is the definition theory, so is the feature theory also a kind of inferential role theory. Feature theories consider those inferences that have high statistical reliability as constitutive to the content of concept. Fodor's argument is that if the content of concept is individuated by its inferential role, then one could not explain compositionality, since inferential role itself is not compositional. (For a more thorough exposition of this argument see section 2).

Fodor's critique of the feature theory rests on the idea as if feature theories say that all concepts are prototypes and he attempts to show that they could not be. As I have understood feature theories, they say no such thing. They claim that there are prototypes for classes of concepts and prototypes are the common type of concepts we use. It appears that feature theories even cannot hold that all concepts are prototypes, for if they were we could not have any prototypes, since being a prototype implies the existence of concepts for which it is a prototype. If all concepts are prototypes, then we would not have prototypical concepts, but we would have prototypical prototypes, i.e., prototypical members of the class of prototypes. And this is not what the feature theory maintains. So it seems that the first part of Fodor's critique against the feature theory misses the target. The second, more general part of the critique will cling on the success of the argument that no inferential role account could account for the compositionality. If it one could develop a version of inferential role theory that could accommodate compositionality and that is consistent with the feature view on concepts, then Fodor's critique would be refuted. As it was pointed out in the section 2, we have a case of such an inferential role theory at hand – Peacocke's possession condition theory of concepts.

5. Fodor's critique of the theory theory

The theory theory is an account of concepts that has come up rather recently, especially in connection with the explanation of the conceptual development in the early childhood. In essence, its idea is that concepts are embedded in the theoretical framework, or even that concepts are like Kuhnian theories (or paradigms) in the respect that during the child's conceptual development there occurs a paradigm shift, i.e., old theories (and concepts they embed) are replaced by new theories whereas new and old theories are incommensurable (see pp. 113 114).

The theory theory is for Fodor not a distinct account of the individuation of concepts. It is usually mixed with the definition or feature theory, and it is clear that the theory theory is a variant of the inferential role account, because it assumes that the inferential relations in the child's primitive theory somehow implicitly define its concepts (p. 114). Fodor reproaches theory theorists for not

giving and even not attempting to give any specific account of which inferential relations in child's theory could be used to implicitly constitute child's concepts. Some theory theorists are happy to acknowledge that all inferences allowed by particular theoretical framework constitute its concepts. This seems to lead directly into unfavourable holistic consequences: no two persons (and the same person at different times) can share the same concept; the same intentional generalizations do not apply to people who have different conceptual frameworks; even stronger, if the individuation of concepts depends on the whole theory and this is usually taken to mean the set of all beliefs, then even for two persons who share the same conceptual framework, it suffices for a slightest change in one person's belief and they do not share the same implicit theory anymore nor are intentional generalizations applicable to both of them. Fodor complains that theory theorists hold this holistic theory without attempting to answer its consequences. In a way, one could understand this, because people who have mostly supported theory theory are not philosophers, but developmental psychologists who do not oblige themselves to answer philosophical criticisms. They are occupied with other issues and Fodor's critique that they do not present well-based philosophical accounts is even unfair with respect of developmental psychologists, since they even do not intend to do philosophy. Moreover, if we restrict the application of the theory theory to infant and child cognition, then it could be the case that there is such a difference between adult concepts and child's concepts that holism is true for the latter but not for the former. If we trust developmental psychologists that there occurs a fundamental shift in the child's theory of mind around the age of 4, it could be true that child's concepts lack salient individuation conditions and that they make sense only in the light of the child's whole theory. It is possible to add some plausibility to theory theory by not treating it as a theory of concepts for adult subjects.

While the argument from the holistic consequences is already well-known, Fodor's second argument has some novelties. Fodor (p. 115) makes a fair point that the holistic individuation is not compatible with theory theory's reliance on implicit definition for concepts. The point is that a concept is implicitly defined if its meaning is determined by its connections with the terms and concepts already defined in the theory. The meaning of defining terms must not change if the concept that is defined is brought into the theory. Otherwise they could not give their contribution to the implicit definition, because their meaning was not fixed. This happens when one tries to couple implicit definition with meaning holism (to be understood in the above-mentioned sense). For it follows from holism that the introduction of a new concept to the theory will result in the change of meaning of those terms that are inferentially related to that new concept, and it is precisely those terms that we need in order to have implicit definition for a new concept. So, holism and implicit definition are incompatible.

