

## **From the pragmatics of classification systems to the metaphysics of concepts**

Jerry A. Fodor. *Concepts: Where Cognitive Science went wrong*. Oxford, UK: Oxford University Press, 1998, 174 pp., ISBN 0-19-823636-0.

Geoffrey C. Bowker and Susan Leigh Star. *Sorting things out: Classification and its consequences*. Cambridge, MA: The MIT Press, 1999, 377 pp., ISBN 0-262-02461-6.

*Commentary by*

***Stella Vosniadou, Costas Pagondiotis and Maria Deliyianni***

Graduate Program in Basic and Applied Cognitive Science

Department of Philosophy and History of Science

National and Kapodistrian University of Athens

Concepts and categories are the building blocks of cognition. They allow us to interpret our experiences, to connect them to prior knowledge, to reason and to make predictions. One could not imagine a theory of learning without a theory of concepts, and sure enough psychologists have been concerned about concepts and categories for a long time now. As we all know, there are three main psychological theories about concepts. Following Smith and Medin (1981), we can distinguish the *classical view*, i.e., that concepts are defined by certain necessary and sufficient properties, from the *prototype and exemplar views*, i.e., that concepts are represented by a prototype or a specific exemplar. More recently psychologists have developed various *theory-based views* according to which concepts are embedded in theory-like structures that constrain them.

Geoffrey Bowker and Susan Leigh Star in *Sorting Things Out: Classification and Its Consequences*, and Jerry Fodor in *Concepts: Where Cognitive Science Went Wrong*, come to tell us that these three approaches to concepts and categories are completely wrong and that we should throw them in the waste basket. Even more interesting is the fact that they come to this radical conclusion from completely different points of view, from the most pragmatist to the most cognitivist.

In *Sorting Things Out*, Bowker and Star, want to look at categories and classifications from a situated, pragmatic, point of view. The book stands, according to its authors, “at the crossroads of the sociology of knowledge and technology, history, and information science” (p. 6). “We want to know empirically how people have designed and used classification systems. We want to understand how political and semantic conflicts are managed over long periods of time and at large levels of scale” (p. 53). Jerry Fodor, on the other hand, is not interested in how concepts are used but in the metaphysical question of how concepts are individuated. In *Concepts* (a book which, despite its subtitle, is deeply committed to the cognitive science program), he presents a powerful critique of all dominant cognitive theories on the nature of concepts, ending with the development of a new theoretical proposal about how concepts are formed, which he calls Informational Atomism (IA).

In this commentary, we will look first at Bowker and Star’s pragmatics of classification. It will be argued that the authors have succeeded in showing that

classification systems are ubiquitous and have a material force that can affect people's lives, but that they have failed to persuade us that categories are conventions that arise solely from situated experience. Continuing with a discussion of some of Fodor's proposals about the way in which concepts may be acquired, we will try to show that his theory leaves many questions unanswered. The usual psychological approaches, we will argue, are still alive and well and have a significant role to play in a theory of concepts. Nevertheless, we will admit, these two books point towards new, promising, and creative ways of looking at concepts that have great potential for enriching our understanding of learning.

### **A pragmatist analysis of classification systems**

Pragmatism is the philosophical school that pays attention not to the logical antecedents of an argument but to its consequences (Dewey, 1929). In the social sciences this has been interpreted as a turn away from people's definitions of a situation to the consequences of the situation. Following the pragmatist turn, Bowker and Star bypass the usual definitions of *classifications* and *standards*, to focus instead on those characteristics that clarify the role that classifications play in people's lives.

Classifications, the authors claim, are *symbolic as well as material, ubiquitous but at the same time invisible* – “enter a modern home and you are surrounded by standards and categories spanning the color of paint on the walls and in the fabric of the furniture, the types of wires strung to appliances, the codes in the building permits allowing the kitchen sink to be properly plumbed and the walls to be adequately fireproofed” (p.1). “These standards and classifications, however imbricated in our lives, are ordinarily invisible” (p.2). Bowker and Star are also interested in the *social and moral implications of classifications*.

Given the above, the purpose of a pragmatist analysis of classification, ranges from methodological and theoretical, to political and ethical – how are decisions about classification made and how they affect people's lives. The authors often make reference to the “new and electronic infrastructures” but in reality they do not say much about electronic forms of classification and their effects. The book is based

instead on a detailed examination of the construction and evolution of a massive classification system, the International Classification of Diseases (ICD), in Part I, and the design of a nursing classification system, the Nursing Interventions Classification (NIC), in Part III. Part II investigates the classification systems of tuberculosis patients and of race in South Africa under apartheid.

The message that comes from Parts I and II is twofold: first, that relatively simple classification schemes cannot capture the complexity of real life producing innumerable contradictions-- “when the work of classification abstracts away the flow of historical time, then the goal of standardization can only be achieved at the price of leakage in these classification systems” (p.168). Second, that in the process “of making people and categories converge, there can be tremendous torque of individual biographies” (p. 225). In Part III, a number of design parameters are discussed, and the authors conclude that designers “develop an economy of knowledge that ensures that all and only relevant features of the object being classified are remembered” (p. 281).

Despite the pragmatist turn, the authors cannot altogether ignore traditional psychological approaches to concepts and categories. In fact, it turns out that these approaches are fundamental in order to best capture the effect that classification systems have in our lives. In the second chapter, Bowker and Star engage in an extensive discussion of the different kinds of classification systems in theories of classification, as they try to capture the interplay between, what they call, the “formal and the informal” or better, the structural and functional aspects of classification. Drawing on John Taylor’s (1995) work in the area of sociolinguistics, they conceptualize this “not so absolute” distinction in terms of the classic divide, so well known to psychologists, between classical or “Aristotelian” and prototype theories of concepts. According to Bowker and Star, an “ideal” Aristotelian classification “works according to a set of binary characteristics that the object being classified either presents or does not present,” whereas in Rosch’s prototype theory “our classifications tend to be much fuzzier than we might at first think” (p. 62).

The text is often ambiguous here, making it difficult to understand whether Bowker and Star believe that classification systems are “in principle Aristotelian”(p. 64) or

“naturally prototypical” (p. 106). It appears that what they want to say is that classification systems usually impose a rigid, Aristotelian, structure on the world, but the world does not behave in this way. In real life, “the classical beauty of the Aristotelian classification gives way to a fuzzier classificatory system that shares in practice key features with common sense prototype classifications – heterogeneous objects linked by metaphor or analogy.” (p. 65).

This interplay between the formal and the informal, the Aristotelian and the prototype, becomes a major theme in the book. For example in Part I, the analysis of the ICD as a classification system focuses on the “implicit narratives” that must be embedded within an Aristotelian classification in order to change it so that it will be able to do its job; in other words, to provide the gateway between the precisely defined world of the laboratory and the workaday world. Similarly, in Part II, the examples provided highlight the ethical and moral problems that arise when rigid classification systems such as apartheid are applied to people. It thus turns out that the psychological theories are useful after all in a pragmatist’s analysis of classification.

### **Extending pragmatism to a theory of concept acquisition**

In the last two chapters of the book, Bowker and Star attempt to sketch a pragmatist theory of concept acquisition. Drawing on previous work by Foucault, Durkheim and Latour, and by psychologists working in the tradition of situated cognition (Lave, Hutchins, Cole, Engeström, etc.), they try “to ground activities previously seen as individual, mental, and nonsocial, as situated, collective, and historically specific” (p. 288) Their basic argument is that categories are historical and political artifacts that arise from work and organized activity, that they are learned through participation in communities of practice, and that they are tied to the things people do.

One of the implications of the position that categories are learned as part of people’s membership in communities of practice is that categories are tied to each community’s particular usages and practical requirements. Such a learning process cannot, however, explain how general classification systems are created, or account for how communication between communities of practice is achieved. The authors

propose various ad hoc solutions in order to solve this problem. For example, they argue that the goal of information transmission across different communities of practice creates “shared objects” that are built across community boundaries, or “boundary objects” (Star & Griesemer, 1989). These boundary objects can be used to resolve anomalies and ensure communication. But, no detail is provided about how these boundary objects become created and what is their relationship to categories and classification systems. Star and Griesemer noticed, for example, that professional biologists gave a different meaning to specimens of dead birds than amateur bird watchers but “the same ‘bird’ was used by each group” (p.297). The authors conclude, “such objects have different meanings in different social worlds but their structure is common enough to more than one world to make them recognizable” (p. 297). If something, this example shows that these two communities somehow share a common concept of ‘bird’ regardless of their different social worlds, and not that a common concept is created as a result of their interaction.