Of course, the theory theorist could abandon the idea of the Kuhnian paradigm shift with its holistic consequences and maintain only the idea that concepts are

implicitly defined by some terms in the theoretical network. In that case, as Fodor points out, the theory theorist has to give a principled basis for deciding which terms or inferential relations are implicitly constitutive for a concept. Such a theorist either accepts one of those views that Fodor has already criticised⁵ (see sect. 2–4) or comes up with a new account of concept individuation that accommodates the idea of implicit definition. However, no such account has been given yet, therefore Fodor does not devote much time to this possibility. It could be concluded that insofar theory theory is not a sufficiently elaborated account of concept individuation, it is not an adequate theory of concepts and this is one of those few points in which I agree with Fodor.

This concludes this part of the essay where I attempted to review Fodor's arguments against theories of concepts that in some way or other were all based on the inferential role semantics. As for conclusion for this part, I have to confess that Fodor has not convinced me that these theories are mistaken. He and Lepore claim that inferential role semantics in general either leads to holism or stumbles in explaining compositionality. Definition theory, on the other hand, allows compositionality since the concepts in the defining inference compose the concept that is defined. However, Fodor argues that there are no definitions. The problem for the feature theory is the converse. There are lot of evidence for prototypes, but prototypes are not compositional (Cf. p. 107). Third, the theory theory, is not a proper theory of concepts at all to deserve extended treatment from Fodor. Against this I have argued that there are viable non-holistic inferential role accounts that could explain compositionality; that there are classes of concepts for which the definition theory is correct; and that it is non sequitur to infer from the non-compositionality of prototypes that the feature theory is incorrect. I agree with Fodor that the theory theory has problems because it is not elaborated enough, but I believe that one could alleviate the force of Fodor's critique if one does not take the theory theory as an account of adult concepts. Given this, I am led to conclude that Fodor's critique is far from being decisive. But perhaps Fodor's own theory is so irresistible that we cannot help but to accept it? The next section attempts to find this out.

6. Informational atomism as a theory of concepts

"Informational atomism" is the Fodor's alternative account and as its name suggests, it consists of two main ideas. First, content is explicated in terms of information. Content expresses properties in the world and this is effected by the nomic "locking" relation between the mind and the world. Fodor has developed

⁵ Whether this critique is decisive is another question and if it is possible to sustain and make sense of the distinction between content-constitutive and non-constitutive definitions or inferential relations, then the theory that conjoins implicit definitions with the possibility of aforementioned distinction is very likely to be viable.

the informational theory of content over some years already (see Fodor 1987, 1990b, 1994) and its essential idea is that the content of mental representation depends constitutively on those properties with which the occurrence of the representation is nomologically covaried. This is roughly how the content is individuated according to the informational theory. The individuation of content, however, is not all that could be said about the individuation of concepts (p. 15). If concepts would be individuated by their contents, then it would have the consequence that all coreferential concepts would be the instances of the same concept. For example, according to the informational theory, concepts WATER and H₂O covary with the same property, since the property of being water and the property of being H₂O are necessarily coinstantiated. Fodor concedes that the content of these concepts is indeed the same and it means that they are synonyms. This does not imply that they are the same concept, since concepts are individuated both by their referents and the mode of presentation of those referents according to Fodor. The mode of presentation is in the form of the syntactically structured mental particular. Accordingly, coreferential concepts have the same conditions of semantic evaluation, but they differ in their syntactic structure that could have different causal effect in one's psychological architecture. That is why different concepts have different roles in our thinking.

The second main idea of the informational atomism is the thesis of conceptual atomism: "most lexical concepts have no internal structure" (p. 121). Atomism follows from the story about the constitution of concepts. If the content of the concept CAT is constituted by causal and nomic relation between the mental/neural tokens of CAT and the instantiations of the property cathood in the environment, then all there is to the possession of the concept CAT is the existence of the respective mind-world connection. As far as I can understand Fodor, this is a metaphysical matter: a mind is either locked onto the property cathood or it is not. This does not carry epistemic commitments with itself. One need not possess any other concepts to possess the concept CAT. Somebody might ask: given the informational atomism with respect to concepts, there are undeniable necessary connections between the properties in the world. Do these connections constrain the concept possession in the sense that in order to have a concept, one must also have those concepts that denote properties that are necessarily connected with the property onto which the formerly mentioned concept is locked? Fodor (p. 14) argues that informational atomism allows conceptually necessary truths but this is not connected with the epistemic conditions of concept possession. This is the central idea of the book: in order to possess a concept one is not thereby committed to possess some other concepts, to master certain inferences or to have knowledge of some theory or analytic truths. An example is needed (see p. 14): assume that "bachelors are unmarried male persons" is a necessary truth, so there is a necessary connection between bachelorhood and the property of being unmarried male person. The concept BACHELOR applies only to bachelors and ipso facto to unmarried male persons.