The narrowly situational theoretical framework developed in the last chapters of the book has additional difficulties accounting for the formation of classification systems that are formal and Aristotelian. As mentioned earlier, one of the main points of the first part of the book is to show how classification systems, that are often formal, rigid and Aristotelian, come in conflict with the messy and fuzzy categories of everyday life, creating serious political and ethical problems. But, if categories arise solely from the fuzziness of a social reality that lacks the classical Aristotelian beauty, why don’t they reflect this fuzziness instead? Where do the formal and systematic aspects of categories come from?

Finally, Bowker and Star’s decision to base their perceptive pragmatic analysis on a radically situated and anti-mentalistic theoretical ground leads them to certain absurd positions, such as the argument, that there is no such thing as “abstract thinking” in mathematics. Mathematics, they argue, is rather “a process of assembling materials close to hand and using them with others in specific contexts.” And, they continue, “those who appear to solve mathematics problems without such outside help are not working in a putative realm of pure number; rather, they and their observers have so naturalised the structures within which they are operating that they have become invisible”(p. 288). The authors fail to consider the possibility that the internalisation

of external symbolic systems is not actually inconsistent with the manipulation of internal mental representations; in fact, it could actually account for the creation of abstract thinking processes.<sup>1</sup>

To conclude, Bowker and Star's analysis of classification systems succeeds in drawing our attention to the material aspects of categories and demonstrates that classification systems can be "powerful technologies" that influence people's actions and lives. *Sorting Things Out* captures the complex interactions between the conceptual and the material, and is persuasive about the materiality of symbolic artefacts like classification systems. But pragmatism does not have to come in conflict with more traditional psychological approaches concerned with definitional aspects of categories. On the contrary, the authors draw on the usual psychological theories of concepts to describe how classification systems are used and how they change over time. Finally, Bowker and Star's preliminary and sketchy attempts to provide a situated and historically specific account of the acquisition of concepts is inadequate and cannot explain phenomena such as the formation of common concepts across different communities of practice and the construction and widespread use of formal and systematic classification systems.

### A cognitivist critique of psychological theories of concepts

Fodor's *Concepts: Where Cognitive Science Went Wrong* consists of three parts. In the first, he presents the Representational Theory of Mind (RTM) as the background theory of cognitive science in the context of which a theory of concepts must be developed. Here Fodor makes five "non-negotiable" assumptions about concepts. That they are mental representations, that is, mental *particulars*; that they are categories, that is, things in the world 'fall under them'; that they exhibit compositionality, that is, they are the constituents of thoughts in the sense that the content of the thoughts is determined by the contents of their constituents and by the

---

<sup>1</sup>

<sup>3</sup> Editors note – For clarity we follow Fodor's convention of presenting names of concepts in all caps and thing signified by the name in italics.

way these constituents are combined; that ‘quite a lot of concepts are learned;’ and finally, that concepts are public and sharable.

In the second part of the book, Fodor argues against the three dominant cognitive theories of concepts, according to which concepts are considered to be definitions, prototypes and theories respectively. What Fodor has against these three theories of concepts is that they are “versions of one and the same idea about content...namely, that primitive concepts and (hence) their possession conditions, are at least partly constituted by their inferential relations.” (p. 35). The theory that content is constituted by inferential relations is known as Inferential Role Semantics (IRS).

Fodor is critical of IRS because it comes in conflict with RTM’s thesis that thinking is computation. More specifically, he claims that IRS cannot give a non-circular naturalistic account of mental representations because it attempts to explain how mental symbols acquire their content by recourse to their computational relations to other symbols while the very notion of computation is understood as “some kind of content-respecting causal relation among symbols”(p. 11). As he characteristically remarks, “for fear of circularity, I can’t both tell a computational story about what inference is and tell an inferential story about what content is” (p. 13), given that content is (or is given by) computation.

Fodor devotes a large part of his critique against the definition theory of concepts. He notes that there are practically no defensible examples of definitions, presenting various arguments to this effect. One of them is based on Quine’s attack on the analytic/synthetic distinction: if there is no un-question-begging way to draw a distinction between conceptual connections that are analytic and conceptual connections that are synthetic, then there cannot be definitions since the conceptual connections between the definientum and the definiens is supposed to be analytic.

Prototype theories, which are also instances of inferential role theories, seem to meet two of Fodor’s non-negotiable conditions – they are public and psychologically real – they do not meet the third condition, namely compositionality. Fodor argues that there are indefinitely many complex concepts that do not have prototypes (see, for example “Boolean” concepts, like NOT A CAT) and “a fortiori they do not inherit



their prototypes from their constituents” (p. 100). But even those complex concepts that do have prototypes, their prototypes aren’t composed from the prototypes of the constituent concepts: the prototype PET FISH is not composed from the prototypical FISH and the prototypical PET.<sup>3</sup>

Finally, the theory theory view of concepts is, according to Fodor, either neutral on the issue of concept individuation or it claims that knowing all or some of the theory is a necessary condition for having the concepts. The latter situation is problematic because it introduces a holism that makes it impossible to understand how two people can share the same concept. Fodor goes on to argue that the claim that holism about content individuation does not even “square with key principles of the theory theory itself” (p. 115). An implicit definition of a new term in a theory can only be effected if at least some of the theory’s vocabulary “is isolated from meaning changes of the sorts that holists say that concept introduction bring about” (p. 115).

What are we to make of Fodor’s critique of psychological theories of concepts? Maybe the first thing to remind our readers is that Fodor’s construction is tied to a very specific interpretation of RTM. Even for those who endorse the idea that thinking is computation, the argument that computation is done on mental representations that are concepts maybe debatable, as it is the case, for example, in connectionist networks where computation is done at a sub-conceptual level. Furthermore, objections can be raised regarding his accusations of IRS on the grounds of circularity. As we saw earlier, Fodor accuses IRS in that it attempts to explain representation in terms of computation, which in turn is defined in terms of representation. Supporters of IRS would retort here that the notions of representation and computation are interdependent and should be explained simultaneously.

Fodor’s arguments against prototypes and theories are serious, given that these two are the most widely held positions on the structure of lexical concepts held by psychologists. However, both in the case of prototypes and in the case of theories Fodor makes certain assumptions that are problematic. For example, in his discussions regarding compositionality, Fodor assumes that if the constituents of a complex concept have a prototype structure, then the complex concept itself must have a prototype structure as well. But, as Laurence and Margolis (1999) note, this

assumption is ironic for Fodor to hold given that his own theory of concepts does not abide by it. Complex concepts, in Fodor's account, acquire their content *compositionally* from the content of their constituents. Unlike their constituents, they do not exhibit any kind of causal locking to what they refer to.. Thus, in the same manner, a prototype theorist could claim that a complex concept acquires its content compositionally from the content of the prototypes –i.e. the primitive concepts– that constitute it.

Another line of argument is to argue that both the prototype and theory theory views should be taken into account in a complete theory of concepts but that they are not adequate by themselves to give an *exhaustive* account of the structure of concepts. Note that Fodor's criticism presupposes that both the prototype and the theory theory views are supposed to give an *exhaustive* account of the structure of concepts. But this does not have to be the case. Some researchers who support the idea that concepts are embedded in theories have argued that theories are not “all the way down” but develop later. According to Carey and Spelke (1996), for example, initial cognitive endowment consists of a set of domain-specific, core systems of knowledge organized around specific principles that have some but not all the properties of later developing intuitive theories. Such views do not of course answer the problem of reference determination, which is the problem that Fodor wants to address. Let us see what he wants to say about this problem.

### **On the view that lexical concepts are mental particulars causally related to the world**

In the third part of the book, Fodor develops his own proposal about concepts, namely Informational Atomism (IA). IA has two parts: *Informational Semantics*, “the claim that content is constituted by some sort of nomic, mind-world relation,” (p. 146) and the *Conceptual Atomism* part, namely, that most lexical concepts have no internal/constituent structure.