It does not follow, however, from this that only when one possesses the concepts UNMARRIED, MALE and PERSON, could one possess BACHELOR.

The informational theory is a metaphysical account of content constitution. It says that there is a lawful connection between a property and a mental/neural structure. Fodor assumes that this nomological connection must be sustained by some causal mechanism and that informational atomism also has to specify how does the instantiated property in the environment cause the tokening of the concept. Fodor (p. 75) names any such specification "a mechanism of semantic access". Anything that effects the tokening of a concept could be regarded as the mechanism of semantic access. The most obvious candidate is perception, or as Fodor (p. 76) expresses it: "[I]f you put a DOG-owner, eyes open, in a dog-filled environment and if you turn up the lights, dog-thoughts will ensue." But there are many mechanisms that mediate one's semantic access to properties. Theoretical inferences, linguistic division of labour, technological means, etc., are also among those mechanisms that could call forth one's thoughts. Since perception is just one mechanism among the many, there is no necessary connection between perception and concept's content according to informational atomism. Although it might turn out that all other mechanisms are dependent on the background of perceptual capacities, it is at least possible that someone who lacks certain particular perceptual capacities could nevertheless be able to entertain contentful thoughts (p. 79).

Fodor is keen to mention that those mechanisms of semantic access neither constitute concept possession nor determine the content of concepts. That one possesses a particular concept is constituted by the fact that there is a nomological relation between the instantiated property and one's particular mental structure. That one's concept has just the content that it has is determined by the fact that it is just that kind of property that the tokenings of one's concept are locked onto and not some other kind of property. The mechanism of semantic access only specifies how this connection might be sustained.

It does not seem very explanatory to say that there is a nomological relation between a particular mental/neural structure and the world mediated by the mechanism of semantic access. What kind of a causal law it is that connects the mind and the world? How are we to make sense of the parallelism between the syntactic mode of presentation and the informational content? Let us see what Fodor himself has said:

The world thus sustains a harmony between the (extrinsic/historical) properties in virtue of which a Mentalese sentence is a mode of presentation of the proposition that P and the (intrinsic/syntactic) propositions in virtue of which a Mentalese sentence is reliably a cause of the sorts of behavioural proclivities that the laws of psychology say that P-believers share. /.../ [C]omputational-syntactic processes can implement broad-intentional ones because the world, and all the other worlds that are nomologically nearby, arranges things so that the syntactic structure of a mode of presentation reliably carries information about its causal history (Fodor 1994: 53-54).

This could be seen as Fodor's attempt to answer his first and fifth condition on concepts. It may seem almost impossible to have concepts that are both public and able to interact causally with computational/mental processes. Fodor's postulated harmony between extrinsic and intrinsic properties of mental representation helps to make sense of that, but at the same time it leaves a lot of questions unanswered. In virtue of what there is such an harmony? The mechanism of semantic access tells us how the mind-world connection was furnished, but what guarantees the harmony? The mechanism of semantic access is neither necessary nor sufficient for the mind-world connection to hold. And what about the normative aspect of concepts, is it satisfactorily captured by the information theory?

Besides semantic value, concepts also have cognitive significance. Even if informational atomism provides an account of concept's semantic value, its account of cognitive significance leaves much wanting. The mode of presentation of a concept modelled as bare syntax does not capture the cognitive significance that is exemplified by the Fregean informativeness of coreferential identity-sentences. It could be the case that Fodor is going in the right direction: if we aim at the naturalised theory of concepts, we need the naturalistic account of the realisation of cognitive significance. But it seems that the mere syntax is not simply enough. Apparently Fodor thinks that we do not need to explain the cognitive significance, since there is no such thing: "The older I get, the more I am inclined to think that there is nothing at all to meaning except denotation; for example, that there is nothing to the meaning of a name except its bearer and nothing to the meaning of a predicate except the property that it expresses." (Fodor 1990a:63.) If so, then his theory must be unattractive to those philosophers who take the cognitive significance to be a well-established notion. It is clear that Fodor's informational atomism is a theory of concepts at the level of reference. When a certain syntactically individuated mental structure is tokened in one's architecture of mental representations in virtue of some mechanism of semantic access, then one entertains a certain concept. Nothing more is needed, especially nothing epistemic. Such an account seems to underestimate the role concepts have in one's thinking. At a certain subpersonal level, it may be correct to describe thinking as computation, but it looks rather obvious that one could also describe and explain concepts at the personal level. And this is where the notion of cognitive significance has its grip. Even if it really is the case that in order to have the concept one needs only proper mind-world connection, one needs to explain also what one has to know in order to master the concept, that is, to use it correctly in the reflective thinking. If thinking is more than the simple recombination of mental representations, that one has to leave some space for conceptual connections in his theory of thinking, then one needs an account of concepts that makes use of some notion of cognitive significance. Fodor's theory of concepts that is constrained by the representational theory of mind does not give us such an account.