If most lexical concepts have no internal structure then they should be primitive. This means that even lexical concepts like DOORKNOB, should be primitive. However,

Fodor's argument in defense of the primitiveness of doorknobs is on shaky grounds. He argues that if DOORKNOB is not primitive, then it must have a definition. But, since linguists and philosophers had no luck so far to define DOORKNOB, it follows that DOORKNOB is undefinable, and since it is undefinable it is primitive. Fodor concludes that DOORKNOB is a primitive appearance concept that works very much like red (i.e., they are both mind-dependent).

The argument that most lexical concepts are primitive is problematic also because it implies a radical conceptual nativism. Primitive concepts must be unlearned and, therefore, must be innate. Thus, even concepts such as DOORKNOB, CARBURETOR, and PHOTON, being atomic, should be innate. This is a consequence embraced by Fodor in his previous works. Fodor was led to this radical conceptual nativism because of his position that learning a concept is an inductive process, that is, a process of devising and testing hypotheses. Fodor's new proposal tries to avoid this radical conceptual nativism by suggesting that concept acquisition is not always a cognitive process and, so, it does not admit of a psychological explanation (like hypothesis formation). Concept acquisition is not explicable in cognitive terms but only in causal terms: "having a concept is (not knowing something but) being in a certain nomic mind-world relation" (p. 124). Thus, primitive concepts can be acquired by some kind of *locking* of the relevant mechanisms of our brain to the properties that the concepts express. In other words, *nativism of concepts* is replaced in Informational Atomism by a *nativism of mechanisms* (see p. 142).

The argument that it is not the concept itself that is innate, but instead a mechanism which is part of an 'innate sensorium', is certainly much more palatable within psychological circles. This proposal brings Fodor much closer to the work of ethologists and psychologists like Gibson, ties nicely with domain specificity views, and is consistent with the views of many researchers who have argued against the innateness of representations. This does not mean that the proposal that "acquiring a concept is getting *nomologically locked* to the property that the concept expresses" (p. 125) is entirely meaningful or particularly illuminating. What exactly is this mechanism? How does it work? How exactly do you go from the specific instances to

the acquisition of the concept if you do not already have the concept?<sup>4</sup> How many concepts are supposed to be acquired in this way?

Even if we accept that there is something to the ‘locking’ idea, there are serious doubts that it is sufficient or even necessary for explaining the contentfulness of concepts. We have concepts of nonexistent and of abstract objects, which (objects), as such, have no causal powers and a fortiori cannot cause the locking of some unspecified brain mechanism. Does the instantiation of properties that are *socially* instituted have the relevant causal powers that Fodor postulates? How could there be, for example, “causal-cum-nomic relations between BACHELOR-tokenings [that is, tokenings of the concept BACHELOR] and tokenings of instantiated bachelorhood” (p. 14)? But even if we limit ourselves to objects and properties that do have causal powers, it is unclear how a causal mechanism that locks, say, to doorknobs leaves room for error. Fodor’s own attempt to deal with this problem –asymmetric dependence theory (Fodor, 1990) –certainly faces many problems (cf., Jacobs, 1997).

Fodor distinguishes three sorts of concepts: concepts of mind-dependent properties, natural-kind concepts and logico-mathematical concepts. About the latter Fodor admits that he has nothing to say (p. 151). We will say something about natural kinds later. Concepts of mind-dependent properties are primitive, and these are of two kinds: concepts of sensory properties, like red, and concepts of appearance properties that are not sensory, like doorknob.

Fodor, in a footnote (pp. 135-136, n.10), remarks that sensory concepts (like RED) are distinguished from the non-sensory appearance concepts (like DOORKNOB) in that we could experience a sensory property even though we did not have the relevant concept. This is because we have sensory organs that can “produce such experiences when they are appropriately stimulated.” On the contrary, the perceptual detection of a non-sensory appearance concept like DOORKNOB is, he claims, always inferential. This is puzzling: if the perceptual detection process is always inferential<sup>5</sup>, then how was “locking” possible in the first place?

---

<sup>4</sup> Fodor calls this the “doorknob/DOORKNOB problem” (p. 127).

<sup>5</sup> That is, involves computations on mental representations.

As far as the natural-kind concepts Fodor has an interesting story to tell. He suggests that “[I]t’s intuitively plausible, phylogenetically, ontogenetically, and even just historically, to think of natural kind concepts as late sophistications that are somehow constructed on a prior cognitive capacity for concepts of mind-dependent properties” (p. 153). He also makes a distinction between the atomistic pre-theoretic concept of water, which is acquired “by locking to being water via its superficial signs,” and the post-theoretic “full-blown, chemical concept of water.” In the latter case, Fodor claims “*we are locked to water via a theory that specifies its essence, so we’re locked to water in every metaphysically possible world*” (p. 157). But what exactly does it mean *to lock to a concept via a theory*? If Fodor believes that it is correct to talk about concepts being embedded in theories that constrain them, how are these theories acquired? A primitive basis of mental particulars, no matter how big, is not adequate to account for the acquisition of theories.

### **Implications for a theory of learning**

We have argued that neither Bowker and Star nor Fodor have succeeded in persuading us that traditional psychological approaches to concepts are useless. They have succeeded, however, in showing that there are important issues about a theory of concepts with implications for learning that have not been adequately approached by traditional psychological views.

Psychologists’ interest on how concepts and categories are formed and what they are has not allowed them to pay attention to the way concepts and categories are used and the consequences they have for our lives. Bowker and Star’s pragmatist analysis of classification is very creative and points in a direction that can be productive for a theory of learning. Cognitive psychologists have shown the importance of prior knowledge in learning something new. Today, social constructivism draws our attention to contextual and social factors on learning. Discourse analysis is used to highlight important social processes of negotiation between students and teachers and between peers that influence the way new information is interpreted and what is remembered. Cultural artifacts have also become important objects of learning

analysis. The time is ripe for yet another move in the direction of a systematic investigation of the material force of something that is less concrete than language and artifacts but equally powerful – the conceptual and organizational systems that underlie the contexts where learning takes place, from formal schooling to information technology, from the broader epistemological frameworks to the most simple systems for data organization in our personal computers. Bowker and Star's book is persuasive in suggesting that such a research program will be rich in its outcomes, providing new understandings and enabling us to design better learning environments in the future.

On the issue of what concepts are and how they acquire their content, Fodor's critique is a reminding that a great deal more work needs to be done. Fodor's bold move opens up new possibilities for a theory of concept acquisition that relates minds to the world in ways that would have been unthinkable a few years ago. Fodor's theory does justice to the intuition that our concepts are about the external world and stem from it, that we are not trapped in an internal world defined exclusively in terms of conceptual relations – it does justice to the intuition that our concepts manage to reach out to the world because they are causally connected to it. It is rather interesting that Fodor, the ultimate cognitivist, defends a position about concept individuation according to which the possession and acquisition of lexical concepts are causal and not cognitive processes. His proposal leaves many questions open but provides interesting ideas that need to be further explored, not only at the philosophical level, but by bringing together all the available evidence from psychology and neuroscience on concept formation, human development and animal cognition.

## References

- Carey, S., & Spelk, E. (1996). Science and core knowledge. *Philosophy of Science*, 63, 515-533.
- Dewey, J. (1929). *The quest for certainty: A study of the relation of knowledge and action*. New York: Minton, Balch.
- Fodor, J. (1990): *A theory of content and other essays*. Cambridge, MA: MIT Press.
- Jacob, P. (1997). *What minds can do: Intentionality in a non-intentional world*. NY: Cambridge University Press.
- Laurence, S., & Margolis, E. (1999). Review of *Concepts: Where Cognitive Science Went Wrong*. *British Journal of Philosophy of Science*, 50, 487-491.
- Smith, E., & Medin, D.L. (1981). *Categories and concepts*. Cambridge: MA, Harvard University Press.
- Star, S.L., & Griesemer, J. (1989). Institutional ecology, 'translations', and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 19, 387-420.
- Taylor, J. R. (1995). *Linguistic categorization: Prototypes in linguistic theory (2<sup>nd</sup> Ed.)*. Oxford: The Clarendon Press.