Finally it may be doubted whether the informational atomism is substantive, i.e., could provide explanation to all kinds of concepts. This theory seems to equalise all

concepts in the sense that there is a property for every concept that carries information about it. Or could there be any concepts that do not carry any information at all? Not for Fodor. Let us have a look at the following concepts: CAT, QUARK, NOT A CAT, RED, PINKISH-LOOKING, ELM, SNOW, GOLDEN MOUNTAIN, IS HAPPY. Of course, it is possible to construe properties that correspond to each aforementioned concept, but this does not say much about those concepts. It would be more informative to say what are the differences between those concepts, and whether they could be divided into classes on the basis of these differences. It is likely that the differences between these concepts are of an epistemic not of a metaphysical kind. Fodor would obviously try to explain the difference in terms of the different semantic access to respective properties. Perhaps it could be possible, but in any case, he has not shown this yet.

One has to conclude that on the one hand, Fodor's arguments against inferential role semantic are not successful enough to tempt us into accepting his informational atomism. On the other hand, the informational atomism is not attractive enough that it would justify one to abandon inferential role semantics.

Acknowledgements

An early version of this essay was read at the seminars of the Estonian Society of Analytic Philosophy in September 1998. I am grateful to the participants in that seminars for interesting discussion. I am also extremely grateful to Martin Davies and Christopher Peacocke for helpful discussions of various aspects in Fodor's philosophy of content.

Address:

Mõisavahe 26–15

Tartu, 50707

Estonia

Phone: +372 (7) 486 230

e-mail: bruno@physic.ut.ee

References

- Block, Ned (1993) "Holism, Hyper-analyticity and Hyper-compositionality". *Mind and Language*, 8, 1, 1–26.
- Boghossian, Paul (1994a) "Does an Inferential Role Semantics Rest Upon a Mistake?" *Mind and Language*, 8, 1, 27–40.
- Boghossian, Paul (1994b) "Inferential Role Semantics and the Analytic/Synthetic Distinction". *Philosophical Studies*, Vol. 73, pp. 109–122.
- Boghossian, Paul (1997) "Analyticity". In *A Companion to the Philosophy of Language*. Crispin Wright and Rob Hale, eds. Oxford: Blackwell, 331–368.

- Devitt, M. (1996). *Coming to Our Senses: A Naturalistic Program of Semantic Localism*. Cambridge: Cambridge University Press.
- Eysenck, M. W. and M. T. Keane (1990) *Cognitive Psychology: A Students Handbook*. Lawrence Erlbaum Association.
- Fodor, Jerry (1987) *Psychosemantics. The Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass.: The MIT Press.
- Fodor, Jerry (1990a) *Substitution arguments and the individuation of beliefs*. In *Meaning and Method*. George Boolos, ed. Cambridge: Cambridge University Press, 63–78.
- Fodor, Jerry (1990b) *A Theory of Content and Other Essays*. Cambridge, Mass.: The MIT Press.
- Fodor, Jerry (1994) *The Elm and the Expert. Mentalese and Its Semantics*. Cambridge, Mass.: The MIT Press.
- Fodor, Jerry (1998) *Concepts: Where Cognitive Science Went Wrong*. Oxford: Clarendon Press.
- Fodor, Jerry and Ernest Lepore (1991) "Why Meaning (Probably) Isn't Conceptual Role". *Mind and Language*, 6, 4, 328–343 (Page references in text are given after the reprint of the article in *Mental Representation: A Reader*, S. Stich and T. Warfield eds, 1994. Oxford: Blackwell, 142–156.)
- Loewer, Barry and Georges Rey (1991) "Editors' Introduction". In *Meaning in Mind: Fodor and His Critics*. Oxford: Blackwell, xi-xxxvii.
- Peacocke, Christopher (1992) *A Study of Concepts*. Cambridge, Mass.: The MIT Press.
- Putnam, Hilary (1983) "Two dogmas' revisited". In *Realism and Reason: Philosophical Papers*. H. Putnam, ed. Vol. III. Cambridge: Cambridge University Press.